Construction and Use of Cognitive Maps in Model-Based Control

Ata Karagoz

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Cognition and Perception Commons, and the Quantitative Psychology Commons

Recommended Citation
Karagoz, Ata, "Construction and Use of Cognitive Maps in Model-Based Control" (2022). Arts & Sciences Electronic Theses and Dissertations. 2824.
https://openscholarship.wustl.edu/art_sci_etds/2824

This Thesis is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Construction and Use of Cognitive Maps in Model-Based Control

by

Ata Baris Karagoz

A thesis presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree of

Master of Arts

December 2022
Saint Louis, Missouri
Contents

List of Tables ......................................................... iv

List of Figures ...................................................... v

Acknowledgments ..................................................... ix

Abstract ............................................................. x

1 Introduction ......................................................... 1

2 Method .............................................................. 6
  2.1 Pre-registration .................................................. 7
  2.2 Behavioral representational similarity task ...................... 7
  2.3 Decision-making task ........................................... 8
  2.4 Memory task .................................................... 10
  2.5 Behavioral representational similarity analysis .................. 11
  2.6 Reinforcement learning model .................................. 12
    2.6.1 Model-free system ........................................... 13
    2.6.2 Model-based system ......................................... 14
    2.6.3 Model fitting procedure ..................................... 17
  2.7 Exploratory PCA ............................................... 18
  2.8 Experiment presentation software .............................. 18
List of Tables

A.1 Parameter recovery ......................................................... 41
A.2 Repeated-measures ANOVA ................................................. 44
List of Figures

1.1 **Task schematic.**  
A. Task transition structure. Four distinct first-stage states (depicted in the top and bottom row) each contain two unique choice objects. Each of these objects deterministically leads to a one of two second-stage states (depicted in the middle row), as depicted by the colored arrows. These first-stage states were associated with different amounts of reward that changed across the task. For the two ‘high-reward’ states (top row), 80% of the trials involved a high stake, with a multiplier cue indicating that any points would be multiplied by five. For the other two ‘low-reward’ first-stage states (bottom row), high-stake trials occurred on only 20% of trials.  
B. Representative trial. First the participant is shown the first-stage state and the stake multiplier and then is shown the objects. After selecting one of these, the trial transitions to the corresponding second-stage state, where they receive 7 space treasure pieces after selecting the second-stage object. The high-stake multiplier amplifies this to 35 points.  
C. Representative trials of the pre- and post-behRSA task. Participants are shown each novel object pairing twice.  
D. A portion of a hypothetical matrix of similarity ratings generated from the behRSA task.
1.2 **Model matrices and example participants.** Representational similarity matrices. The top row depicts hypothetical model matrices. The bottom row depicts sample behRSA data from participants that are well captured by the corresponding model matrix. In each matrix, each cell corresponds to a similarity measure between two objects (defined by row and column). The diagonal cells correspond to self-similarity and are assigned full similarity by default. **A.** Visual cooccurrence model of representational change in which objects that appear together in the same first-stage state become highly similar, and objects that appear in second-stage states become mildly similar. **B.** Direct item association model of representational change where first-stage objects become more similar to the second-stage item they lead to. **C.** Indirect item association model where first-stage objects that all lead to the same second-stage state become more similar to each other. **D.** Example participant with high visual cooccurrence model fit. **E.** Example participant with high direct item association model fit. **F.** Example participant with high indirect item association model fit.

3.1 **Model matrix fits.** **A.** Participants’ model matrix fits for the three hypothesized models. Model matrix fit is calculated using a multiple regression that produces 3 coefficients for each participant, one for each hypothesized model. All three of the models are represented within the sample. **B.** Participants on average show a higher degree of cognitive map-based abstraction for the high-stake vs low-stake context. (* represents $p < 0.05$, error bars are 95% CI)
3.2 **Principal components analysis of similarity ratings.** Inspecting the first 3 principal components of participants’ behRSA data we find that the principal component that drives most variance (Principal component 1) is similar to an equal mixture of our *a-priori* direct and indirect item association models. The second principal component reflects a mixture of a negative similarity for objects visual cooccurrence combined with indirect and direct item associations. Finally, the third principal component is a mixture of direct item association and visual cooccurrence models. Colorbar depicts similarity structure in arbitrary units.  

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

3.4 **Post-task surprise memory probe.** Participants show a difference in their mismatch $d'$ sensitivity scores indicating that mismatch trials involving high-reward context backgrounds were easier to discriminate than mismatch trials involving low-reward context backgrounds. This effect is not seen in the lure $d'$ sensitivity scores where low- and high-reward context backgrounds are equally discriminable. (* represents $p < 0.05$, error bars are 95% CI.)
A.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)

A.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 $p < 0.05$)
Acknowledgments

I would like to thank my committee: Dr. Zachariah Reagh, Dr. Wouter Kool, and Dr. Jeff Zacks for their helpful insights and guidance in this work.

This work would not have been possible without the help of partner, my family, and my incredibly supportive lab. Thank you for all you do. Mostly I want to thank my mentors for their guidance and willingness to take this project on.

A special thanks goes to the many graduate students, staff, post-doctoral fellows, and faculty that have discussed this work and reviewed this thesis and helped support the related research.

Ata Baris Karagoz

Washington University in Saint Louis

December 2022
When making decisions, we sometimes rely on habit and at other times plan towards goals. Planning requires the construction and use of an internal representation of the environment, a cognitive map. How are these maps constructed, and how do they guide goal-directed decisions? We coupled a sequential decision-making task with a behavioral representational similarity analysis approach to examine how relationships between choice options change when people build a cognitive map of the task structure. We found that participants who encoded stronger higher-order relationships among choice options showed increased planning and better performance. These higher-order relationships were more strongly encoded among objects encountered in high-reward contexts, indicating a role for motivation during cognitive map construction. In contrast, lower-order relationships such as simple visual co-occurrence of objects did not predict goal-directed planning. These results show that humans actively construct and use cognitive maps of task structure to make goal-directed decisions.
Chapter 1

Introduction

A fundamental question of human cognition is how we are able to effectively plan towards goals. Decisions are not made in a vacuum, but rather capitalize on the structure of the world around us. Imagine moving to a new city and learning about its structure as you navigate it. Ideally, your internal model will be set up to guide effective planning, strongly encoding relationships among major streets that connect different neighborhoods. However, this model may also incorporate features that are less relevant for planning. For example, you may strongly encode relationships between streets with similar sounding names or that happen to be located close together. Here, we ask how humans learn which features of the environment are relevant for planning and apply these representations to optimize behavior. In other words: how do we build and use internal models to guide our decisions?

These internal models are commonly referred to as 'cognitive maps', a term coined by Tolman to explain how rodents were able to use spatial features of a maze to navigate towards goals[65]. The discovery of place cells and grid cells, which indicate an animal’s current location in space[41, 28] and which are used in simulating possible future trajectories[30] provide the neural underpinnings for cognitive maps. Recent work suggests that cognitive maps are not restricted to spatial domains (see Behrens et al. for review[1]). Instead, they seem to also encode more abstract maps, such as those involving social relationships[42,
43, 63], transitions between tasks[55, 56], and abstract associations between items[10, 64]. However, much remains unclear about how humans construct abstract cognitive maps, how they use them to plan towards goals, and whether the quality of these maps relates to goal-directed behavior.

Recent studies of decision making formalize goal-directed planning over a cognitive map as "model-based" reinforcement learning[62, 13, 20, 25]. A model-based learner computes expected values for available actions by using a representation of the task structure. This flexible but computationally costly strategy enables it to apply the values learned at a given goal to every path that leads to that goal. Meanwhile, a model-free learner uses a more efficient but inflexible strategy, only updating the value of actions that led to reward, without considering the structure of the task.

Here, we assess the construction of cognitive maps to drive goal-directed decisions. To do so, participants performed a variant of the 'two-step' task[13, 12, 33, 34]. This task dissociates model-based and model-free control by exploiting the ability of the model-based system to plan using an internal representation of the task structure, which contrasts with the model-free reliance on direct action-reward associations. Many prior studies using this task assume or ensure that an effective representation of the task is present (but see Feher da Silva and Hare[21]). However, individual differences in cognitive maps may critically influence differences in model-based control. How does the nature and quality of one’s cognitive map influence behavior?

To index cognitive map structure in a purely behavioral setting, we developed a novel approach, which we termed behavioral representational similarity analysis (behRSA). Our approach was inspired by neuroimaging analyses[36, 17] that assess pairwise neural pattern similarity between stimuli. Here, participants simply rated the perceived relatedness of pairs
of components before and after learning the two-step task. This allowed direct insights into representational similarity among these components as defined by the participants’ behavioral output. We assessed the amount of planning-relevant versus planning-irrelevant information incorporated into their maps, and whether motivation influenced map formation.

We found that participants whose similarity ratings reflected the higher-order transition structure earned more reward and used more model-based control, but we found no such relationship for components reflecting more superficial structure. Moreover, such structure learning was amplified for contexts in which more reward was at stake. Our work introduces a method for inferring internal representations on the basis of behavioral ratings alone, and provides insight into the way people construct and plan over these representations.
Figure 1.1: **Task schematic.** A. Task transition structure. Four distinct first-stage states (depicted in the top and bottom row) each contain two unique choice objects. Each of these objects deterministically leads to one of two second-stage states (depicted in the middle row), as depicted by the colored arrows. These first-stage states were associated with different amounts of reward that changed across the task. For the two ‘high-reward’ states (top row), 80% of the trials involved a high stake, with a multiplier cue indicating that any points would be multiplied by five. For the other two ‘low-reward’ first-stage states (bottom row), high-stake trials occurred on only 20% of trials. B. Representative trial. First the participant is shown the first-stage state and the stake multiplier and then is shown the objects. After selecting one of these, the trial transitions to the corresponding second-stage state, where they receive 7 space treasure pieces after selecting the second-stage object. The high-stake multiplier amplifies this to 35 points. C. Representative trials of the pre- and post-behRSA task. Participants are shown each novel object pairing twice. D. A portion of a hypothetical matrix of similarity ratings generated from the behRSA task.
Figure 1.2: Model matrices and example participants. Representational similarity matrices. The top row depicts hypothetical model matrices. The bottom row depicts sample behRSA data from participants that are well captured by the corresponding model matrix. In each matrix, each cell corresponds to a similarity measure between two objects (defined by row and column). The diagonal cells correspond to self-similarity and are assigned full similarity by default. A. Visual cooccurrence model of representational change in which objects that appear together in the same first-stage state become highly similar, and objects that appear in second-stage states become mildly similar. B. Direct item association model of representational change where first-stage objects become more similar to the second-stage item they lead to. C. Indirect item association model where first-stage objects that all lead to the same second-stage state become more similar to each other. D. Example participant with high visual cooccurrence model fit. E. Example participant with high direct item association model fit. F. Example participant with high indirect item association model fit.
Chapter 2

Method

We recruited 209 healthy younger (range: 18-35 years old) adults for this study using the Cloud Research Platform and Amazon Mechanical Turk. We excluded 48 participants based on preregistered criteria. We excluded 32 participants because they failed to respond in more than 20% of the trials in the decision-making task. We excluded 12 participants because they failed to respond in more than 20% of trials in the similarity task. We also excluded participants if they engaged in purely random action selection[44]. To determine this, we simulated 1000 agents that responded randomly on each trial, yielding a distribution of log-likelihoods expected under random action selection. We excluded 2 participants whose best-fitting log-likelihood exceeded the 5th percentile of this distribution (negative log-likelihood score > 172.846). We also excluded 2 participants because of missing data in their behavioral representation similarity task, rendering certain pairwise comparisons impossible. This resulted in an effective sample of 161 younger adults (102 male, 58 female, and 1 non-binary, age range: 18-35 years, mean age: 29.2 years). We stopped collecting data once our effective sample size reached 160, which allowed us to detect a correlation with the same strength of our initial study ($r = 0.2764$) with 95% power.
All participants were compensated with a base payment of $9 and additional performance-related payments of 1 cent for every 25 points obtained in the decision-making task. All participants gave informed consent, and procedures were approved by the Washington University in St. Louis Institutional Review Board.

2.1 Pre-registration

This paper describes a pre-registered replication of a pilot sample which can be found here https://osf.io/uw3p7

2.2 Behavioral representational similarity task

In order to test how task experience and motivation affects structure learning, participants provided relatedness rating of pairs of novel 3D objects that were also used as choice options in the decision-making task (adapted with permission from[29, 53]). On each trial, they were shown an initial object for 1 second, followed by a 500ms fixation cross, and then another object for 1 second. Participants were then shown a ‘slider’ bar and were asked to move the slider to their perceived level of ‘relatedness’ between the two objects they had just seen (Figure 1.1C). Participants had 5.5 seconds to provide a response. All objects were presented in the center of the screen, with image sizes of 400 x 400 pixels. Participants performed all pairwise ratings of all 10 objects in both orders, resulting in 90 trials. Participants performed the behavioral representational similarity task both before and after the decision-making task. This allowed for a pre- vs. post-learning comparison of object relatedness judgments.
2.3 Decision-making task

The decision-making task was designed to dissociate between model-free and model-based
decisions in a setting where participants need to learn about the structure of the task, and
was based on a recently developed “two-stage” task19. Each trial of the decision-making task
started randomly in one of four first-stage states. Each of these states offered a choice
between a unique pair of ‘teleporters’, presented side-by-side. Participants used the ‘F’ key
on their keyboard to choose the left teleporter, and the ‘J’ key to choose the right teleporter.
This choice determined which one of two second-stage states would be encountered. For
each pair, one of the teleporters deterministically led to a purple second-stage state, and the
other deterministically led to a red second-stage state. Importantly, each teleporter always
led to the same second-stage state (Figure 1.1A).

Each second-stage state contained a unique ‘generator’ that was associated with a scalar
reward. Participants were instructed to press the spacebar key to interact with the generator
so that it provided them with ‘space treasure’, and they were told that the fuel rods used
by the generators would sometimes yield more or less space treasure. The payoffs of the
generators changed over the course of the experiment according to independent random
walks. Their reward distributions were initialized randomly for each participant within a
range of 0 to 9 points and then varied according to a Gaussian random walk ($\sigma = 2$) with
reflecting bounds at 0 and 9.

The 10 objects were randomly assigned as teleporters and generators for each participant
separately.

The first-stage states were not only uniquely identifiable by the pair of teleporters, but they
also were associated with a set of background images. Specifically, each of the first-stage
states was always presented with a background image that belonged to one of four ‘location’
categories. Two of the first-stage states were always shown with background images of ‘inside’
locations (libraries and restaurants), and the other two with images of ‘outside locations’
(deserts and forests). Each of these categories contained four images. Their presentation
was selected pseudorandomly, so that each background image was shown equally frequently.

In order to introduce differing incentives for model-based control and demands for learning
the task structure, we introduced a “stakes” manipulation in this task. At the start of each
trial, an incentive stake cue indicated by how much the reward obtained at the end of the
trial would be multiplied. On some trials, the stake cue was a 5x multiplier (high-stakes)
and on others a 1x multiplier (low-stakes). For example, if a participant earned 5 space
treasure pieces on a high-stakes trial, the multiplier would result in a total of 25 points.
Importantly the chances of a high-stake trial were different between the first-stage states.
For each participant, either the ‘inside’ or ‘outside’ first-stage states were selected as ‘high-
stakes’ states, and the other as the ‘low-stakes’ states. In the high-stakes states there was
an 80% chance of a 5x multiplier, while in the low-stakes states there was a 20% chance of
a 5x multiplier.

At the start of each trial, participants saw the category background and the stake multiplier
for 1 second. Then, the stake moved to the top left corner and the teleporters were presented,
and participants were given a time limit of 1.5 seconds to choose between them. After the
response, the selected option was highlighted and the non-selected option was greyed-out for
the remainder of the response period. There was a 500ms interval between the end of the
first stage response period and the onset of the second stage. Following a 200ms interval
after the generator was selected, participants were shown how many space treasure pieces
they earned for 1.5 seconds. Then each piece was converted to points and added to the
total score (100ms each point). There was a 500ms intertrial interval (ITI). Participants completed a total of 256 trials with an optional short break in the middle.

Participants were extensively instructed on the decision-making task. They first were first familiarized with a set of practice objects and contexts. There were 2 first-stage practice contexts and 2 second-stage practice contexts. Each practice first-stage had a pair of unique practice objects, and each practice second-stage had a unique practice object. Participants would learn that in each initial state there are a pair of unique novel objects that appear side by side and that these objects are called ‘teleporters’. They were taught that each teleporter would lead deterministically to one planet or another. Participants were then asked to practice which object led to which second stage state. We informed participants that the practice objects and contexts are different from those used in the main task, but that the rules governing the transition structure would be similar.

2.4 Memory task

Participants also completed a surprise memory test where they were shown a pairing of a background image and a pair of objects (Figure 1.1C). They were asked to indicate whether they saw this pairing before. Specifically, they were given 5 seconds to respond on a four point scale consisting of ‘Yes, Sure’, ‘Yes, Unsure’, ‘No, Unsure’, ‘No, Sure’. After a 500ms ITI, a new stimulus/background pairing was shown. Participants were shown all 16 background/object pairings they saw in the task (4 for each planet x 4 planets). They were also presented with ‘lure’ trials. In 12 of these, participants saw choice objects of one state presented on a background of another state (i.e. the desert objects in front of a forest they have seen). In the remaining 16 lure trials, a pair of choice objects were shown on a new
image from each category of background (i.e., desert objects in front of a desert they have never seen). This results in a total of 44 trials.

2.5 Behavioral representational similarity analysis

We hypothesized that experience with the decision-making task would yield participants to represent the task structure across different levels of abstraction. To test this hypothesis, we compared their matrices of relatedness ratings to three *a-priori* model matrices that reflected different levels of representation of the decision-making task structure (Figure 1.2A, 1.2B, 1.2C). The first of these (visual cooccurrence) was designed to capture an increase in relatedness between objects that were viewed together in the same first-stage states, (e.g., the two teleporters from the desert first-stage state). The second (direct outcome association) captured any increase in relatedness between first-stage objects and the second-stage object that they led to, (i.e. the teleporter from the desert that led to the red planet and the red planet’s generator). The third model matrix (Indirect item-outcome association) captured increases in relatedness between all the teleporters that led to the same second-stage state. Note that these model matrices differ in the degree to which they reflect associations that are useful for goal-directed control. Specifically, the direct outcome association and the indirect item-outcome models encode the consequences of choices, whereas the visual cooccurrence model encodes a more superficial aspect of the task that does not reflect such higher-order structure. For each model matrix, cells that reflect the hypothesized relatedness relationship were coded as 1, and cells that did not were coded as 0 (with the exception that the two second-state stages were coded as 0.5 with relation to one another in the visual cooccurrence model). It is important to note that these model matrices are agnostic to the identity
of specific objects in the task structure, because these were randomly assigned for each participant. In sum, each matrix cell indicates a pairwise similarity rating between two objects, and the hypothesized matrices reflect three predicted ways in which participants could change their perceived similarity among objects based on learning this task structure.

We calculated the average relatedness ratings for each pair of objects separately for the pre- and post-task ratings. We then subtracted the pre-task averages from the post-task averages to get a measure of representational change. Finally, we fit a multiple regression model for each participant using the three hypothesized models of behavioral similarity, allowing us to extract $\beta$ coefficient values for each participant’s data fit to each model matrix:

$$y = \beta_{\text{visual cooccurrence}} x_0 + \beta_{\text{direct item}} x_1 + \beta_{\text{indirect item}} x_2$$

We also tested whether the task structure was learned differently for the high-stakes and low-stakes first-stage states. To do so, we split the model relatedness matrices along high-stakes contexts and low-stakes contexts and fit separate multiple regression models, resulting in six coefficient values per subject (one for the low-stakes and high-stakes arm for each of the three models). We then subtracted the high-stakes arm coefficient from the low-stakes arm coefficient to provide a difference score (Figure 3.1B).

## 2.6 Reinforcement learning model

We adapted an established hybrid reinforcement-learning model to assess participants’ behavior in the decision-making task, specifically dissociating model-free and model-based decision making. Every trial $t$ started out in one of four first-stage states $(s_{1,t})$ where one of
two possible actions $a_A$ and $a_B$ could be selected ($a_{1,t}$). Depending on their selection, the participant deterministically transitioned to one of two second-stage states ($s_{2,t}$) where they could perform only one action ($a_{2,t}$) and then obtain a reward ($r_t$).

The model described here contains both a model-free learner and a model-based learner that learn expectations of long-term future reward $Q(s, a)$ for each combination of state and action. The model-free system learns reward expectations for each of the eight teleporters and two generators, by updating their values based on reward prediction errors. The model-based system, on the other hand, learns a transition structure that represents to which planet each teleporter leads. It then combines this with the model-free reward expectations of the terminal, second-stage states to select between teleporters.

### 2.6.1 Model-free system

All model-free reward expectations were instantiated with a reward expectation of 4.5 (arithmetic mean of minimum and maximum possible reward) to all actions and states. The model-free learner would then use the $SARSA()$ temporal difference learning algorithm to update its cached reward expectations based on the difference between predicted and received rewards. In the decision-making task this resulted in a reward prediction error ($\delta$) being calculated at each stage according to:

\[
\delta_{1,t} = Q_{MF}(s_{2,t}, a_{2,t}) - Q_{MF}(s_{1,t}, a_{1,t})
\]

\[
\delta_{2,t} = r_t - Q_{MF}(s_{2,t}, a_{2,t})
\]
Notice that the second-stage prediction error incorporates the immediate reward outcome for that trial, but that the first-stage prediction error only incorporates expectations of future reward. The values of each prediction error were then used to update the reward expectations of the model-free learner at both the first and second stage:

$$Q_{MF}(s_{1,t}, a_{1,t}) \leftarrow Q_{MF}(s_{1,t}, a_{1,t}) + \alpha \delta_{1,t} + \alpha \lambda \delta_{2,t}$$

$$Q_{MF}(s_{2,t}, a_{2,t}) \leftarrow Q_{MF}(s_{2,t}, a_{2,t}) + \alpha \delta_{2,t}$$

Here, $\alpha$ is the reward learning rate (between 0 to 1) that determines how quickly new information about rewards is incorporated into the model-free learner expectations. The eligibility trace decay parameter $\lambda$ (between 0 to 1) determines how much a reward prediction error experienced after the second stage choice changes first-stage reward expectations.

### 2.6.2 Model-based system

The model-based system combines learns the transitions structure of the task, and uses this to flexibly compute reward expectations for each available teleporter. Specifically, it learns a transition matrix $T(s_{1}, a_{1})$ that encodes the probability of moving to the second-stage state $s_{2}$ after choosing the action $a_{1}$ in the first-stage state $s_{1}$. In order to compute the model-based reward expectations, these probabilities were then combined with the reward expectations at the second stage:

$$Q_{MB}(s_{1,t}, a_{1,t}) = \sum_{s_{2}} T(s_{1,t}, a_{1,t})Q_{MB}(s_{2}, a_{2})$$
\[ Q_{MB}(s_{2,t}, a_{2,t}) = Q_{MF}(s_{2,t}, a_{2,t}) \]

Because participants were not trained on the transition structure of the task, we initialized their model of the structure without any knowledge of the correct transitions, setting \( T(s_1, a_1) = 0.5 \) for all combinations of first-stage actions and second-stage states. We assumed that participants learned of the transition structure by computing a state prediction error \( \delta^{SPE} \) as

\[ \delta_t^{SPE} = 1 - T(s_{1,t}, a_{1,t}) \]

reflecting the difference between the actual transition and its expectation.

The state prediction error is then used to update the cell in the transition matrix corresponding to the experienced transition. Transition probabilities must sum to 1, so the cell corresponding to a transition to the alternative second stage \( \neg s_{2,t} \) is reduced by the same value. Because participants may learn the transition structure at different rates a transition learning rate parameter \( \eta \) (ranging from 0 to 1) was defined to weight how quickly the transition structure is learned. This results in the following transition matrix update equations:

\[ T(s_{1,t}, a_{1,t}) \leftarrow T(s_{1,t}, a_{1,t}) + \eta \delta_t^{SPE} \]

\[ T(\neg s_{1,t}, a_{1,t}) \leftarrow T(\neg s_{1,t}, a_{1,t})(1 - \eta) \]
In addition to updating the transition structure for the experienced action, we reasoned that some participants would make an inference about where the alternate action in the first-stage state would have led. Therefore, we implemented a parameter, $\kappa$ (ranging from 0 to 1) which would update the probability of the counterfactual transition according to the following rules:

$$\delta_t^{CF-SPE} = 1 - T(s_{1,t}, a_{1,t})$$

$$T(s_{1,t}, -a_{1,t}) \leftarrow T(s_{1,t}, -a_{1,t}) + \kappa \delta_t^{CF-SPE}$$

$$T(s_{2,t}|s_{1,t}, -a_{1,t}) \leftarrow T(s_{2,t}|s_{1,t}, -a_{1,t})(1 - \kappa)$$

Choice rule. The model-free and model-based learners reward expectations in the first stage are integrated using a model-based weighting parameter $w$ (ranging from 0 to 1) using the following rule:

$$Q_{net}(s_1, a_1) = (1 - w)Q_{MF}(s_1, a_1) + wQ_{MB}(s_1, a_1)$$

We then used a softmax function to map the reward expectation to choice probabilities:

$$P(a_{1,t} = a_1|s_{1,t}) = \frac{\exp(\beta Q_{net}(s_{1,t}, a_1) + \pi^{rep}(a_1) + \rho^{resp}(a_1))}{\sum_{a_1'} \exp(\beta Q_{net}(s_{1,t}, a_1') + \pi^{rep}(a_1') + \rho^{resp}(a_1'))}$$

Here, $\beta$ is the inverse softmax temperature (left-bounded to 0) that determines how much influence reward expectations have on choice probabilities and can be thought of as a measure.
of exploration and exploitation. High softmax temperatures mean that the model is more likely to explore and low softmax temperatures mean the model more commonly exploits its knowledge. In order to capture choice perseveration, we included two parameters to capture both response key and stimulus ‘stickiness’. The variable $\text{rep}(a_1)$ is defined as 1 if $a_1$ was the action that was chosen on the previous trial and 0 otherwise. The choice stickiness parameter $\pi$ (left unbounded) related to choice perseveration when positive and choice switching when negative. The variable $\text{resp}(a_1)$ was defined as 1 if the action $a_1$ could be selected with the same response key that was used in the previous trial and 0 otherwise. The response stickiness parameter $\rho$ captured perseveration of the response key press when positive and switching of response key press when negative. Together this results in a model with 8 free parameters.

In addition to the standard model we sought to ascertain if there were differences in the amount of model-based control participants were exhibiting across the different conditions we fit a model with a separate $w$ parameter for each combination of trial type (low or high) with context (low stakes common or high stakes common). This resulted in four model-based weight parameters ($w_{\text{lowTrial,lowContext}}, w_{\text{highTrial,lowContext}}, w_{\text{lowTrial,highContext}}, w_{\text{highTrial,highContext}}$) which when combined with the other parameters resulted in 11 free parameters.

### 2.6.3 Model fitting procedure

For each participant we obtained maximum a posteriori (MAP) estimates of the free parameters in the model, using custom scripts coupled with the scipy.optimize.minimize function. All parameters bounded between 0 and 1 ($\alpha, \lambda, \eta, \kappa, w$) used a $\text{Beta}(2, 2)$ prior, for the inverse
softmax temperature $\beta$ we used a $Gamma(3, 0.2)$ prior, and for the two stickiness parameters ($\pi$ and $\rho$) we used $Normal(0, 1)$ priors. In order to avoid local optima we randomly initialized the parameters and performed the optimization procedure 10 times per participant. We then selected the parameters of the run with the highest posterior probability. Analyses on parameter recovery were performed to assess the model fitting procedure. These are described in Appendix: Parameter Recovery.

2.7 Exploratory PCA

To reconstruct the data-driven model matrices that would describe the most shared variance among the participants we used a principal components analysis (PCA). The lower diagonal, excluding the diagonal itself, for each participants representational similarity matrix was flattened into a vector. These vectors were combined into a subject x data matrix which we extracted the principal components from using singular value decomposition (SVD). We created a scree plot to assess the explained variance ratios for all of the principal components output by the function (Fig 4B). This plot indicates which of the principal components explain more than 5% of the variance, in our data this was the first 3 principal components. We then took each of these and remapped them into the original symmetric matrix to get an assessment of what the change along that principal component dimension was (Fig 4A).

2.8 Experiment presentation software

Both the experiment presented here and the pilot experiment were programmed in JavaScript for online presentation using the jsPsych (6.3.0) package [15].
2.9 Task data analysis

All data were analyzed using Python. Statistical analyses were performed using SciPy (1.6.1), pingouin (0.5.0) and scikit-learn (1.0.0) packages [69, 67, 46]. All reported statistical tests were two-sided. Additional analyses were performed using custom Python scripts which are available in the repository on GitHub (github.com/cdm-lab/mb-cog-maps-paper/).

2.9.1 Analysis of decision-making data

We computed average performance on the decision-making task as the average number of points earned per trial. To correct for baseline differences in available reward (as a result of the random Gaussian walks), we then subtracted the average available reward across both second-stage states. Participants’ data were also fit using a reinforcement learning model (described in Reinforcement learning model).

2.9.2 Analysis of memory probe data

To assess memory precision for the participants we used $d'$. First, one obtains the hit rate by taking the hit rate ($H$) which is the proportion of target trials that a subject correctly identifies as old. Next, one calculates the false alarm rate ($F$) which is the proportion of lure trials that a subject incorrectly identifies as old. The $d'$ measure is then calculated as the following equation, $d' = z(H) - z(F)$. We used this equation to calculate two separate $d'$ measures, one for lure trials and one for mismatch trials. The first was calculated using a false alarm rate for true lure trials, which we call $d'_{\text{lure}}$. The second was calculated using a false alarm rate from the mismatch trials, we called this $d'_{\text{mismatch}}$. We calculated these
measures separately for high and low stakes contexts. The $d_{\text{mismatch}}$ value was then compared for the high and low stakes contexts using a paired samples t-test.

2.10 Initial Experiment

The Appendix describes the initial experiment in more detail. This study was largely similar to the pre-registered replication experiment, but there were a few differences in their design.
Chapter 3

Results

We combined a reinforcement-learning task with behRSA to measure how experience and motivation influence the representation of task structure. Participants (n = 161) performed pairwise similarity ratings on ten novel objects[29, 53] before and after encountering them in a sequential decision-making task that distinguishes model-free and model-based control (Figure 1). This is a pre-registered replication of a prior experiment (see Appendix: Cohort 1).

The decision-making task, modeled after an established paradigm[12, 33, 34, 19], required participants to learn its structure in order to plan towards reward (Figure 1.1A, 1.1B). Each trial started pseudorandomly in one of four first-stage states, where participants chose between two objects. This choice determined which second-stage state – a red or a purple ‘planet’ – would then be encountered. For each first-stage state, one object always led to the red planet, and the other always to the purple planet. On each planet, participants then interacted with an object that provided a scalar reward which slowly changed over time.

This task distinguishes between model-based and model-free strategies, since only a model-based decision maker generalizes experiences from one starting state to all other starting
states. That is, after receiving a high reward on a planet, a model-based learner can use its knowledge of the transitions to plan its way back to that same second-stage state. A model-free agent, on the other hand, learns through action-reward associations and will only become more likely to choose that same action in the same first-stage state, not transferring experiences from one first-stage state to the others [33, 19].

We used maximum \textit{a-posteriori} estimation [4, 22] to fit a dual-system reinforcement-learning model to behavior on this task (Methods: Reinforcement learning model). This model describes behavior as a mixture of model-free and model-based control weighted according to a mixture parameter \( w \). This parameter is fit closer to 1 for pure model-based control and closer to 0 for pure model-free control. Mirroring prior work [12, 33], we found that participants’ behavior reflected a mixture of model-free and model-based control (mean \( w = 0.57 \)). This suggests that participants learned an internal representation of the task, a cognitive map, and that they used this for goal-directed decision making.

### 3.1 Behavioral indices of representational change track task structure.

We used participants’ relatedness ratings of objects to measure the structure of their cognitive maps (Figure 1.1C, 1.1D). Specifically, we formulated three ways in which objects could become more ‘related’ through task experience. First, we hypothesized that objects could become related if they co-occurred in a first-stage state (Figure 1.2A). Such a representational shift, although capturing some task structure, is not useful for planning because it does not reflect the consequences of actions. We also hypothesized two forms of representational
shifts that related to the task’s transitions (Figure 1.2B, 1.2C). In one case, we hypothesized that first-stage state objects and the second-stage objects they lead would become related. This representation, which we call a “direct item association”, allows for planning from a first-stage action to the related second-stage state. We also hypothesized that all first-stage objects leading to the same second-stage state would become related. We call this representation an “indirect item association”, because it encodes relations between objects that never occurred on the same trial. Critically, these indirect associations would indicate that participants are abstracting beyond immediate experience in a way that goes above a simple understanding of action-outcome contingencies in the task.

To measure how strongly participants encoded these aspects of the task structure, we formalized these three representational shifts from pre- to post-task behRSA data with pre-registered hypothetical model matrices. Then, we used multiple linear regressions models to fit the changes in behRSA data to these matrices, for each participant separately (Figure 1.2). This produces three regression coefficients ($\bar{\beta}_{\text{visual cooccurrence}}$, $\bar{\beta}_{\text{direct item}}$, $\bar{\beta}_{\text{indirect item}}$), each reflecting the strengths of one of the hypothesized representational shifts.

At the group level, participants judged item similarity in a manner consistent with each model matrix (example subjects with strong fits to each coefficient can be seen in Figure 1.2D, 1.2E, and 1.2F respectively). As depicted in Figure 3.1A, participants judged objects as more related when they had occurred in the same first-stage state ($t_{(160)} = 4.64$, $p < 0.001$, $d = 0.37$), when they constituted a pair where one object transitioned to the other ($t_{(160)} = 5.98$, $p < 0.001$, $d = 0.47$), and when they both led to the same second-stage state ($t_{(160)} = 5.29$, $p < 0.001$, $d = 0.42$). In other words, all three hypothesized components of the task were represented at the group-level.
Figure 3.1: Model matrix fits. A. Participants’ model matrix fits for the three hypothesized models. Model matrix fit is calculated using a multiple regression that produces 3 coefficients for each participant, one for each hypothesized model. All three of the models are represented within the sample. B. Participants on average show a higher degree of cognitive map-based abstraction for the high-stake vs low-stake context. (* represents $p < 0.05$, error bars are 95% CI)
Figure 3.2: **Principal components analysis of similarity ratings.** Inspecting the first 3 principal components of participants’ behRSA data we find that the principal component that drives most variance (Principal component 1) is similar to an equal mixture of our *a-priori* direct and indirect item association models. The second principal component reflects a mixture of a negative similarity for objects visual cooccurrence combined with indirect and direct item associations. Finally, the third principal component is a mixture of direct item association and visual cooccurrence models. Colorbar depicts similarity structure in arbitrary units.

Next, we conducted a data-driven test of whether our hypothesized components captured the primary patterns that emerged from the relatedness ratings. We performed a principal components analysis on the aggregate behRSA data and inspected the three dimensions that described the most variance (Fig 3.2). The results were largely consistent with our hypothesized relationships. The first principal component corresponded to a mixture of our models of direct and indirect item association. The second corresponded to a combination of dissimilarity of cooccurrence, coupled with increased similarity for the direct and indirect item associations. Finally, the third principal component resembled a combination of increased similarity for visual cooccurrence and direct item associations. This provides strong support for our predicted representations of task structure.
3.2 Representations of task structure correlate with task performance and model-based control.

If the behRSA data measures the cognitive map, and cognitive maps enable planning, then individual differences in these representations of similarity structure should correlate with task performance (average reward rate) and reliance on model-based control (Figure 3.3). We predicted positive correlations for aspects of the task structure that are important for goal-directed planning but not for lower-order relationships.

Consistent with our hypotheses, we did not observe a relationship between performance and the strength of the visual cooccurrence component ($r_{(159)} = 0.004, 95\% CI = [-0.15, 0.16], p = 0.96$). However, we found that performance was positively correlated with the encoding of direct item associations ($r_{(159)} = 0.32, 95\% CI = [0.18, 0.46], p < 0.001$) and indirect item associations ($r_{(159)} = 0.45, 95\% CI = [0.31, 0.56], p < 0.001$).

Next, we found a trending but non-significant negative correlation between model-based control and the strength of the visual cooccurrence component ($r_{(159)} = -0.14, 95\% CI = [-0.29, 0.01], p = 0.0753$). Critically, however, we found that use of model-based control was positively correlated with the encoding of direct item associations ($r_{(159)} = 0.33, 95\% CI = [0.18, 0.46], p < 0.001$), and indirect item associations ($r_{(159)} = 0.48, 95\% CI = [0.35, 0.59], p < 0.001$).

In short, the encoding of higher-order representations of task structure – such as a first-stage item to its second-stage counterpart, or two first-stage items that lead to the same second-stage item – enabled goal-directed planning towards goals, whereas representations of low-level features such as item cooccurrence did not. Taken together, these results suggest that
the behRSA approach captures the construction of internal representation of the transition structure, the cognitive map, and the individual differences in its fidelity.

3.3 Motivation affects representational change.

In order to test the effect of motivation on cognitive map construction, we implemented an incentive manipulation (Figure 1.1). At the start of each trial, a cue indicated by what factor the points earned on that trial would be multiplied (1 vs. 5). Importantly, two first-stage states were associated with a high probability of a high-stakes trial (80%; high-stakes context), and the other with a low probability (20%; low-stakes context). Replicating prior work[34, 4, 45, 57], model-based control was increased on high-stakes trials (mean $w_{\text{high}} = 0.56$) compared to low-stakes trials ($w_{\text{low}} = 0.54$) ($F_{(1,160)} = 6.83, p = 0.008$, see SI: Stake x Arm ANOVA)$\textsuperscript{1}$

We sought to understand whether the difference in incentives between contexts affect the representation of task structure. We predicted that aspects of the task related to higher-order structure - both direct item associations and indirect item associations - would be more strongly encoded for the high-stake compared to the low-stake context items (Figure 1.1A).

To test this, we ran new multiple linear regressions, estimating the coefficients separately for each stake context. We found no difference between the high- and low-stake context representations of visual cooccurrence ($t_{(160)} = 0.45, p = 0.65, d = 0.04$) (Figure 3.1B). We found a non-significant trend toward a context-driven difference in the direct item outcome

$\textsuperscript{1}$Note that these estimates of model-based control are lower than the estimate obtained from the single parameter model. This happens because only a quarter of the data contributes to each of the four model-based weighting parameter estimates, and therefore the prior (with a mode at 0.5) has a stronger influence on the estimates.
representations \( t(160) = 1.74, \ p = 0.08, \ d = 0.14 \). However, the indirect item associations were encoded more strongly for the high-stakes context compared to the low-stake context \( t(160) = 3.22, \ p = 0.0015, \ d = 0.25 \). These results indicate that incentives lead to stronger representations for goal-directed information in task structure.

**Figure 3.3: Model-based representations correlate with decision-making task and reinforcement learning model.** A. Comparison of behRSA representations of subjects with their measure of performance in the decision-making task. Direct item associations and indirect item associations correlate with performance in the decision-making task whereas simple visual cooccurrence does not. B. Reconstruction of task-relevant representations in behRSA also correlate with increased use of model-based control, whereas visual cooccurrence does not. (* represents \( p < 0.05 \))
3.4 Memory for object-background pairings is better in higher-stakes contexts.

Finally, we predicted that enhanced encoding of the task structure in the higher-stakes context would also enhance encoding of peripheral elements of trials in this condition. Therefore, we tested participants’ memory of the object-background pairings encountered in the first stage of the two-step task. Some of these pairings were indeed encountered before (targets), but others consisted of two first-stage objects on a new exemplar from the correct background category (lures), or two first-stage objects on a background from a different first-stage state (mismatch). The mismatch trials probed highly specific memory for object-scene pairs, because participants had encountered all of that information during the task. For each trial, participants indicated whether they had seen that combination of object and background before.

To test this, we computed $d'$ sensitivity scores separately for the lure and mismatch trials. We found no effect of incentive condition for the lure trials ($t = 0.82, p = 0.4152$). For the mismatch condition, however, discrimination was higher for high-stakes compared to low-stakes trials ($t = 2.29, p = 0.0234$) (Figure 3.4). This suggests that components of the task – even for elements that were incidental to maximizing reward – were better learned in high-stakes contexts. In particular, incentives drove enhanced memory for highly specific item-in-context information.
Figure 3.4: **Post-task surprise memory probe.** Participants show a difference in their mismatch $d'$ sensitivity scores indicating that mismatch trials involving high-reward context backgrounds were easier to discriminate than mismatch trials involving low-reward context backgrounds. This effect is not seen in the lure $d'$ sensitivity scores where low- and high-reward context backgrounds are equally discriminable. (* represents $p < 0.05$, error bars are 95% CI.)
Chapter 4

Discussion

Understanding the cognitive mechanisms underlying goal-directed planning is fundamental to the study of human behavior[65, 12, 5, 68, 37, 18, 70]. Here, we aimed to measure one of its key components: the internal representation of task structure, also known as the cognitive map. We aimed to measure these representations at large scale using a purely behavioral approach. To do this, we developed a novel behavioral variant of RSA to determine different sources of variability in how cognitive maps are constructed. Specifically, participants told us how related they thought the choice options they encountered in a sequential decision-making task were. Next, we correlated their ratings with three a-priori hypotheses about how one might infer task structure. This allowed us to assess to which degree participants’ cognitive maps reflected both more superficial and higher-order components of task structure. Strikingly, the principal ways in which participants’ relatedness ratings varied were largely accounted for by our a-priori models. Participants tracked aspects of task structure irrelevant for planning, such as which choice options merely cooccurred. More importantly, they also tracked components of the task that reflect its higher-order nature, such as which options are directly and indirectly linked through the transition structure.
Consistent with the idea that these relatedness ratings allow insights into cognitive maps, we found that participants whose ratings reflect the higher-order structure showed increased deployment of model-based control and better performance on the decision-making task. In contrast, planning-irrelevant components (such as item co-occurrence) did not relate to strategy selection or task performance. Moreover, the representations of indirectly related associations were more strongly present in contexts with increased incentives, further indicating the relevance of these cognitive maps for goal-directed control. The relationship between a first-stage item and the item encountered in the subsequent reward state is more complex than mere co-occurrence in the two-step task. However, one might nonetheless take this kind of representation to indicate that participants simply understand the task’s transition structure. In contrast, the representation of indirect associations suggests not only an understanding of basic causality in the task, but also mapping out the space of possibilities in a more abstract sense. Importantly, this more abstract mapping is not merely a given for individuals who perform well at the task. That is, such a representation goes beyond immediate action-outcome relationships and suggests that some participants are linking across indirectly related experiences. Together, these findings suggest that structured representations of a task are used to guide planning toward decisions.

4.1 Insights into model-based control

The last decade has seen an explosion of research on how humans exert model-based control, often relying on variants of the two-stage task and the dual-system reinforcement-learning model that we used here[12, 33, 34, 4, 45, 57]. Even though the insights from this field of research have been rich, this approach does not typically probe the cognitive map. First, in
the majority of studies, it has either been assumed or ensured that the structure of a task was fully learned (for example[34]). Second, behavior in this class of tasks does not lend insight into the structure and task relevance of the cognitive map a participant has constructed. As a consequence, prior work has overlooked these representations as important sources of variance (but see Feher da Silva and Hare[21]). Our behRSA approach complements this prior work. Despite participants receiving minimal instruction during behRSA task and having to learn the transition structure during task performance, we nonetheless found that their relatedness ratings revealed components of the task structure. Critically, these representations - particularly for abstractions about task structure - correlated with goal-directed control.

4.2 Comparison with successor representation

Representations of task structure play a crucial role in the successor representation[14, 24, 23, 39], a class of reinforcement-learning algorithms that combines the advantages of model-free and model-based control. Specifically, the successor representation stores cached predictions about state transitions. This type of reinforcement-learning strategy explains a variety of neuroscientific data from the hippocampus[60]. Moreover, direct evidence for the successor representation has been found in behavioral and neuroimaging work in humans[39, 50, 38, 51]. This raises an important question: to what extent do our behRSA data reflect the cached long-term predictions about future states? Of course, the direct associations between first-stage objects and their associated second-stage objects are accounted for by this type of reinforcement-learning algorithm. However, the successor representation does not predict the existence of the “indirect” relations between items that share a second-stage goal. To
see this, note that the classic successor representation only caches the future consequences of individual actions, and that items with an indirect association are never reached from each other. However, cached successor representations may still be useful for constructing cognitive maps. For example, agents may use probabilistic reverse inference [58, 59] over their successor representations to infer indirect associations between choice options, allowing them to build up sophisticated cognitive maps from cheaply formed cached transition counts. This leads to the intriguing possibility that map construction occurs by switching between two distinct modes. In the first, agents learn cached successor representations from direct experience with the world. In the second, they engage in offline simulations using reverse inference [58], to infer relationships between items that never occur together. This calls to mind the DYNA reinforcement-learning algorithm, which learns how to choose actions by integrating real and simulated experience [61]. It also makes specific predictions about which components of the cognitive map require the engagement of deliberative processing. Together, this provides a rich framework for future theoretical and behavioral work.

4.3 Cost of structure learning

Consistent with this perspective, previous work [9, 8] has shown that the learning of a task structure carries a cognitive cost. For example, participants show learning-related increases in response time and reductions in performance when a task affords structure learning. In line with this finding, we found that when participants more strongly represented goal-directed components of the task structure for regions where increased stakes were more likely. This suggests that structure learning in our two-step task also imposed a cost, since participants became more willing to encode transitions when it payed off more. This implies
that the brain engages in a cost-benefit tradeoff to decide how strongly to encode the task structure. A similar cost-benefit tradeoff has been proposed to explain how people decide which reinforcement-learning strategy to use[32]. Taken together, this indicates that there may be two, parallel, cost-benefit tradeoffs, one to determine how much to plan, and another to determine whether it is worth to learn the structures over which planning occurs. Future work may investigate what learning signals the brain uses to determine whether to engage in more extensive structure learning, and whether these two tradeoffs are truly parallel or rely on similar estimations of the value of model-based control. In other words, when model-based control is estimated as worthwhile, the brain may in turn decide that it should not only rely on goal-directed planning more, but also refine its cognitive map.

This prior work on the cost of structure learning provides some strong constraints for this tradeoff: participants are prone to learn a structure even when it does not provide them direct benefits. Indeed, we also found that participants learned components of the structure that are not beneficial for goal-directed decisions (i.e., cooccurrence-based similarity). This invites a few interpretations. Maybe some forms of structure learning occur without deliberate control. Indeed, work on statistical learning suggests that people are able to pick up on sequentially structured information, even in the absence of conscious awareness[66, 52]. Alternatively, this cost-benefit tradeoff of structure learning may consider variables beyond effort costs and task-related reward. People may simply like to learn about the world, because they value information[2, 26, 70], they may want to reduce uncertainty[13, 11], or because they feel empowered by their increased knowledge about the world[6]. A plausible hypothesis, consistent with our results, is that structure learning is more costly when higher-order relationships need to be learned[7], and therefore that cost-benefit analyses are engaged when costs are particularly high.
4.4 Benefits of behRSA approach

The goal of our study was to obtain representations of task structure using purely behavioral measures. Drawing inspiration from human fMRI techniques, we assessed pairwise similarity for objects encountered in the decision-making task. Though this technique does not directly measure neuronal firing, we suggest that its power lies in its more immediate relationship with each participant’s mental representation of the task. This makes the behRSA particularly powerful and flexible for studying the structure and format of internal representations. Regardless of the specific task design or hypotheses, in any task with a learnable structure, behRSA can be applied to test multiple dissociable hypotheses about participants’ representation of the task, and these can be explicitly investigated in relation to task performance. This approach can be useful across a number of research domains where task representation may underlie or inform other measures. With regard to future studies, we suggest that behRSA is a highly cost-effective alternative for or precursor to neuroimaging approaches.

Our behRSA results demonstrate that there exists substantial variability in the degree to which participants represent the structure of our decision-making tasks. Recent work by Feher da Silva and Hare[21] suggests similar variability in a more conventional two-stage decision-making paradigm, which is reduced when participants are guided through a more thorough explanation of the structure of their task. In other words, constraining the space of possible cognitive maps increases the engagement of successful model-based control. By providing only instructions on the rules of the paradigm but leaving exact transition structure unspecified, our behRSA allowed us to measure the individual difference in the formation of cognitive maps. We did not impose constraints on participants’ representations of the task. Rather, we simply tracked how each participant constructed their own model of task structure. This flexibility may be particularly important when linking model-based decision
making to broader aspects of cognition. In particular, learning and reasoning over relational maps may be linked to the deployment of model-based control in naturalistic scenarios in which the structure is neither known nor instructed.

In accordance with this idea, recent work by Rmus and colleagues[48] found a link between the deployment of model-based control and general mapping ability. In addition to performing a two-step task, participants also completed a graph learning task in which they viewed pairs of stimuli from a graph structure and judged the relative distances between them. Interestingly, the performance in the graph learning task positively predicted the use of model-based control in the two-step task. One benefit from our, more direct, approach is that it probes the cognitive map underlying decisions within the task itself. One promising area for research would be to investigate whether general mapping ability explains the observed differences in higher-order structure learning, which then in turn explains reliance on model-based control. In other words, we predict that structure learning in the two-step task mediates the previously observed relationship between general mapping ability and the use of model-based control.

4.5 Caveats of the approach

Following from the issue of variability, note that there was a fair amount of heteroskedasticity in whether participants generated task-related representational similarities. One of the reasons for this may be that some participants lost interest in providing ratings for pairs of objects as the task continued. Indeed, it should be noted that data collection occurred at the height of the coronavirus pandemic. It is also possible that we observed some form of retroactive interference of the behRSA task on the task structure. That is, after viewing
and rating many of the possible pairings of objects, the strength of some of the previously observed relationships may have weakened. Finally, it is possible that some participants struggled with the relatively sparse nature of the behRSA instructions. We do note, however, that the results reported here are pre-registered and replicate effects found in an initial sample. Future work will assess whether the strength of behRSA fits can be enhanced across a sample with more directed instructions.

### 4.6 Neural correlates and next steps

Though the behRSA approach is useful for behaviorally assessing participant’s models and even useful for assessing cost of structure learning, future work is needed to assess the underlying neural correlates that are allowing for map construction. In line with prior work[55, 53], we believe medial prefrontal cortex (mPFC), orbitofrontal cortex (OFC), and hippocampus are the primary regions involved in building and using representations for our task. Firstly, mPFC has been implicated in biasing hippocampal sampling in goal-directed situations[40, 54]. On the other hand, OFC has been implicated in computing task structure[55, 42, 71]. Finally, the hippocampus is thought to be involved in using and traversing cognitive maps[41, 1, 3]. Work assessing regional computations underlying our task can be complemented by work investigating the time course of representations. For instance, planning relevant and irrelevant features may come online at distinct time courses after viewing the stimuli. Tracking the onset of separable features would be possible using a time-resolved EEG approach[27]. In recent work, Kikumoto and Mayr[31] have shown decodability of different conjunctive features of a stimulus using a time-resolved RSA technique. One could extend this approach to cover different objects stored in a cognitive
map and compare it with the behavioral similarity reported by participants. Relatedly, in this paper we have mainly demonstrated that participants are constructing and using their maps, but are unable to speak to how participants are doing this “online”. Neuroimaging approaches are well suited to repeatedly probe representations of choice options. Another outstanding issue is the degree to which neural and behavioral RSA provide complimentary (or unique) sources of information. This comparison will provide increased understanding of the connection between reported mental and specific circuit representations, as well as the degree to which these representations are explicitly accessible.

4.7 Conclusion

In sum, our findings demonstrate that it is possible to behaviorally assess the features of participants’ cognitive maps. Higher-order features of these cognitive maps predict use of model-based control and performance in an established decision making task. Finally, the quality of these abstract cognitive maps is enhanced by the presence of higher reward, indicating that they can be flexibly built and used in pursuit of goals.
Appendix A

Supplement

Pilot cohort

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:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$r_{(498)}$</th>
<th>95% CI</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>learning rate: $\alpha$</td>
<td>0.61</td>
<td>[0.56, 0.67]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>inverse temperature: $\beta$</td>
<td>0.54</td>
<td>[0.47, 0.60]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>trace decay: $\lambda$</td>
<td>0.50</td>
<td>[0.43, 0.56]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>stickiness: $\pi$</td>
<td>0.70</td>
<td>[0.66, 0.74]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>response stickiness: $\rho$</td>
<td>0.24</td>
<td>[0.15, 0.32]</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>transition learning rate: $\eta$</td>
<td>0.08</td>
<td>[-0.01, 0.16]</td>
<td>= 0.085</td>
</tr>
<tr>
<td>counterfactual transitions: $\kappa$</td>
<td>0.02</td>
<td>[-0.07, 0.1]</td>
<td>= 0.724</td>
</tr>
</tbody>
</table>

Note the inability to recover the $\eta$ and $\kappa$ 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 $\eta$ and $\kappa$ were hardcoded to 1 (indicating immediate update of transition probabilities).
We then assessed the correlations between the other parameters in the model with \( \eta \) and \( \kappa \) as free parameters and the model with \( \eta \) and \( \kappa \) as hardcoded. The results reported in the main paper do not differ when analyzed with the hardcoded model (correlations among all other parameters \( r_{(498)} = 0.99, 95\% CI = [0.99, 1.0], p < 0.001 \)).

**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 (A.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 (A.1B).

**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 \((r_{(91)} = 0.04, \ 95\% \ CI = [-0.17, 0.24], \ p = 0.735)\).

**Main experiment**

**Points earned, RT, and \(w\) parameters**

In line with previous work\[33, 34\] we observed a correlation between the points earned in the task and the model-based weighting parameter \(w\) \((r_{159} = 0.6876, \ 95\% \ CI = [0.6, 0.76], \ p < 0.001)\). 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 \((r_{159} = 0.3919, \ 95\% \ CI = [0.25, 0.52], p < 0.001)\) and the model-based weighting parameter and response time \((r_{159} = 0.3186, \ 95\% \ CI = [0.17, 0.45], p < 0.001)\).

**ANOVA for 4 \(w\) parameter model**

Based on previous work exhibiting an effect of stakes on model-based control\[34\] 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
\( w_{\text{high}} = 0.56 \)) compared to low-stakes trials (\( w_{\text{low}} = 0.54 \)) \( (F(1, 160) = 7.265, p = 0.008) \). We found no effect of first-stage reward context, and no interaction effect (Supplementary Table 2: Repeated-measures ANOVA).

<p>| Table A.2: Repeated-measures ANOVA |</p>
<table>
<thead>
<tr>
<th>SS</th>
<th>Ddof1</th>
<th>Ddof2</th>
<th>MS</th>
<th>( F )</th>
<th>( p )</th>
<th>partial ( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>arm</td>
<td>0.015</td>
<td>1</td>
<td>160</td>
<td>0.015</td>
<td>1.346</td>
<td>0.248</td>
</tr>
<tr>
<td>stakes</td>
<td>0.091</td>
<td>1</td>
<td>160</td>
<td>0.091</td>
<td>7.265</td>
<td>0.008</td>
</tr>
<tr>
<td>arm * stakes</td>
<td>0.024</td>
<td>1</td>
<td>160</td>
<td>0.024</td>
<td>1.286</td>
<td>0.258</td>
</tr>
</tbody>
</table>

\( d' \) 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 \( w \) fit and \( d'_{\text{mismatch}} \) high-arm
\( (r_{(159)} = -0.01851, \ 95\% \ CI = [-0.17, 0.14], \ p = 0.8157) \). Correlation between \( w \) fit and
\( d'_{\text{mismatch}} \) low-arm \( (r_{(159)} = 0.01812, \ 95\% \ CI = [-0.14, 0.17], \ p = 0.8196) \). Correlation between \( w \) fit and \( d'_{\text{ lure}} \) high-arm \( (r_{(159)} = 0.09555, \ 95\% \ CI = [-0.06, 0.25], \ p = 0.2280) \). Correlation between \( w \) fit and \( d'_{\text{lure}} \) low-arm \( (r_{(159)} = 0.02351, \ 95\% \ CI = [-0.13, 0.18], \ p = 0.7672) \).
Figure A.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)
Figure A.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 $p < 0.05$)
References


[47] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent


[59] Alec Solway, Carlos Diuk, Natalia Córdova, Debbie Yee, Andrew G. Barto, Yael Niv, and
Matthew M. Botvinick. Optimal Behavioral Hierarchy. *PLOS Computational Biology*,
10(8):e1003779, August 2014.

[60] Kimberly L Stachenfeld, Matthew M Botvinick, and Samuel J Gershman. The hip-


[63] Rita Morais Tavares, Avi Mendelsohn, Yael Grossman, Christian Hamilton Williams,
Matthew Shapiro, Yaacov Trope, and Daniela Schiller. A Map for Social Navigation in

[64] Stephanie Theves, Guillén Fernández, and Christian F. Doeller. The Hippocampus
Maps Concept Space, Not Feature Space. *Journal of Neuroscience*, 40(38):7318–7325,
September 2020.

208, 1948.

Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities With-

3(31):1026, November 2018.

[68] Bas van Opheusden, Gianni Galbiati, Zahy Bnaya, Yunqi Li, and Wei Ji Ma. A com-
putational model for decision tree search. In *Proceedings of the 39th Annual Meeting of

[69] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy,
David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan
Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman,
Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef
Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M.
Archibald, António H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0:
272, March 2020.
