bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Similarities and differences in spatial and non-spatial cognitive maps

Charley M. Wu[1,2*], Eric Schulz[3], Mona M. Garvert[4,5,6] Björn Meder[2,7,8], Nicolas W. Schuck[5]

1 Department of Psychology, Harvard University, Cambridge, MA, USA
2 Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany
3 Max Planck Research Group Computational Principles of Intelligence, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
4 Department of Psychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
5 Max Planck Research Group NeuroCode, Max Planck Institute for Human Development, Berlin, Germany
6 Wellcome Centre for Integrative Neuroimaging, University of Oxford, John Radcliffe Hospital, Oxford, UK
7 Max Planck Research Group iSearch, Max Planck Institute for Human Development, Berlin, Germany
8 Department of Psychology, University of Erfurt, Erfurt, Germany

* charleywu@fas.harvard.edu

Abstract

There is a resurgence of interest in "cognitive maps" based on recent evidence that the hippocampal-entorhinal system encodes both spatial and non-spatial information, with far-reaching implications for human behavior. Yet little is known about the commonalities and differences in the computational principles underlying human learning and decision making in spatial and non-spatial domains. We use a within-subject design to examine how humans search for either spatially or conceptually correlated rewards. Using a Bayesian learning model, we find evidence for the same computational mechanisms of generalization across domains. While participants were sensitive to expected rewards and uncertainty in both tasks, how they leveraged this knowledge to guide exploration was different: participants displayed less uncertainty-directed and more random exploration in the conceptual domain. Moreover, experience with the spatial task improved conceptual performance, but not vice versa. These results provide important insights about the degree of overlap between spatial and conceptual cognition.

Introduction

Thinking spatially is intuitive. We remember things in terms of places [1–3], describe the world using spatial metaphors [4, 5], and commonly use concepts like "space" or "distance" in mathematical descriptions of abstract phenomena.

In line with these observations, previous theories have argued that reasoning about abstract conceptual information follows the same computational principles as spatial

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reasoning [6–8]. This has recently gained new support from neuroscientific evidence suggesting that common neural substrates are the basis for knowledge representation across domains [9–13].

One important implication of these accounts is that reinforcement learning [14] in non-spatial domains may rely on a map-like organization of information, supported by the computation of distances or similarities between experiences. Here, we ask to what extent does the search for rewards depend on the same distance-dependent generalization across domains? We formalize a computational model that incorporates distance-dependent generalization and test it in a within-subject experiment, where either spatial features or abstract conceptual features are predictive of rewards. This allows us to study learning, decision making, and exploration in spatial versus conceptual domains, in order to gain insights into the organizational structure of cognitive representations in both domains.

Whereas early psychological theories described reinforcement learning as merely developing an association between stimuli, responses and rewards [15–17], more recent studies have recognized that the structure of representations plays an important role in making value-based decisions [11, 18] and is particularly important for knowing how to generalize from limited data to novel situations [19, 20]. This idea dates back to Tolman, who famously argued that both rats and humans extract a "cognitive map" of the environment [21]. This cognitive map encodes relationships between experiences or options, such as the distances between locations in space [22], and—crucially—facilitates flexible planning and generalization. While cognitive maps were first identified as representations of physical spaces, Tolman hypothesized that similar principles may underlie the organization of knowledge in broader and more complex cognitive domains [21].

As was the case with Tolman, neuroscientific evidence for a cognitive map was initially found in the spatial domain, in particular, with the discovery of spatially selective place cells in the hippocampus [23, 24] and entorhinal grid cells that fire along a spatial hexagonal lattice [25]. Together with a variety of other specialized cell types that encode spatial orientation [26, 27], boundaries [28, 29], and distances to objects [30], this hippocampal-entorhinal machinery is often considered to provide a cognitive map facilitating navigation and self-location. Yet more recent evidence has shown that the same neural mechanisms are also active when reasoning about more abstract, conceptual relationships [31–36], characterized by arbitrary feature dimensions [37] or temporal relationships [38, 39]. For example, using a technique developed to detect spatial hexagonal grid-like codes in fMRI signals [40], Constantinescu et al. found that human participants displayed a pattern of activity in the entorhinal cortex consistent with mental travel through a 2D coordinate system defined by the length of a bird's legs and neck [9]. Similarly, the same entorhinal-hippocampal system has also been found to reflect the graph structure underlying sequences of stimuli [10] or the structure of social networks [41], and even to replay non-spatial representations in the sequential order that characterized a previous decision-making task [42]. At the same time, much evidence indicates that cognitive map-related representations are not limited to medial temporal areas, but also include ventral and orbital medial prefrontal areas [9, 11, 40, 43–45].

Based on these findings, we asked whether learning and searching for rewards in spatial and conceptual domains is governed by similar computational principles. Using a within-subject design comparing spatial and non-spatial reward learning, we tested whether participants used perceptual similarities in the same way as spatial distances to generalize from previous experiences and inform the exploration of novel options. In both domains, rewards were correlated (see Fig. S2), such that nearby or similar options tended to yield similar rewards. To model how participants generalize and explore using either perceptual similarities or spatial distances, we used Gaussian Process (GP)

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regression [46, 47] as a Bayesian model of generalization. The Bayesian predictions of the GP model generalize about novel options using a common notion of similarity across domains, and provide estimates of expected reward and uncertainty. We tested out-of-sample predictions of the GP model against a Bayesian learner that incorporates uncertainty-guided exploration but without generalization, and investigated differences in parameters governing value-based decision making and uncertainty-directed exploration [48–50].

Participant performance was correlated across tasks and was best captured by the GP model in both domains. We were also able to reliably predict participant judgments about unobserved options based on parameters estimated from the bandit task. Whereas the model parameters indicated similar levels of generalization in both domains, we found lower levels of directed exploration in the conceptual domain, where participants instead showed increased levels of random exploration. Moreover, we also observed an asymmetric task order effect, where performing the spatial task first boosted performance on the conceptual task but not vice versa. These findings provide a clearer picture of both the commonalities and differences in how people reason about and represent both spatial and abstract phenomena in complex reinforcement learning tasks.

Results

129 participants searched for rewards in two successive multi-armed bandit tasks (Fig 1). The *spatial* task was represented as an 8×8 grid, where participants used the arrow keys to move a highlighted square to one of the 64 locations, with each location representing one option (i.e., arm of the bandit). The *conceptual* task was represented using Gabor patches, where a single patch was displayed on the screen and the arrow keys changed the tilt and stripe frequency (each having 8 discrete values; see Fig. S1), providing a non-spatial domain where similarities are relatively well defined. Each of the 64 options in both tasks produced normally distributed rewards, where the means of each option were correlated, such that similar locations or Gabor patches with similar stripes and tilts yielded similar rewards (Fig. S2), thus providing traction for similarity-guided generalization and search. The strength of reward correlations were manipulated between subjects, with one half assigned to *smooth* environments (with higher reward correlations) and the other assigned to *rough* environments (with lower reward correlations), although both classes of environments had the same expectation of rewards across options.

The spatial and conceptual tasks were performed in counter-balanced order, with each task consisting of an initial training phase (see Methods; Fig 1c) and then 10 rounds of bandits. Each round had a different reward distribution (drawn without replacement from the assigned class of environments), and participants were given 20 choices to acquire as many points as possible (later converted to monetary rewards). The search horizon was much smaller than the total number of options and therefore induced an explore-exploit dilemma and motivated the need for generalization and efficient exploration. The last round of each task was a "bonus round", where after 15 choices, participants were shown 10 unobserved options (selected at random) and asked to make judgments about the expected reward and their level of confidence (i.e., uncertainty about the expected rewards). These judgments were used to validate the internal belief representations of our models. All data and code, including interactive notebooks containing all analyses in the paper, is publicly available at `https://github.com/charleywu/cognitivemaps`.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

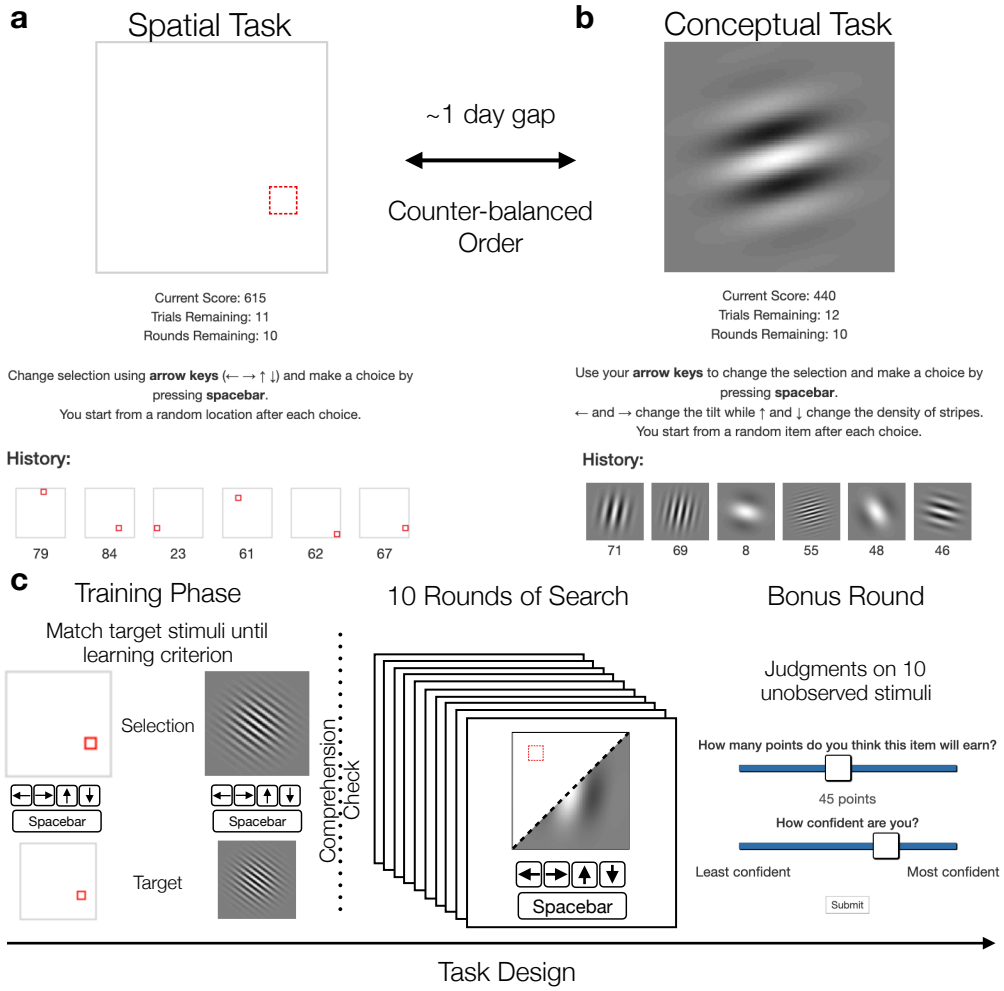

Figure 1. Experiment design. **a**) In the spatial task, options were defined as a highlighted square in a 8×8 grid, where the arrow keys were used to move the highlighted location. **b**) In the conceptual task, each option was represented as a Gabor patch, where the arrow keys changed the tilt and the number of stripes (Fig S1). Both tasks corresponded to correlated reward distributions, where choices in similar locations or having similar Gabor features predicted similar rewards (Fig S2). **c**) The same design was used in both tasks. Participants first completed a training phase where they were asked to match a series of target stimuli. This used the same inputs and stimuli as the main task, where the arrow keys modified either the spatial or conceptual features, and the spacebar was used to make a selection. After reaching the learning criterion of at least 32 training trials and a run of 9 out of 10 correct, participants were shown instructions for the main task and asked to complete a comprehension check. The main task was 10 rounds long, where participants were given 20 selections in each round to maximize their cumulative reward (shown in panels a and b). The 10th round was a "bonus round" where after 15 selections participants were asked to make 10 judgments about the expected reward and associated uncertainty for unobserved stimuli from that round. After judgments were made, participants selected one of the options, observed the reward, and continued the round as usual.

Computational Models of Learning, Generalization, and Search

Multi-armed bandit problems [51, 52] are a prominent framework for studying learning, where various reinforcement learning (RL) models [14] are used to model the learning of

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reward valuations and to predict behavior. A common element of most RL models is some form of prediction-error learning [53, 54], where model predictions are updated based on the difference between the predicted and experienced outcome. One classic example of learning from prediction errors is the Rescorla-Wagner [54] model, in which the expected reward $V(\cdot)$ of each bandit is described as a linear combination of weights $\mathbf{w}_t$ and a one-hot stimuli vector $\mathbf{x}_t$ representing the current state s_t:

$$V(\mathbf{x}_t) = \mathbf{w}_t^\top \mathbf{x}_t \tag{1}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta \delta_t \mathbf{x}_t \tag{2}$$

Learning occurs by updating the weights $\mathbf{w}$ as a function of the prediction error $\delta_t = r_t - V(\mathbf{x}_t)$, where r_t is the observed reward, $V(\mathbf{x}_t)$ is the reward expectation, and $0 < \eta \leq 1$ is the learning rate parameter. In our task, we used a *Bayesian Mean Tracker* (BMT) as a Bayesian variant of the Rescorla-Wagner model [54, 55]. Rather than making point estimates of reward, the BMT makes independent and normally distributed predictions $V(s_{i,t}) \sim \mathcal{N}(m_{i,t}, v_{i,t})$ for each state $s_{i,t}$, which are characterized by a mean m and variance v and updated on each trial t via the delta rule (see Methods for details).

Generalization using Gaussian process regression

Yet, an essential aspect of human cognition is the ability to generalize from limited experiences to novel options. Rather than learning independent reward representations for each state, we adopt a function learning approach to generalization [19, 56], where continuous functions represent candidate hypotheses about the world, mapping the space of possible options to some outcome value. For example, a function can map how pressure on the gas pedal is related to the acceleration of a car, or how different amounts of water and fertilizer influence the growth rate of a plant. Crucially, the learned mapping provides estimates even for outcomes that have not been observed, by interpolating or extrapolating from previous experiences.

While the literature on how humans explicitly learn functions extends back to the 1960s [57], more recent approaches have proposed Gaussian Process (GP) regression [46] as a candidate model of human function learning [58–60]. GPs unite previous proposals of rule-based [61, i.e., learning the weights of a particular parametric function] and exemplar-based theories [62, i.e., neural networks predicting similar inputs will produce similar outputs], while also predicting the perceived difficulty of learning different functions [63] and explaining biases in how people extrapolate from limited data [58].

Formally, a GP defines a multivariate-normal distribution $P(f)$ over possible value functions $f(s)$ that map inputs s to output $y = f(s)$.

$$P(f) \sim \mathcal{GP}\left(m(s), k(s, s')\right) \tag{3}$$

The GP is fully defined by the mean function $m(s)$, which is frequently set to 0 for convenience without loss of generality [46], and kernel function $k(s, s')$ encoding prior assumptions (or inductive biases) about the underlying function. Here we use the *radial basis function* (RBF) kernel:

$$k(s, s') = \exp\left(-\frac{||s - s'||^2}{2\lambda^2}\right) \tag{4}$$

encoding similarity as a smoothly decaying function of the squared Euclidean distance between stimuli s and s', measured either in spatial or conceptual distance. The

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

length-scale parameter λ encodes the rate of decay, where larger values correspond to broader generalization over larger distances.

Given a set of observations $\mathcal{D}_t = [\mathbf{s}_t, \mathbf{y}_t]$ about previously observed states and associated rewards, the GP makes normally distributed posterior predictions for any novel stimuli $s^\star$, defined in terms of a posterior mean and variance:

$$m(s^\star|\mathcal{D}_t) = K(s^\star, \mathbf{s}_t)\left[K(\mathbf{s}_t, \mathbf{s}_t) + \sigma_\epsilon^2 \mathbf{I}\right]^{-1} \mathbf{y}_t \tag{5}$$

$$v(s^\star|\mathcal{D}_t) = k(s^\star, s^\star) - K(s^\star, \mathbf{s}_t)\left[K(\mathbf{s}_t, \mathbf{s}_t) + \sigma_\epsilon^2 \mathbf{I}\right]^{-1} K(\mathbf{s}_t, s^\star) \tag{6}$$

The posterior mean corresponds to the expected value of $s^\star$ while the posterior variance captures the underlying uncertainty in the prediction. Note that the posterior mean can also be rewritten as a similarity-weighted sum:

$$m(s^\star|\mathcal{D}_t) = \sum_{i=1}^{t} w_i k(s^\star, s_i) \tag{7}$$

where each s_i is a previously observed input in $\mathbf{s}_t$ and the weights are collected in the vector $\mathbf{w} = \left[K(\mathbf{s}_t, \mathbf{s}_t) + \sigma_\epsilon^2 \mathbf{I}\right]^{-1} \mathbf{y}_t$. Intuitively, this means that GP regression is equivalent to a linearly weighted sum, but uses basis functions $k(\cdot, \cdot)$ that project the inputs into a feature space, instead of the discrete state vectors. To generate new predictions, every observed reward y_i in $\mathbf{y}_t$ is weighted by the similarity of the associated state s_i to the candidate state $s^\star$ based on the kernel similarity. This similarity-weighted sum (Eq 7) is equivalent to a RBF network [64], which has featured prominently in machine learning approaches to value function approximation [14] and as a theory of the neural architecture of human generalization [65] in vision and motor control.

Uncertainty-directed exploration

In order to transform the Bayesian reward predictions of the BMT and GP models into predictions about participant choices, we use upper confidence bound (UCB) sampling together with a softmax choice rule as a combined model of both *directed* and *random* exploration [19, 49, 50].

UCB sampling uses a simple weighted sum of expected reward and uncertainty:

$$q_{UCB}(s) = m(s) + \beta\sqrt{v(s)} \tag{8}$$

to compute a value q for each option s, where the exploration bonus β determines how to trade off exploring highly uncertain options against exploiting high expected rewards. This simple heuristic—although myopic—produces highly efficient learning by preferentially guiding exploration towards uncertain yet promising options, making it one of the only algorithms with known performance bounds in Bayesian optimization [66]. Recent studies have provided converging evidence for directed exploration in human behavior across a number of domains [19, 49, 67–69].

The UCB values are then put into a softmax choice rule:

$$P(s_i) = \frac{\exp\left(q(s_i)/\tau\right)}{\sum_j \exp\left(q(s_j)/\tau\right)} \tag{9}$$

where the temperature parameter τ controls the amount of random exploration. Higher temperature sampling leads to more random choice predictions, with $\tau \to \infty$ converging on uniform sampling. Lower temperature values make more precise predictions, where $\tau \to 0$ converges on an $\arg\max$ choice rule. Taken together, the exploration bonus β and temperature τ parameters estimated on participant data allow us to assess the relative contributions of directed and undirected exploration, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Behavioral Results

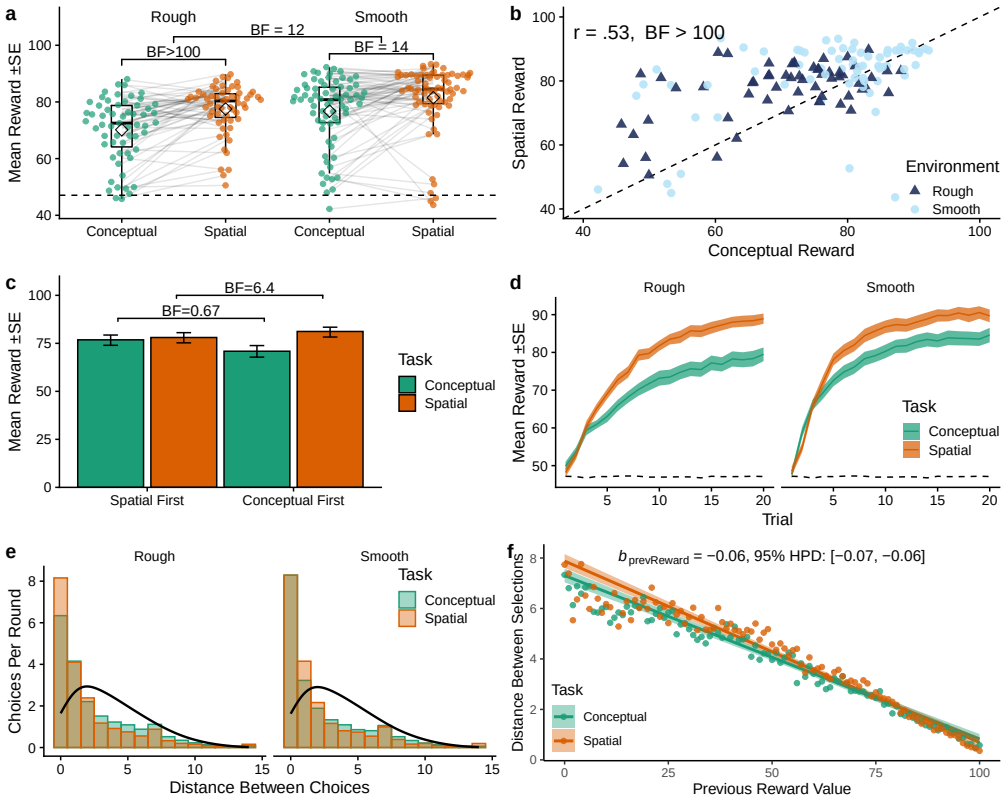

Figure 2. Behavioral results. **a**) Mean reward in each task, where each dot is a participant and lines connect the same participant across tasks. Tukey boxplots show median (horizontal line) and 1.5x IQR, while diamonds indicate the group mean. The dashed line indicates chance performance. Bayes Factors (BF) indicate the evidence against a specified null hypothesis for either two sample (rough vs. smooth) or paired (conceptual vs. spatial) t-tests (see Methods). **b**) Correspondence between tasks, where each dot represents the average reward of a single participant and the dotted line indicates $y = x$. **c**) Task order effect, where experience with spatial search boosted performance on conceptual search, but not vice versa. Bayes factors correspond to paired t-tests. **d**) Average learning curves over trials, showing the mean (line) and standard error (ribbon) aggregated across rounds and participants. The dashed line indicates chance performance. **e**) The Manhattan distance between selections compared to a random baseline (black line). **f**) Distance between selections as a function of the previous observed reward value, showing the means (points) and the group-level predictions of a mixed-effects regression (Table S1), where the ribbons indicate the 95% CI.

We first analyzed participant performance in the bandit tasks before turning to model-based analyses. Participants achieved much higher rewards than chance in both conceptual (one-sample t-test: $t(128) = 24.6$, $p < .001$, $d = 2.2$, $BF > 100$) and spatial tasks ($t(128) = 34.6$, $p < .001$, $d = 3.0$, $BF > 100$; Fig. 2a)[1]. Using a two-way mixed ANOVA, we found that both environment (smooth vs. rough: $F(1, 127) = 9.4$, $p = .003$, $\eta^2 = .05$, $BF = 13$) and task (spatial vs. conceptual: $F(1, 127) = 35.8$, $p < .001$, $\eta^2 = .06$, $BF > 100$) influenced performance. The stronger reward correlations present

[1]Bayes Factors (BF) accompany each frequentist test to indicate the evidence against a specified null hypothesis. See Methods for further details.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in smooth environments facilitated higher performance (two sample t-test: $t(127) = 3.1$, $p = .003$, $d = 0.5$, $BF = 12$), even though both environments had the same expected reward.

While performance was strongly correlated between the spatial and conceptual tasks (Pearson's $r = .53$, $p < .001$, $BF > 100$; Fig. 2b), participants performed systematically better in the spatial version (paired t-test: $t(128) = 6.0$, $p < .001$, $d = 0.5$, $BF > 100$). This difference in task performance can largely be explained by a one-directional transfer effect (Fig. 2c). Participants performed better on the conceptual task after having experienced the spatial task ($t(127) = 2.8$, $p = .006$, $d = 0.5$, $BF = 6.4$). This was not the case for the spatial task, where performance did not differ whether performed first or second ($t(127) = -1.7$, $p = .096$, $d = 0.3$, $BF = .67$). Thus, experience with spatial search boosted performance on conceptual search, but not vice versa.

Participants learned effectively within each round and obtained higher rewards with each successive choice (Pearson correlation between reward and trial: $r = .88$, $p < .001$, $BF > 100$; Fig 2d). We also found evidence for learning across rounds in the spatial task (Pearson correlation between reward and round: $r = .91$, $p < .001$, $BF = 15$), but not in the conceptual task ($r = .58$, $p = .104$, $BF = 1.5$).

Patterns of search also differed across domains. Comparing the average Manhattan distance between consecutive choices in a two-way mixed ANOVA showed an influence of task (within: $F(1, 127) = 13.8$, $p < .001$, $\eta^2 = .02$, $BF = 67$) but not environment (between: $F(1, 127) = 0.12$, $p = .73$, $\eta^2 = .001$, $BF = 0.25$, Fig. 2e). This reflected that participants searched in smaller step sizes in the spatial task ($t(128) = -3.7$, $p < .001$, $d = 0.3$, $BF = 59$), corresponding to a more local search strategy, but did not adapt their search distance to the environment. Note that each trial began with a randomly sampled initial stimuli, such that participants did not begin near the previous selection (see Methods). The bias towards local search (one-sample t-test comparing search distance against chance: $t(128) = -16.3$, $p < .001$, $d = 1.4$, $BF > 100$) is therefore not a side effect of the task characteristics, but both purposeful and effortful (see Fig S4 for additional analysis of search trajectories).

Participants also adapted their search patterns based on reward values (Fig. 2f), where lower rewards predicted a larger search distance on the next trial (correlation between previous reward and search distance: $r = -.66$, $p < .001$, $BF > 100$). We analyzed this relationship using a Bayesian mixed-effects regression, where we found previous reward value to be a reliable predictor of search distance ($b_{\text{prevReward}} = -0.06$, 95% HPD: $[-0.07, -0.06]$; see Table S1), while treating participants as random effects. This provides initial evidence for generalization-like behavior, where participants actively avoided areas with poor rewards and stayed near areas with rich rewards.

In summary, we find correlated performance across tasks, but also differences in both performance and patterns of search. Participants were boosted by a one-directional transfer effect, where experience with the spatial task improved performance on the conceptual task, but not the other way around. In addition, participants made larger jumps between choices in the conceptual task and searched more locally in the spatial task. However, participants adapted these patterns in both domains in response to reward values, where lower rewards predicted a larger jump to the next choice.

Modeling Results

To better understand how participants navigated the spatial and conceptual tasks, we used computational models to predict participant choices and judgments. Both GP and BMT models implement directed and undirected exploration using the UCB exploration bonus β and softmax temperature τ as free parameters. The models differed in terms of learning, where the GP generalized about novel options using the length-scale parameter

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

λ to modulate the extent of generalization over spatial or conceptual distances, while the BMT learns the rewards of each option independently (see Methods).

Both models were estimated using leave-one-round-out cross validation, where we compare goodness of fit using out-of-sample prediction accuracy, described using a pseudo-R^2 (Fig 3a). The differences between models were reliable and meaningful, with the GP model making better predictions than the BMT in both the conceptual ($t(128) = 3.9$, $p < .001$, $d = 0.06$, $BF > 100$) and spatial tasks ($t(128) = 4.3$, $p < .001$, $d = 0.1$, $BF > 100$). In total, the GP model best predicted 85 participants in the conceptual task and 93 participants in the spatial task (out of 129 in total). A Bayesian model selection framework [70, 71] confirmed that the GP had the highest posterior probability (corrected for chance) of being the best model in both tasks (protected exceedance probability; conceptual: $pxp(GP) = .997$; spatial: $pxp(GP, spatial) = 1.000$; Fig 3b). superiority of the GP model suggests that generalization about novel options via the use of structural information played a guiding role in how participants searched for rewards (see Fig S6 for additional analyses).

Learning Curves

To confirm that the GP model indeed captured learning behavior better in both tasks, we simulated learning curves from each model using participant parameter estimates (Fig. 3c; see Methods). The GP model achieved human-like performance in all tasks and environments (comparing aggregate GP and human learning curves: conceptual MSE=17.7; spatial MSE=16.6), whereas BMT learning curves were substantially less similar (conceptual MSE=150.6; spatial MSE=330.7).

Parameter Estimates

To understand how generalization and exploration differed between domains, Fig. 3d compares the estimated model parameters from the conceptual and spatial tasks. The GP model had three free parameters: the extent of generalization (λ) of the RBF kernel, the exploration bonus (β) of UCB sampling, and the temperature (τ) of the softmax choice rule (see Fig. S9 for BMT parameters). Note that the exploration bonus captures exploration *directed* towards uncertainty, whereas temperature captures random, *undirected* exploration, which have been shown to be distinct and recoverable parameters [19, 69].

We do not find reliable differences in λ estimates across tasks (Wilcoxon signed-rank test: $Z = -1.2$, $p = .115$, $r = -.11$, $BF = .13$), although the removal of outliers revealed a pattern of narrower generalization in the conceptual domain (paired t-test with outliers removed: $t(104) = -3.8$, $p < .001$, $d = 0.4$, $BF = 75$, see Methods). In all cases, we observed lower levels of generalization relative to the true generative model of the underlying reward distributions ($\lambda_{\text{rough}} = 2$, $\lambda_{\text{smooth}} = 4$; min-$BF = 1456$), replicating previous findings [19] that found undergeneralization to be largely beneficial in similar settings. Generalization was also correlated across tasks (Kendall rank correlation: $r_\tau = .13$, $p = .028$, $BF = 1.3$; Pearson correlation with outliers removed: $r = .30$, $p = .002$, $BF = 22$), suggesting participants tended to generalize similarly across domains.

Whereas generalization was similar between tasks, there were intriguing differences in exploration. We found substantially lower exploration bonuses (β) in the conceptual task ($Z = -5.0$, $p < .001$, $r = -.44$, $BF > 100$; outliers removed: $t(116) = -6.6$, $p < .001$, $d = 0.8$, $BF > 100$), indicating a large reduction of directed exploration, relative to the spatial task. At the same time, there was an increase in temperature (τ) in the conceptual task ($Z = 6.9$, $p < .001$, $r = -.61$, $BF > 100$; outliers removed: $t(105) = 6.9$, $p < .001$, $d = 0.8$, $BF > 100$), corresponding to an increase in random,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

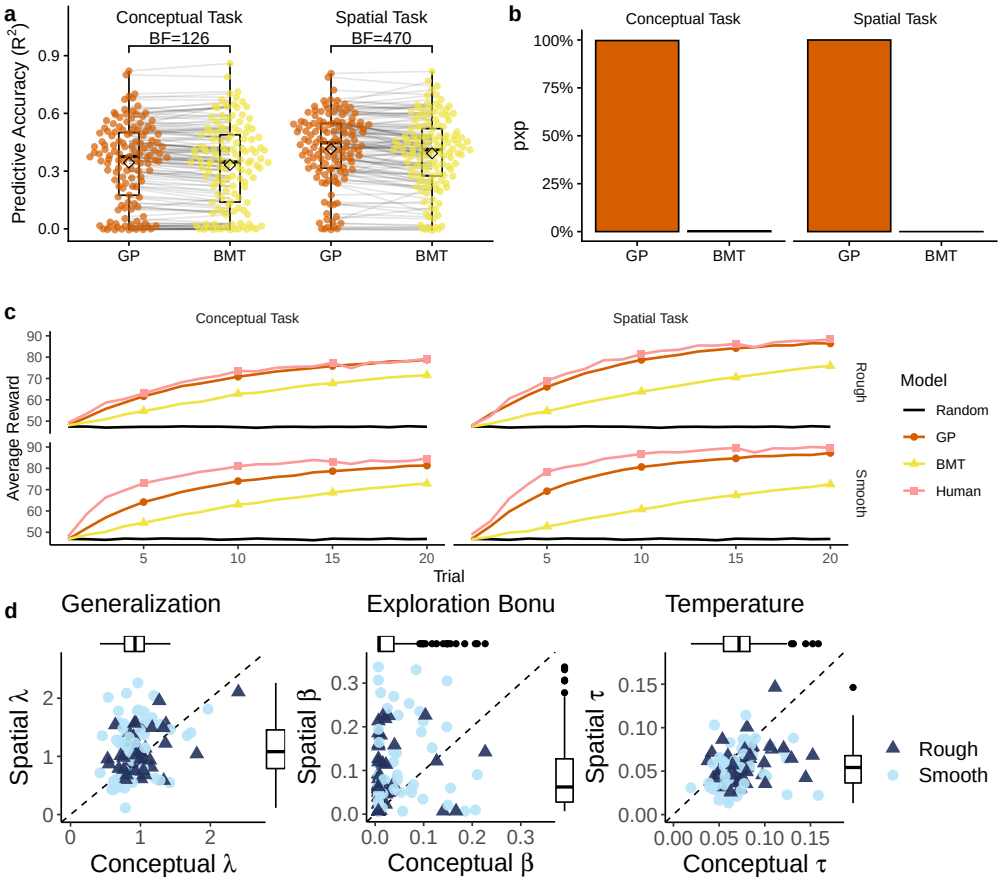

Figure 3. Modeling results. **a**) Predictive accuracy of each model, where 1 is a perfect model and 0 is equivalent to chance. Each dot is a single participant, with lines indicating the difference between models. Tukey boxplot shows the median (line) and 1.5 IQR, with the group mean indicated as a diamond. **b**) Protected Exceedence Probability (pxp), which provides a hierarchical estimate of model prevalence in the population (corrected for chance). **c**) Simulated learning curves. Each line is the averaged performance over 10,000 replications, where we sampled participant parameter estimates and simulated behavior on the task. The pink line is the group mean of our human participants, while the black line provides a random baseline. **d**) GP parameter estimates from the conceptual (x-axis) and spatial (y-axis) tasks. Each point is the mean estimate for a single participant and the dotted line indicates $y = x$. Outliers (using the larger Tukey criteria for the two tasks) are excluded from the plot but not from the rank correlations.

undirected exploration. Despite these differences, we find some evidence of correlations across tasks for directed exploration ($r_\tau = .18$, $p = .002$, $BF = 13$; outliers removed: $r = .15$, $p = .109$, $BF = .73$) and substantial evidence for correlations between random exploration across domains ($r_\tau = .43$, $p < .001$, $BF > 100$; outliers removed: $r = .30$, $p = .002$, $BF = 23$).

Thus, participants displayed correlated and similar levels of generalization in both tasks, but with markedly different patterns of exploration. Whereas participants engaged in typical levels of directed exploration in the spatial domain (replicating previous studies [19, 69]), they displayed reduced levels of directed exploration in the conceptual task, substituting instead an increase in undirected exploration. Again, this is not due to a lack of effort, because participants made substantially longer search

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

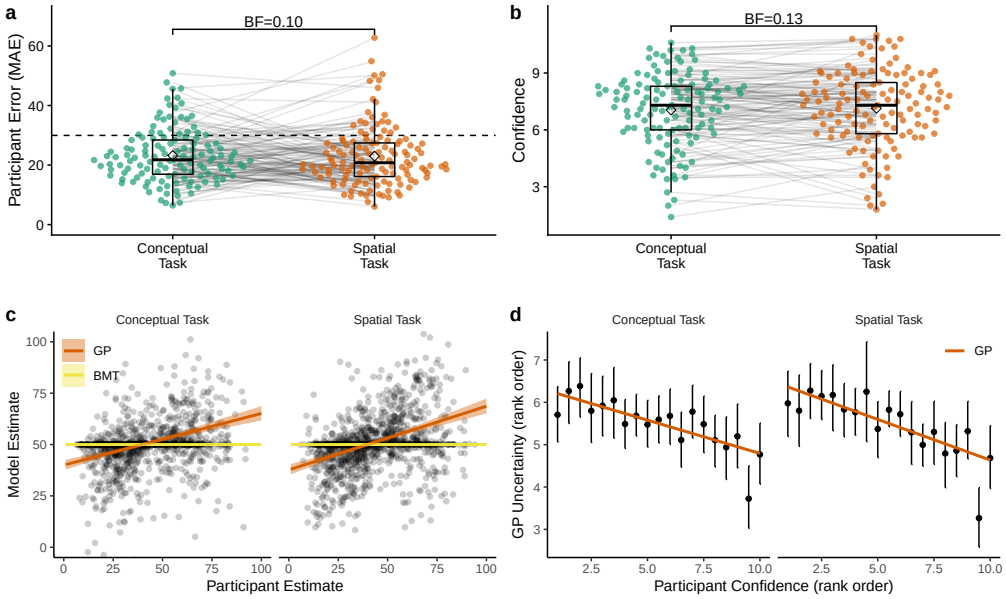

Figure 4. Bonus Round. **a**) Mean absolute error (MAE) of judgments in the bonus round, where each dot is a single participant and lines connect performance across tasks. Tukey boxplot show median and 1.5× IQR, with the diamonds indicating group mean and the dashed line providing a comparison to chance. Bayes factor indicates the evidence against the null hypothesis for a paired t-test. **b**) Average confidence ratings (Likert scale: [0,10]). **c**) Comparison between participant judgments and model predictions (based on the parameters estimated from the search task). Each point is a single participant judgment, with color lines representing the predicted group-level effect of a mixed effect regression (Table S2 and ribbons show the 95% CI (undefined for the BMT model, which makes identical predictions for all unobserved options). **d**) Correspondence between participant confidence ratings and GP uncertainty, where both are rank-ordered at the individual level. Black dots show aggregate means and 95% CI, while the colored line is a linear regression.

trajectories in the conceptual domain (see Fig S4a). Rather, this indicates a fundamental difference in how people represent or reason about spatial and conceptual domains in order to decide which are the most promising options to explore.

Bonus Round

In order to further validate our behavioral and modeling results, we analyzed participants' judgments of expected rewards and perceived confidence for 10 unobserved options they were shown during the final "bonus" round of each task (see Methods and Fig. 1c). Participants made equally accurate judgments in both tasks (comparing mean absolute error to the ground truth: $t(128) = -0.2$, $p = .827$, $d = 0.02$, $BF = .10$; Fig. 4a), which were far better than chance (conceptual: $t(128) = -9.2$, $p < .001$, $d = 0.8$, $BF > 100$; spatial: $t(128) = -8.4$, $p < .001$, $d = 0.7$, $BF > 100$) and correlated between tasks ($r = .27$, $p = .002$, $BF = 20$). Judgment errors were also correlated with performance in the bandit task ($r = -.45$, $p < .001$, $BF > 100$), such that participants who earned higher rewards also made more accurate judgments.

Participants were equally confident in both domains ($t(128) = -0.8$, $p = .452$, $d = 0.04$, $BF = .13$; Fig. 4b), with correlated confidence across tasks ($r = .79$, $p < .001$, $BF > 100$) suggesting some participants were consistently more confident than others. Ironically, more confident participants also had larger judgment errors ($r = .31$, $p < .001$,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$BF = 91$) and performed worse in the bandit task ($r = -.28$, $p = .001$, $BF = 28$).

Using parameters estimates from the search task (excluding the entire bonus round), we computed model predictions for each of the bonus round judgments as an out-of-task prediction analysis. Whereas the BMT invariably made the same predictions for all unobserved options since it does not generalize (Fig. 4c), the GP predictions were correlated with participant judgments in both conceptual (mean individual correlation: $\hat{r} = .35$; single sample t-test of z-transformed correlation coefficients against $\mu = 0$: $t(128) = 11.0$, $p < .001$, $d = 1.0$, $BF > 100$) and spatial tasks ($\hat{r} = .43$; $t(128) = 11.0$, $p < .001$, $d = 1.0$, $BF > 100$). This correspondence between human judgments and model predictions was also confirmed using a Bayesian mixed effects model, where we again treated participants as random effects ($b_{\text{participantJudgment}} = .25$, 95% HPD: [0.20, 0.31]; see Table S2 for details).

Not only was the GP able to predict judgments about expected reward, but it also captured confidence ratings. Fig 4d shows how the highest confidence ratings corresponded to the lowest uncertainty estimates made by the GP model. This effect was also found in the raw data, where we again used a Bayesian mixed effects model to regress confidence judgments onto the GP uncertainty predictions ($b_{\text{participantJudgment}} = -0.02$, 95% HPD: [-0.03, -0.01]; see Table S2).

Thus, participant search behavior was consistent with our GP model and we were also able to make accurate out-of-task predictions about both expected reward and confidence judgments using parameters estimated from the search task. These predictions validate the internal learning model of the GP, since reward predictions depend only on the generalization parameter λ. All together, our results suggest domain differences were not due to differences in how participants computed or represented expected reward and uncertainty, since they were equally good judging their uncertainty in the bonus rounds for both domains. Rather, these diverging patterns of search arose from differences in exploration, where participants substantially reduced their level of exploration directed towards uncertain options in the conceptual domain.

Discussion

Previous theories of cognitive maps [21, 32–34] have argued that reasoning in abstract domains follows similar computational principles as in spatial domains, for instance, sharing a common approach to computing similarities between experiences. These accounts imply that the shared notion of similarity should influence how people generalize from past outcomes, and also how they balance between sampling new and informative options as opposed to known options with high expected rewards.

Here, we investigated to what extent learning and searching for rewards are governed by similar computational principles in spatial and conceptual domains. Using a within-subject design, we studied participant behavior in both spatially and conceptually correlated reward environments. Comparing different computational models of learning and exploration, we found that a Gaussian Process model that incorporated distance-based generalization, and hence a cognitive map of similarities, best predicted participants behavior in both domains. This model also generated human-like learning curves and made accurate out-of-task predictions about participant reward estimations and confidence ratings in a final bonus round. The model evidence for distance-based decision making in non spatial domains was in line with our behavioral results. Performance was correlated across domains and benefited from higher outcome correlations between similar bandit options. Subsequent choices tended to be more local than expected by chance, and similar options where more likely to be chosen after a high reward than a low reward outcome.

In addition to revealing similarities, our modelling and behavioral analyses provided

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a diagnostic lens into differences between spatial and conceptual domains. Whereas we found similar levels of generalization in both tasks, patterns of exploration were substantially different. Although participants showed clear signs of directed exploration (i.e., seeking out more uncertain options) in the spatial domain, this was notably reduced in the conceptual task. However, as if in compensation, participants increased their random exploration in the conceptual task. This implies a reliable shift in sampling strategies but not in generalization. Thus, even though the computational principles underpinning reasoning in both domains are indeed similar, how these computations are mapped onto actions can vary substantially. Moreover, participants obtained more rewards and sampled more locally in the spatial domain. We also find a one-directional transfer effect, where experience with the spatial task boosted performance on the conceptual task, but not vice versa. These findings shed new light onto the computational mechanisms of generalization and decision making, suggesting a universality of generalization and a situation-specific adaptation of decision making policies.

Several questions about the link between cognitive maps across domains remain unanswered by our current study and are open for future investigations. One question is whether participants adapted their exploration strategies due to domain-specific cognitive or representational differences, or rather, if the conceptual domain was simply harder, causing participants to revert to a less taxing form of random exploration. Our pre-task training phase (Fig. 1c) giving participants familiarity with the spatial and conceptual stimuli certainly reduced this effect, but some performance differences remained. In addition, evidence that domain differences are not a mere consequence of differences in difficulty comes from participant performance in the bonus round. We found no differences in their predictions and uncertainty estimates about unseen options. This means that participants generalized and managed to track the uncertainties of unobserved options similarly in both domains. However, they did not or could not leverage their representations of uncertainty for performing directed exploration as effectively in the conceptual task. Similar changes in exploration strategies without a change in generalization have also been observed in risky search domains, where participants flexibly shifted from eagerly seeking out uncertainty in a positive reward condition to actively avoiding uncertainty in conditions where negative outcomes had to be avoided [72].

Exploration differences could also have been influenced by task constraints which were different in the two tasks. In the conceptual task only a single stimulus out of the 64 available options was displayed, while the entire set of available options was displayed in the spatial task. This may have made the conceptual task more difficult and played a role in inducing a shift to a simpler, random exploration strategy. Previous work used a task where both spatial and conceptual features were simultaneously presented [73, i.e., conceptual stimuli were shuffled and arranged on a grid], yet only spatial or only conceptual features predicted rewards. However, differences in the saliency of spatial and conceptual features meant participants were highly influenced by spatial features, even when they were irrelevant. This present study was designed to overcome these issues by presenting only task-specific features. Spatial relationships may also be easier to learn in principle, due to the extended nature of our awareness of the world around us, whereas conceptual stimuli are more commonly experienced one at a time, such as a single idea in a sequential train of thought. Currently, our model can capture but not fully explain these differences in search behavior, since it treats both domains as equivalent generalization and exploration problems.

Our model also does not account for attentional mechanisms [74] or working memory constraints [75, 76], which may play a crucial role in influencing how people integrate information differently across domains. Indeed, the “stretchy birds” paradigm used

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by [9] as evidence for a common neural representation of spatial and conceptual knowledge required several hours of training before being measured in the scanner. Similar shifts from directed to random exploration have also been observed under direct cognitive load manipulations, such as by adding working memory load [77] or by limiting the available decision time [78].

Another interesting finding is the one directional transfer from the spatial to the conceptual domain but not vice versa. This finding supports the argument that spatial representations have been "exapted" to other more abstract domains [6–8]. For example, experience of different resource distributions in a spatial search task was found to influence behavior in a word generation task, where participants exposed to sparser rewards in space generated sparser semantic clusters of words [79]. Thus, while both spatial and conceptual knowledge are capable of being organized into a common map-like representation, there may something special or central about spatial encoding [80], producing domain differences in terms of the ease of learning such a map and asymmetries in the transfer of knowledge.

Finally, our current experiment only looked at similarities between spatial and conceptual domains if the underlying structure was the same in both tasks. Future studies could expand this approach across different domains such as logical rule-learning, numerical comparisons, or semantic similarities. Additionally, structure learned in one domain could be transferable to structures encountered in either the same domain with slightly changed structures or even to totally different domains with different structures. A truly all-encompassing model of generalization should capture transfer across domains and structural changes. Even though several recent studies have advanced our understanding of how people transfer knowledge across graph structures [81], state similarities in multi-task reinforcement learning [82], and target hypotheses supporting generalization [83], whether or not all of these recruit the same computational principles and neural machinery remains to be seen.

Conclusion

We used a rich experimental paradigm to study how people generalize and explore both spatially and conceptually correlated reward environments. While people employed similar principles of generalization in both domains, we found a substantial shift in exploration, from more uncertainty-directed exploration in the spatial task to more random exploration in the conceptual domain. These results enrich our understanding of the principles connecting generalization and search across different domains and pave the way for future cognitive and neuroscientific investigations into principles of generalization and search across domains.

Methods

Participants and Design

140 participants were recruited through Amazon Mechanical Turk (requiring a 95% approval rate and 100 previously approved HITs) for a two part experiment, where only those who had completed part one were invited back for part two. In total 129 participants completed both parts and were included in the analyses (55 female; mean age=35, SD=9.5). Participants were paid \$4.00 for each part of the experiment, with those completing both parts being paid an additional performance-contingent bonus of up to \$10.00. Participants earned \$15.6 $\pm$ 1.0 and spent 54 $\pm$ 19 minutes completing both parts. There was an average gap of 18 $\pm$ 8.5 hours between the two parts of the experiment. Informed consent was obtained from all participants.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We varied the task order between subjects, with participants completing the spatial and conceptual task in counterbalanced order in separate sessions. We also varied between subjects the extent of reward correlations in the search space by randomly assigning participants to one of two different classes of environments (*smooth* vs. *rough*), with smooth environments corresponding to stronger correlations, and the same environment class used for both tasks (see below).

Materials and Procedure

Each session consisted of a training phase, the main search task, and a bonus round. At the beginning of each session participants were required to complete a training task to familiarize themselves with the stimuli (spatial or conceptual), the inputs (arrow keys and spacebar), and the search space (8×8 feature space). Participants were shown a series of randomly selected targets and were instructed to use the arrow keys to modify a single selected stimuli to match the target (i.e., adjusting the stripe frequency and angle of a Gabor patch or moving the location of a spatial selector, Fig. 1c). The space bar was used to make a selection and feedback was provided for 800ms (correct or incorrect). Participants were required to complete at least 32 training trials and were allowed to proceed to the main task once they had achieved at least 90% accuracy on a run of 10 trials (i.e., 9 out of 10). See Fig S3 for analysis of the training data.

After completing the training, participants were shown instructions for the main search task and had to complete three comprehension questions (Figs S11-S12) to ensure full understanding of the task. Specifically, the questions were designed to ensure participants understood that the spatial or conceptual features predicted reward. Each search task comprised 10 rounds of 20 trials each, with a different reward function sampled without replacement from the set of assigned environments. The reward function specified how rewards mapped onto either the spatial or conceptual features, where participants were told that options with either similar spatial features (Spatial task) [19, 84] or similar conceptual features (Conceptual task) [20, 56] would yield similar rewards. Participants were instructed to accumulate as many points as possible, which were later converted into monetary payoffs.

The tenth round of each sessions was a "bonus round", with additional instructions shown at the beginning of the round. The round began as usual, but after 15 choices, participants were asked to make judgments about the expected rewards (input range: [1,100]) and their level of confidence (Likert scale from least to most confident: [0,10]) for 10 unrevealed targets. These targets were uniformly sampled from the set of unselected options during the current round. After the 10 judgments, participants were asked to make a forced choice between the 10 options. The reward for the selected option was displayed and the round continued as normal. All behavioral and computational modeling analyses exclude the last round, except for the analysis of the bonus round judgments.

Spatial and Conceptual Search Tasks

Participants used the arrow keys to either move a highlighted selector in the spatial task or change the features (tilt and stripe frequency) of the Gabor stimuli in the conceptual task (Fig S1). On each round, participants were given 20 trials to acquire as many cumulative rewards as possible. A selection was made by pressing the space bar, and then participants were given feedback about the reward for 800 ms, with the chosen option and reward value added to the history at the bottom of the screen. At the beginning of each trial, the starting position of the spatial selector or the displayed conceptual stimulus was randomly sampled from a uniform distribution. Each reward observation included normally distributed noise, $\epsilon \sim \mathcal{N}(0, 1)$, where the rewards for

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

each round were scaled to a uniformly sampled maximum value in the range of 80 to 95, so that the value of the global optima in each round could not be easily guessed.

Participants were given feedback about their performance at the end of each round in terms of the ratio of their average reward to the global maximum, expressed as a percentage (e.g., "You have earned 80% of the maximum reward you could have earned on this round"). The performance bonus (up to $10.00) was calculated based on the cumulative performance of each round and across both tasks.

Bonus Round Judgments

In both tasks the last round was a "bonus round", which solicited judgments about the expected reward and their level of confidence for 10 unrevealed options. Participants were informed that the goal of the task remained the same (maximize cumulative rewards), but that after 15 selections, they would be asked to provide judgments about 10 randomly selected options, which had not yet been explored. Judgments about expected rewards were elicited using a slider from 1 to 100 (in increments of 1), while judgments about confidence were elicited using a slider from 0 to 10 (in increments of 1), with the endpoints labeled 'Least confident' and 'Most confident'. After providing the 10 judgments, participants were asked to select one of the options they just rated, and subsequently completed the round like all others.

Environments

All environments were sampled from a GP prior parameterized with a *radial basis function* (RBF) kernel (Eq 4), where the length-scale parameter (λ) determines the rate at which the correlations of rewards decay over (spatial or conceptual) distance. Higher λ-values correspond to stronger correlations. We generated 40 samples of each type of environments, using $\lambda_{rough} = 2$ and $\lambda_{smooth} = 4$, which were sampled without replacement and used as the underlying reward function in each task (Fig S2). Environment type was manipulated between subjects, with the same environment type used in both conceptual and spatial tasks.

Models

Bayesian Mean Tracker

The Bayesian Mean Tracker (BMT) is a simple but widely-applied associative learning model [68, 85, 86], which is a special case of the Kalman Filter with time-invariant reward distributions. The BMT can also be interpreted as a Bayesian variant of the Rescorla-Wagner model [55], making predictions about the rewards of each option j in the form of a normally distributed posterior:

$$P(\mu_{j,t}|\mathcal{D}_t) = \mathcal{N}(m_{j,t}, v_{j,t}) \tag{10}$$

The posterior mean $m_{j,t}$ and variance $v_{j,t}$ are updated iteratively using a delta-rule update based on the observed reward y_t when option j is selected at trial t:

$$m_{j,t} = m_{j,t-1} + \delta_{j,t} G_{j,t} \left[y_t - m_{j,t-1} \right] \tag{11}$$

$$v_{j,t} = \left[1 - \delta_{j,t} G_{j,t} \right] v_{j,t-1} \tag{12}$$

where $\delta_{j,t} = 1$ if option j was chosen on trial t, and 0 otherwise. Rather than having a fixed learning rate, the BMT scales updates based on the Kalman Gain $G_{j,t}$, which is defined as:

$$G_{j,t} = \frac{v_{j,t-1}}{v_{j,t-1} + \theta_\epsilon^2} \tag{13}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where θ_ϵ^2 is the error variance, which is estimated as a free parameter. Intuitively, the estimated mean of the chosen option $m_{j,t}$ is updated based on the prediction error $y_t - m_{j,t-1}$ and scaled by the Kalman Gain $G_{j,t}$ (Eq 11). At the same time, the estimated variance $v_{j,t}$ is reduced by a factor of $1 - G_{j,t}$, which is in the range $[0, 1]$ (Eq 12). The error variance θ_ϵ^2 can be interpreted as an inverse sensitivity, where smaller values result in more substantial updates to the mean $m_{j,t}$, and larger reductions of uncertainty $v_{j,t}$.

Model Cross-validation

As with the behavioral analyses, we omit the 10th "bonus round" in our model cross-validation. For each of the other nine rounds, we use cross validation to iteratively hold out a single round as a test set, and compute the maximum likelihood estimate using differential evolution [87] on the remaining eight rounds. Model comparisons use the summed out-of-sample prediction error on the test set, defined in terms of log loss (i.e., negative log likelihood).

Predictive accuracy

As an intuitive statistic for goodness of fit, we report *predictive accuracy* as a pseudo-R^2:

$$R^2 = 1 - \frac{\log \mathcal{L}(M_k)}{\log \mathcal{L}(M_{rand})} \tag{14}$$

comparing the out-of-sample log loss of a given model M_k against a random model M_{rand}. $R^2 = 0$ indicates chance performance, while $R^2 = 1$ is a theoretically perfect model.

Protected exceedance probability

The protected exceedance probability (pxp) is defined in terms of a Bayesian model selection framework for group studies [70, 71]. Intuitively, it can be described as a random-effect analysis, where models are treated as random effects and are allowed to differ between subjects. Inspired by a Polya's urn model, we can imagine a population containing K different types of models (i.e., people best described by each model) much like an urn containing different colored marbles. If we assume that there is a fixed but unknown distribution of models in the population, what is the probability of each model being more frequent in the population than all other models in consideration?

This is modelled hierarchically, using variational Bayes to estimate the parameters of a Dirichlet distribution describing the posterior probabilities of each model $P(m_k|\mathbf{y})$ given the data $\mathbf{y}$. The exceedance probability is thus defined as the posterior probability that the frequency of a model r_{m_k} is larger than all other models $r_{m_{k' \neq k}}$ under consideration:

$$xp(m_k) = p(r_{m_k} > r_{m_{k' \neq k}}|\mathbf{y}) \tag{15}$$

[71] extends this approach by correcting for chance, based on the Bayesian Omnibus Risk (BOR), which is the posterior probability that all model frequencies are equal:

$$pxp(m_k) = xp(m_k)(1 - BOR) + \frac{BOR}{K} \tag{16}$$

This produces the *protected exceedance probability* (pxp) reported throughout this chapter, and is implemented using `https://github.com/sjgershm/mfit/blob/master/bms.m`.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Simulated learning curves

We simulated each model by sampling (with replacement) from the set of cross-validated participant parameter estimates, and performing search on a simulated bandit task. We performed 10,000 simulations for each combination of model, environment, and domain (spatial vs. conceptual).

Bonus round predictions

Bonus round predictions used each participant's estimated parameters to predict their judgments about expected reward and confidence. Because rewards in each round were randomly scaled to a different global maximum, we also rescaled the model predictions in order to align model predictions with the observed rewards and participant judgments.

Statistical tests

Comparisons

We report both frequentist and Bayesian statistics. Frequentist tests are reported as Student's t-tests (specified as either paired or independent) for parametric comparisons, while the Mann-Whitney-U test or Wilcoxon signed-rank test are used for non-parametric comparisons (for independent samples or paired samples, respectively). Each of these tests are accompanied by a Bayes factors (BF) to quantify the relative evidence the data provide in favor of the alternative hypothesis (H_A) over the null (H_0).

Parametric comparison are tested using the default two-sided Bayesian t-test for either independent or dependent samples, where both use a Jeffreys-Zellner-Siow prior with its scale set to $\sqrt{2}/2$, as suggested by [88]. All statistical tests are non-directional as defined by a symmetric prior (unless otherwise indicated).

Non-parametric comparisons are tested using either the frequentist Mann-Whitney-U test for *independent samples*, or the Wilcoxon signed-rank test for *paired samples*. In both cases, the Bayesian test is based on performing posterior inference over the test statistics (Kendall's r_τ for the Mann-Whitney-U test and standardized effect size $r = \frac{Z}{\sqrt{N}}$ for the Wilcoxon signed-rank test) and assigning a prior using parametric yoking [89]. This leads to a posterior distribution for Kendall's r_τ or the standardized effect size r, which yields an interpretable Bayes factor via the Savage-Dickey density ratio test. The null hypothesis posits that parameters do not differ between the two groups, while the alternative hypothesis posits an effect and assigns an effect size using a Cauchy distribution with the scale parameter set to $1/\sqrt{2}$.

Correlations

For testing linear correlations with Pearson's r, the Bayesian test is based on Jeffrey's [90] test for linear correlation and assumes a shifted, scaled beta prior distribution $B(\frac{1}{k}, \frac{1}{k})$ for r, where the scale parameter is set to $k = \frac{1}{3}$ [91].

For testing rank correlations with Kendall's tau, the Bayesian test is based on parametric yoking to define a prior over the test statistic [92], and performing Bayesian inference to arrive at a posterior distribution for r_τ. The Savage-Dickey density ratio test is used to produce an interpretable Bayes Factor.

Outlier removal

For the analysis of model parameters, we first report the full results using non-parametric tests, comparing differences in parameter estimates using the Wilcoxon

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

signed-rank test and correlations using Kendall's τ. In addition, we also applied a conservative outlier removal procedure based on Tukey's outlier removal criterion, where we removed values larger than Q3 + 1.5×IQR and ran standard t-tests and Pearson's correlations.

ANOVA

We use a two-way mixed-design analysis of variance (ANOVA) to compare the means of both a fixed effects factor (smooth vs. rough environments) as a between-subjects variable and a random effects factor (conceptual vs. spatial) as a within-subjects variable. To compute the Bayes Factor, we assume independent g-priors [93] for each effect size $\theta_1 \sim \mathcal{N}(0, g_1\sigma^2), \cdots, \theta_p \sim \mathcal{N}(0, g_p\sigma^2)$, where each g-value is drawn from an inverse chi-square prior with a single degree of freedom $g_i \overset{\text{i.i.d}}{\sim}$ inverse-$\chi^2(1)$, and assuming a Jeffreys prior on the aggregate mean and scale factor. Following [94], we compute the Bayes factor by integrating the likelihoods with respect to the prior on parameters, where Monte Carlo sampling was used to approximate the g-priors. The Bayes factor reported in the text can be interpreted as the log-odds of the model relative to an intercept-only null model.

Acknowledgments

ES is supported by the Harvard Data Science Initiative. We thank Daniel Reznik, Nicholas Franklin, Samuel Gershman, Christian Doeller, and Fiery Cushman for helpful discussions.

Author contributions statement

C.M.W., E.S., B.M., and N.W.S. conceived the experiment, C.M.W. conducted the experiment and analysed the results. All authors wrote and reviewed the manuscript.

References

1. William James. *The Principles of Psychology.* Dover, New York, 1890.
2. Frances Amelia Yates. *Art of Memory.* Routledge, 2013.
3. Martin Dresler, William R Shirer, Boris N Konrad, Nils CJ Müller, Isabella C Wagner, Guillén Fernández, Michael Czisch, and Michael D Greicius. Mnemonic training reshapes brain networks to support superior memory. *Neuron*, 93:1227–1235, 2017.
4. Barbara Landau and Ray Jackendoff. Whence and whither in spatial language and spatial cognition? *Behavioral and Brain Sciences*, 16:255–265, 1993.
5. George Lakoff and Mark Johnson. *Metaphors We Live By.* University of Chicago press, 2008.
6. Peter M Todd, Thomas T Hills, and Trevor W Robbins. *Cognitive search: Evolution, algorithms, and the brain.* MIT press, 2012.
7. Thomas T Hills, Peter M Todd, and Robert L Goldstone. Search in external and internal spaces: Evidence for generalized cognitive search processes. *Psychological Science*, 19:802–808, 2008.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8. Thomas T Hills. Animal foraging and the evolution of goal-directed cognition. *Cognitive Science*, 30:3–41, 2006.

9. Alexandra O Constantinescu, Jill X O'Reilly, and Timothy EJ Behrens. Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352:1464–1468, 2016.

10. Mona M Garvert, Raymond J Dolan, and Timothy EJ Behrens. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife*, 6:e17086, 2017.

11. Nicolas W Schuck, Ming Bo Cai, Robert C Wilson, and Yael Niv. Human orbitofrontal cortex represents a cognitive map of state space. *Neuron*, 91:1402–1412, 2016.

12. Dmitriy Aronov, Rhino Nevers, and David W Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719, 2017.

13. Ethan A Solomon, Bradley C Lega, Michael R Sperling, and Michael J Kahana. Hippocampal theta codes for distances in semantic and temporal spaces. *Proceedings of the National Academy of Sciences*, 116(48):24343–24352, 2019.

14. Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. Cambridge: MIT Press, 1998.

15. Edward L Thorndike. Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, 2(4):i, 1898.

16. Ivan Petrovich Pavlov. *Conditional reflexes: an investigation of the physiological activity of the cerebral cortex.* Oxford University Press, 1927.

17. Burrhus Frederic Skinner. *The behavior of organisms: An experimental analysis*. Appleton-Century, New York, 1938.

18. Peter Dayan. Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5:613–624, 1993.

19. Charley M Wu, Eric Schulz, Maarten Speekenbrink, Jonathan D Nelson, and Björn Meder. Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, 2:915—-924, 2018.

20. Hrvoje Stojic, Eric Schulz, Pantelis P Analytis, and Maarten Speekenbrink. It's new, but is it good? how generalization and uncertainty guide the exploration of novel options. 2018.

21. Edward C Tolman. Cognitive maps in rats and men. *Psychological Review*, 55:189–208, 1948.

22. Perry W Thorndyke. Distance estimation from cognitive maps. *Cognitive psychology*, 13(4):526–550, 1981.

23. John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.

24. John O'Keefe. A review of the hippocampal place cells. *Progress in neurobiology*, 13(4):419–439, 1979.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. Torkel Hafting, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801, 2005.

26. Jeffrey S Taube, Robert U Muller, and James B Ranck. Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, 10(2):420–435, 1990.

27. Jeffrey S Taube. Head direction cells and the neurophysiological basis for a sense of direction. *Progress in neurobiology*, 55(3):225–256, 1998.

28. Colin Lever, Stephen Burton, Ali Jeewajee, John O'Keefe, and Neil Burgess. Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, 29(31):9771–9777, 2009.

29. Trygve Solstad, Charlotte N Boccara, Emilio Kropff, May-Britt Moser, and Edvard I Moser. Representation of geometric borders in the entorhinal cortex. *Science*, 322(5909):1865–1868, 2008.

30. Øyvind Arne Høydal, Emilie Ranheim Skytøen, Sebastian Ola Andersson, May-Britt Moser, and Edvard I Moser. Object-vector coding in the medial entorhinal cortex. *Nature*, 568(7752):400, 2019.

31. Russell A Epstein, Eva Zita Patai, Joshua B Julian, and Hugo J Spiers. The cognitive map in humans: spatial navigation and beyond. *Nature neuroscience*, 20(11):1504, 2017.

32. Timothy EJ Behrens, Timothy H Muller, James CR Whittington, Shirley Mark, Alon B Baram, Kimberly L Stachenfeld, and Zeb Kurth-Nelson. What is a cognitive map? organizing knowledge for flexible behavior. *Neuron*, 100(2):490–509, 2018.

33. Raphael Kaplan, Nicolas W Schuck, and Christian F Doeller. The role of mental maps in decision-making. *Trends in Neurosciences*, 40:256–259, 2017.

34. Jacob LS Bellmund, Peter Gärdenfors, Edvard I Moser, and Christian F Doeller. Navigating cognition: Spatial codes for human thinking. *Science*, 362(6415):eaat6766, 2018.

35. Howard Eichenbaum. Hippocampus: remembering the choices. *Neuron*, 77(6):999–1001, 2013.

36. Hugo J. Spiers. The hippocampal cognitive map: One space or many? *Trends in Cognitive Sciences*, 2020.

37. Daniela Schiller, Howard Eichenbaum, Elizabeth A Buffalo, Lila Davachi, David J Foster, Stefan Leutgeb, and Charan Ranganath. Memory and space: towards an understanding of the cognitive map. *Journal of Neuroscience*, 35(41):13904–13911, 2015.

38. Benjamin J Kraus, Robert J Robinson II, John A White, Howard Eichenbaum, and Michael E Hasselmo. Hippocampal "time cells": time versus path integration. *Neuron*, 78(6):1090–1101, 2013.

39. Christopher J MacDonald, Stephen Carrow, Ryan Place, and Howard Eichenbaum. Distinct hippocampal time cell sequences represent odor memories in immobilized rats. *Journal of Neuroscience*, 33(36):14607–14616, 2013.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. Christian F Doeller, Caswell Barry, and Neil Burgess. Evidence for grid cells in a human memory network. *Nature*, 463(7281):657, 2010.

41. Rita Morais Tavares, Avi Mendelsohn, Yael Grossman, Christian Hamilton Williams, Matthew Shapiro, Yaacov Trope, and Daniela Schiller. A map for social navigation in the human brain. *Neuron*, 87(1):231–243, 2015.

42. Nicolas W Schuck and Yael Niv. Sequential replay of nonspatial task states in the human hippocampus. *Science*, 364(6447):eaaw5181, 2019.

43. Joshua Jacobs, Christoph T Weidemann, Jonathan F Miller, Alec Solway, John F Burke, Xue-Xin Wei, Nanthia Suthana, Michael R Sperling, Ashwini D Sharan, Itzhak Fried, et al. Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature neuroscience*, 16(9):1188, 2013.

44. Nicolas W Schuck, Robert Wilson, and Yael Niv. A state representation for reinforcement learning and decision-making in the orbitofrontal cortex. In *Goal-Directed Decision Making*, pages 259–278. Elsevier, 2018.

45. Yael Niv. Learning task-state representations. *Nature neuroscience*, 22(10):1544–1553, 2019.

46. CE. Rasmussen and CKI. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, 2006.

47. Eric Schulz, Maarten Speekenbrink, and Andreas Krause. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *bioRxiv*, 2017.

48. Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

49. Robert C Wilson, Andra Geana, John M White, Elliot A Ludvig, and Jonathan D Cohen. Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, 143:155–164, 2014.

50. Eric Schulz and Samuel J Gershman. The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology*, 55:7–14, 2019.

51. Mark Steyvers, Michael D Lee, and Eric-Jan Wagenmakers. A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, 53:168–179, 2009.

52. Daniel Acuna and Paul Schrater. Bayesian modeling of human sequential decision-making on the multi-armed bandit problem. In *Proceedings of the 30th annual conference of the cognitive science society*, volume 100, pages 200–300. Washington, DC: Cognitive Science Society, 2008.

53. Robert R Bush and Frederick Mosteller. A mathematical model for simple learning. *Psychological Review*, 58:313, 1951.

54. Robert A Rescorla and Allan R Wagner. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 2:64–99, 1972.

55. Samuel J Gershman. A unifying probabilistic view of associative learning. *PLoS Computational Biology*, 11(11):e1004567, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

56. Eric Schulz, Emmanouil Konstantinidis, and Maarten Speekenbrink. Putting bandits into context: How function learning supports decision making. *Journal of experimental psychology: learning, memory, and cognition*, 44(6):927, 2018.

57. J Douglas Carroll. Functional learning: The learning of continuous functional mappings relating stimulus and response continua. *ETS Research Bulletin Series*, 1963:i–144, 1963.

58. Christopher G Lucas, Thomas L Griffiths, Joseph J Williams, and Michael L Kalish. A rational model of function learning. *Psychonomic Bulletin & Review*, 22(5):1193–1215, 2015.

59. Thomas L Griffiths, Chris Lucas, Joseph Williams, and Michael L Kalish. Modeling human function learning with gaussian processes. In *Advances in Neural Information Processing Systems*, pages 553–560, 2009.

60. Eric Schulz, Joshua B. Tenenbaum, David Duvenaud, Maarten Speekenbrink, and Samuel J. Gershman. Compositional inductive biases in function learning. *Cognitive Psychology*, 99:44 – 79, 2017.

61. Kyunghee Koh and David E Meyer. Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17:811, 1991.

62. J. R. Busemeyer, E. Byun, E. L. DeLosh, and M. A. McDaniel. Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In K. Lamberts and D. Shanks, editors, *Concepts and Categories*, pages 405–437. MIT Press, Cambridge, 1997.

63. Eric Schulz, Joshua B Tenenbaum, David N Reshef, Maarten Speekenbrink, and Samuel Gershman. Assessing the perceived predictability of functions. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, pages 2116–2121, Cognitive Science Society, 2015.

64. Frank Jäkel, Bernhard Schölkopf, and Felix A Wichmann. Similarity, kernels, and the triangle inequality. *Journal of Mathematical Psychology*, 52:297–303, 2008.

65. Tomaso Poggio and Emilio Bizzi. Generalization in vision and motor control. *Nature*, 431:768–774, 2004.

66. Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. *Proceedings of the 27th International Conference on Machine Learning (ICML 2010)*, pages 1015–1022, 2010.

67. Samuel J Gershman. Deconstructing the human algorithms for exploration. *Cognition*, 173:34–42, 2018.

68. Maarten Speekenbrink and Emmanouil Konstantinidis. Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, 7:351–367, 2015.

69. Eric Schulz, Charley M Wu, Azzurra Ruggeri, and Bjoern Meder. Searching for rewards like a child means less generalization and more directed exploration. *Psychological Science*, 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

70. Klaas Enno Stephan, Will D Penny, Jean Daunizeau, Rosalyn J Moran, and Karl J Friston. Bayesian model selection for group studies. *Neuroimage*, 46:1004–1017, 2009.

71. Lionel Rigoux, Klaas Enno Stephan, Karl J Friston, and Jean Daunizeau. Bayesian model selection for group studies—revisited. *Neuroimage*, 84:971–985, 2014.

72. Eric Schulz, Charley M Wu, Quentin JM Huys, Andreas Krause, and Maarten Speekenbrink. Generalization and search in risky environments. *Cognitive Science*, 42:2592–2620, 2018.

73. Charley M Wu, Eric Schulz, Mona M Garvert, Björn Meder, and Nicolas W Schuck. Connecting conceptual and spatial search via a model of generalization. In T. T. Rogers, M. Rau, X. Zhu, and C. W. Kalish, editors, *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, pages 1183–1188, Austin, TX, 2018. Cognitive Science Society.

74. Angela Radulescu, Yael Niv, and Ian Ballard. Holistic reinforcement learning: The role of structure and attention. *Trends in Cognitive Sciences*, 23:278–292, 2019.

75. Anne GE Collins and Michael J Frank. Within-and across-trial dynamics of human eeg reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences*, 115:2502–2507, 2018.

76. Sven Ohl and Martin Rolfs. Saccadic selection of stabilized items in visuospatial working memory. *Consciousness and Cognition*, 64:32–44, 2018.

77. Irene Cogliati Dezza, Axel Cleeremans, and William Alexander. Should we control? the interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma. *Journal of Experimental Psychology: General*, 2019.

78. Charley M Wu, Eric Schulz, Kimberly Gerbaulet, Timothy J Pleskac, and Maarten Speekenbrink. Under pressure: The influence of time limits on human exploration. In A.K. Goel, C.M. Seifert, and C. Freksa, editors, *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, pages 1219—-1225, Montreal, QB, 2019. Cognitive Science Society.

79. Thomas T Hills, Peter M Todd, and Robert L Goldstone. The central executive as a search process: Priming exploration and exploitation across domains. *Journal of Experimental Psychology: General*, 139(4):590, 2010.

80. Lynn Nadel. The hippocampus and space revisited. *Hippocampus*, 1(3):221–229, 1991.

81. Shirley Mark, Rani Moran, Thomas Parr, Steve Kennerley, and Tim Behrens. Transferring structural knowledge across cognitive maps in humans and models. *bioRxiv*, 2019.

82. Momchil Tomov, Eric Schulz, and Samuel J Gershman. Multi-task reinforcement learning in humans. *bioRxiv*, page 815332, 2019.

83. Joseph L Austerweil, Sophia Sanborn, and Thomas L Griffiths. Learning how to generalize. *Cognitive science*, 43(8), 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

84. Charley M Wu, Eric Schulz, Maarten Speekenbrink, Jonathan D Nelson, and Björn Meder. Mapping the unknown: The spatially correlated multi-armed bandit. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, pages 1357–1362, 2017.

85. Aaron C Courville and Nathaniel D Daw. The rat as particle filter. In *Advances in neural information processing systems*, pages 369–376, 2008.

86. Danielle J Navarro, Peter Tran, and Nicole Baz. Aversion to option loss in a restless bandit task. *Computational Brain & Behavior*, 1(2):151–164, 2018.

87. Katharine Mullen, David Ardia, David L Gil, Donald Windover, and James Cline. Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26, 2011.

88. Jeffrey N Rouder, Paul L Speckman, Dongchu Sun, Richard D Morey, and Geoffrey Iverson. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16:225–237, 2009.

89. Johnny van Doorn, Alexander Ly, Maarten Marsman, and Eric-Jan Wagenmakers. Bayesian latent-normal inference for the rank sum test, the signed rank test, and Spearman's ρ. *arXiv preprint arXiv:1712.06941*, 2017.

90. Harold Jeffreys. *The Theory of Probability*. Oxford, UK: Oxford University Press, 1961.

91. Alexander Ly, Josine Verhagen, and Eric-Jan Wagenmakers. Harold jeffreys's default bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, 72:19–32, 2016.

92. Johnny van Doorn, Alexander Ly, Maarten Marsman, and Eric-Jan Wagenmakers. Bayesian inference for kendall's rank correlation coefficient. *The American Statistician*, 72:303–308, 2018.

93. Arnold Zellner and Aloysius Siow. Posterior odds ratios for selected regression hypotheses. In J. M. Bernardo, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics: Proceedings of the First International Meeting held in Valencia (Spain)*, pages 585–603, University of Valencia, 1980.

94. Jeffrey N Rouder, Richard D Morey, Paul L Speckman, and Jordan M Province. Default bayes factors for anova designs. *Journal of Mathematical Psychology*, 56:356–374, 2012.

95. Joseph Austerweil and Thomas Griffiths. Learning hypothesis spaces and dimensions through concept learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32, 2010.

96. Paul-Christian Bürkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting information

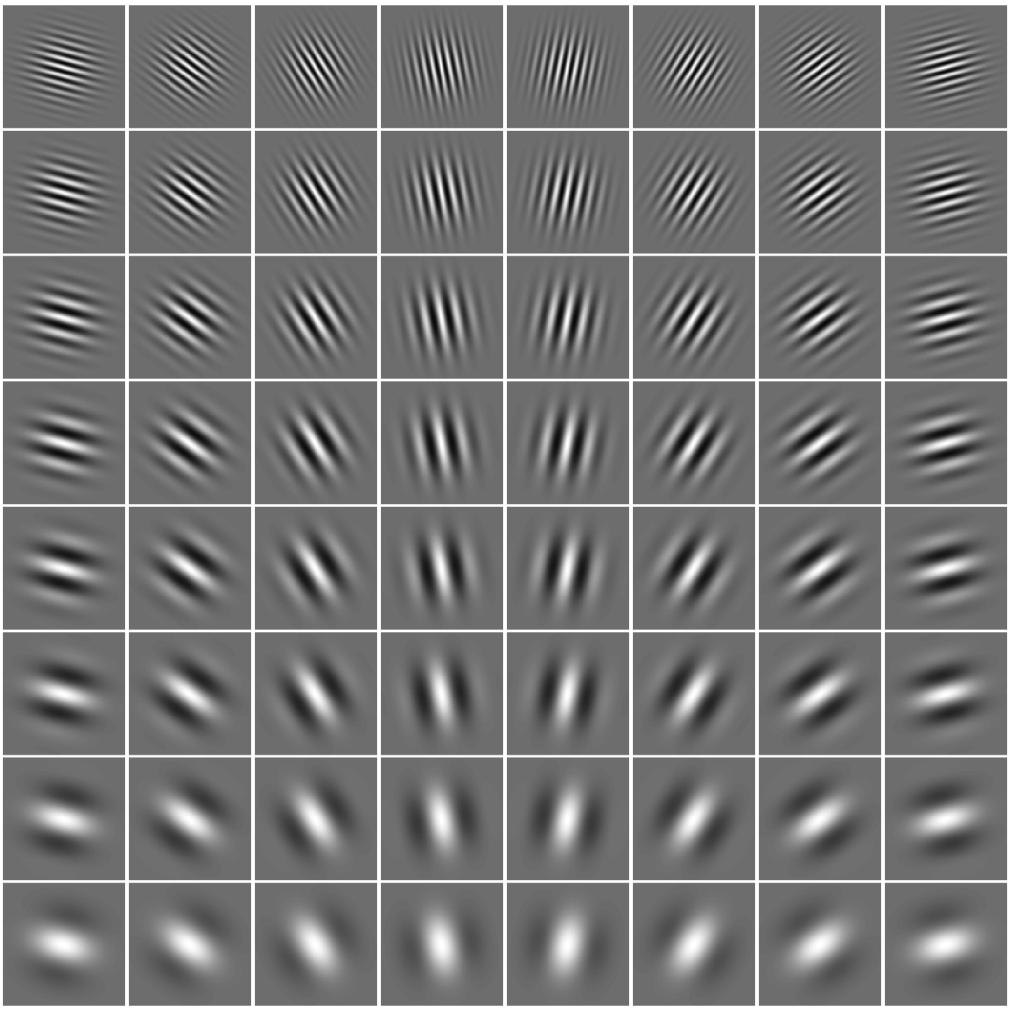

Figure S1. Gabor stimuli. Tilt varies from left to right from 105° to 255° in equally spaced intervals, while stripe frequency increases moving upwards from 1.5 to 15 in log intervals.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

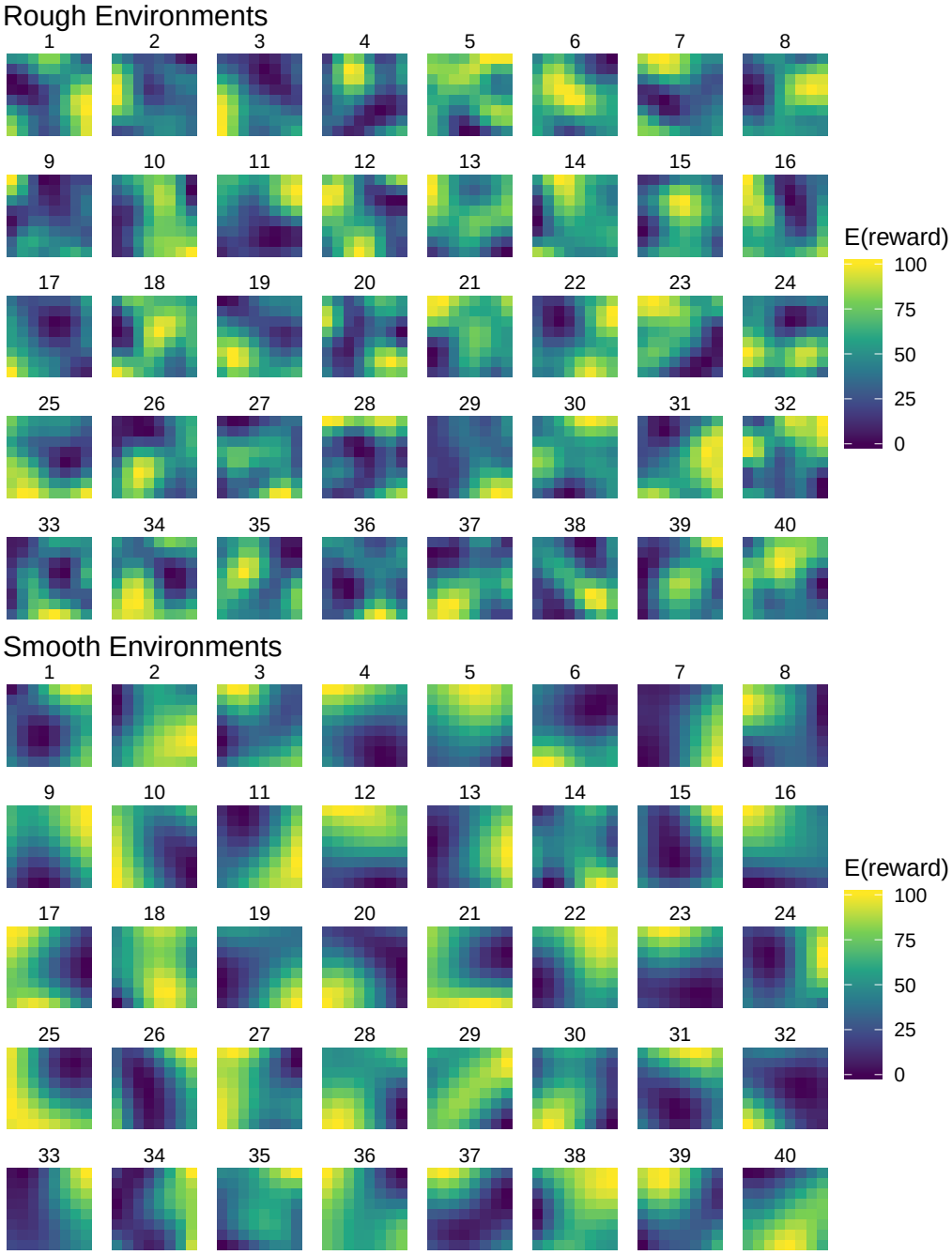

Figure S2. Correlated reward environments. Heatmaps of the reward environments used in both spatial and conceptual domains. The color of each tile represents the expected reward of the bandit, where the x-axis and y-axis were mapped to the spatial location or the tilt and stripe frequency (respectively).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

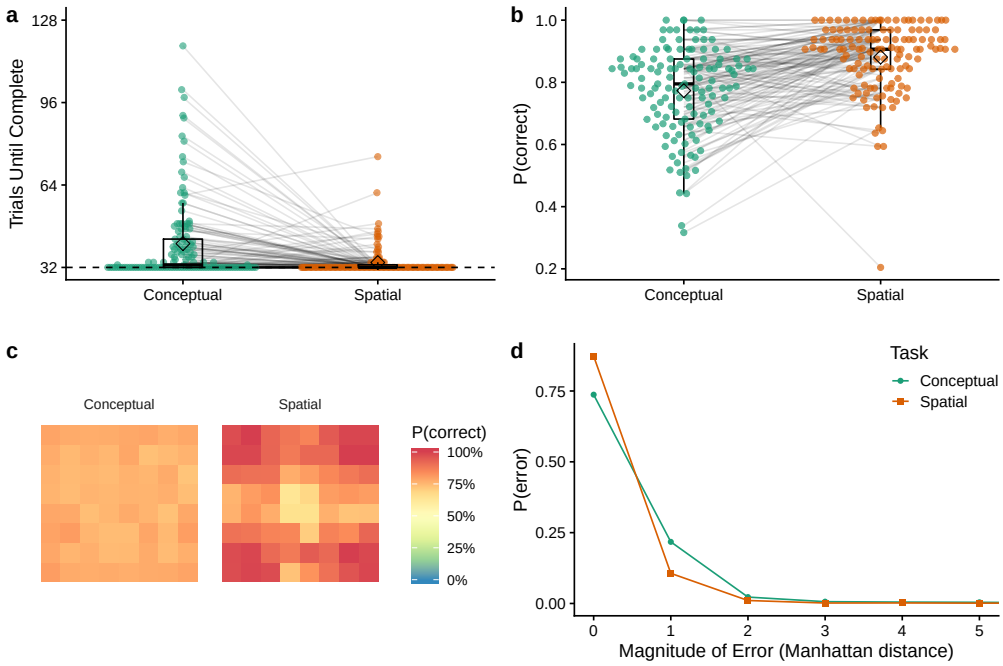

Figure S3. Training Phase. **a**) Trials needed to reach the learning criterion (90% accuracy over 10 trials) in the training phase, where the dotted line indicates the 32 trial minimum. Each dot is a single participant with lines connecting the same participant. Tukey boxplots show median (line) and 1.5x IQR, with diamonds indicating group means. **b**) Average correct choices during the training phase. **c**) Heatmaps of the accuracy of different target stimuli, where the x and y-axes of the conceptual heatmap indicate tilt and stripe frequency, respectively. **d**) The probability of error as a function of the magnitude of error (Manhattan distance from the correct response). Thus, most errors were close to the target, with higher magnitude errors being monotonically less likely to occur.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

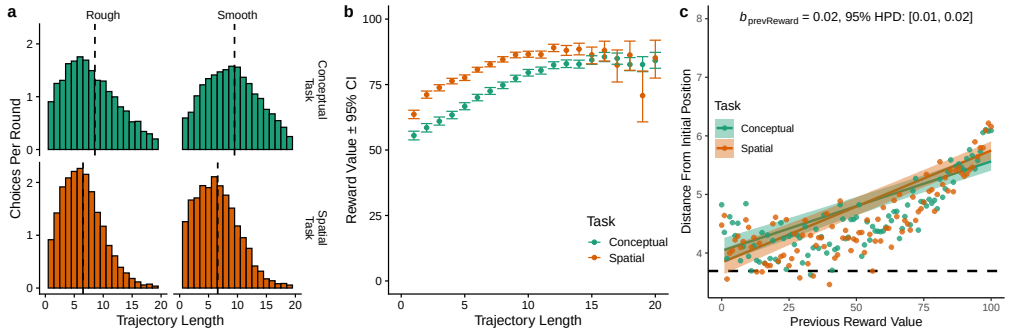

Figure S4. Search Trajectories. **a**) Distribution of trajectory length, separated by task and environment. The dashed vertical line indicates the median for each category. Participants had longer trajectories in the contextual task ($t(128) = -10.7$, $p < .001$, $d = 1.0$, $BF > 100$), but there were no differences across environments ($t(127) = 1.3$, $p = .213$, $d = 0.2$, $BF = .38$). **b**) Average reward value as a function of trajectory length, showing how longer trajectories generally resulted in higher rewards ($r = .23$, $p < .001$, $BF > 100$). Each dot is a mean with error bars showing the 95% CI. **c**) Distance from the random initial starting point in each trial as function of the previous reward value. Each dot is the aggregate mean and the dashed line indicates random chance. Lines are the fixed effects of a Bayesian mixed-effects model (see Table S1), with the ribbons indicating the 95% HPD.

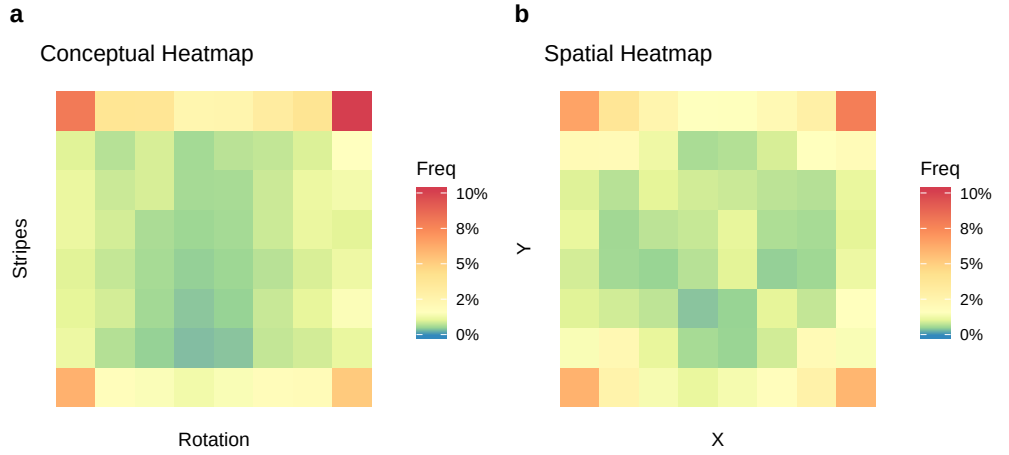

Figure S5. Heatmaps of choice frequency. Heatmaps of chosen options in **a**) the Gabor feature of the conceptual task and **b**) the spatial location of the spatial task, aggregated over all participants. The color shows the frequency of each option centered on yellow representing random chance (1/64), with orange and red indicating higher than chance, while green and blue were lower than chance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

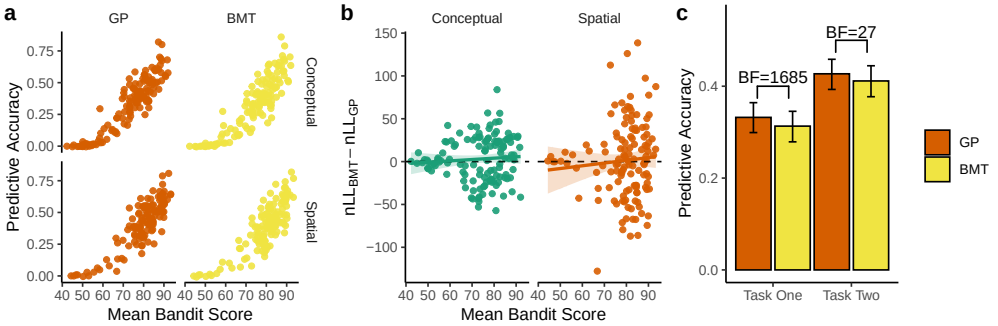

Figure S6. Additional Modeling Results. **a**) The relationship between mean performance and predictive accuracy, where in all cases, the best performing participants were also the best described. **b**) The best performing participants were also the most diagnostic between models, but not substantially skewed towards either model. Linear regression lines strongly overlap with the dotted line at $y = 0$, where participants above the line were better described by the GP model. **c** Model comparison split by which task was performed first vs. second. In both cases, participants were better described on their second task, although the superiority of the GP over the BMT remains, comparing only task one (paired t-test: $t(128) = 4.6$, $p < .001$, $d = 0.10$, $BF = 1685$) or only task two ($t(128) = 3.5$, $p < .001$, $d = 0.08$, $BF = 27$).

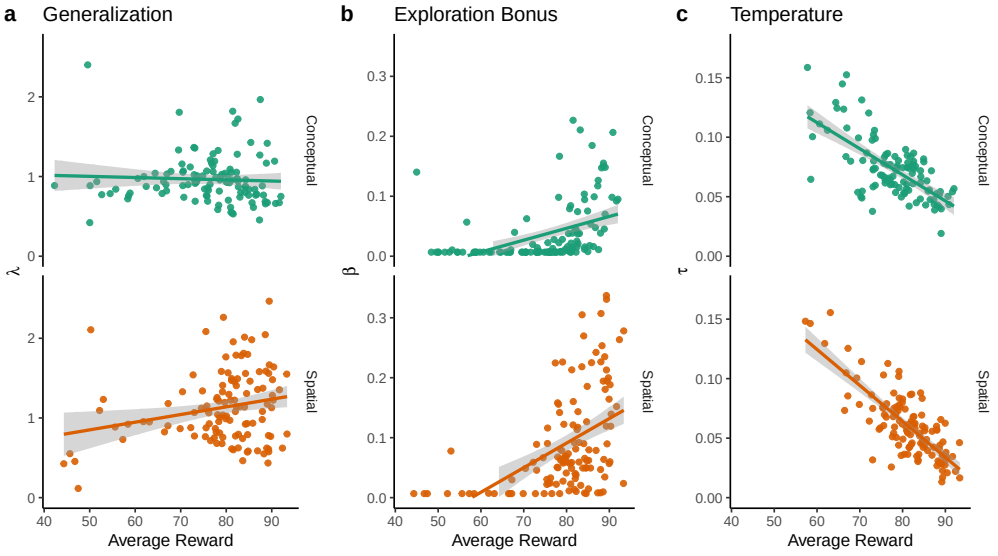

Figure S7. GP parameters and performance. **a**) We do not find a reliable relationship between λ estimates and performance in either the spatial task ($r_\tau = .13$, $p = .030$, $BF = 1.2$; outliers removed: $r = -.05$, $p = .600$, $BF = .25$) or the conceptual task ($r_\tau = -.22$, $p < .001$, $BF > 100$; outliers removed: $r = .23$, $p = .012$, $BF = 4.3$; note the opposite signs of the correlation coefficients). **b**) Higher beta estimates were strongly predictive of better performance in both conceptual ($r_\tau = .32$, $p < .001$, $BF > 100$; outliers removed: $r = .43$, $p < .001$, $BF > 100$) and spatial tasks ($r_\tau = .31$, $p < .001$, $BF > 100$; outliers removed: $r = .48$, $p < .001$, $BF > 100$). **c**) On the other hand, high temperature values predicted lower performance in both conceptual($r_\tau = -.59$, $p < .001$, $BF > 100$; outliers removed: $r = -.68$, $p < .001$, $BF > 100$) and spatial tasks ($r_\tau = -.58$, $p < .001$, $BF > 100$; outliers removed: $r = -.78$, $p < .001$, $BF > 100$).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

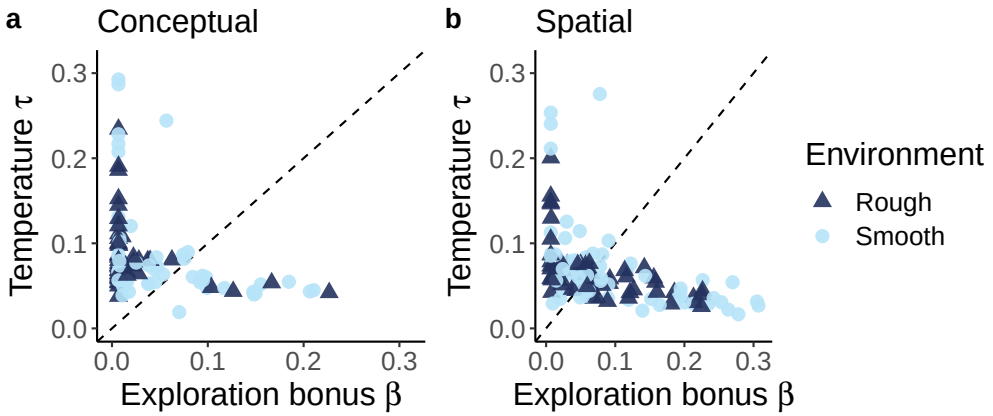

Figure S8. GP exploration bonus and temperature. We check here whether there exists any inverse relationship between directed and undirected exploration, implemented using the UCB exploration bonus β (x-axis) and the softmax temperature τ (y-axis), respectively. Results are split into conceptual (**a**) and spatial tasks (**b**), where each dot is a single participant and the dotted line indicates $y = x$. The upper axis limits are set to the largest 1.5×IQR, for both β and τ, across both conceptual and spatial tasks.

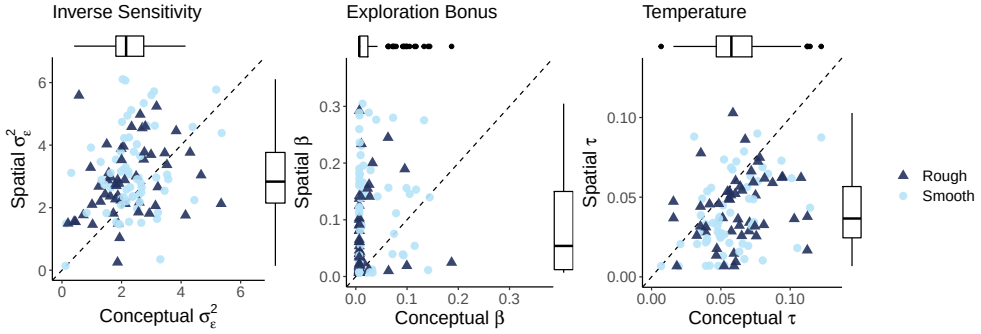

Figure S9. BMT parameters. Each dot is a single participant and the dotted line indicates $y = x$. **a**) We found lower error variance (σ^2_ϵ) estimates in the conceptual task (Wilcoxon signed-rank test: $Z = -4.8$, $p < .001$, $r = -.42$, $BF > 100$; with outliers removed: $t(119) = -5.5$, $p < .001$, $d = 0.6$, $BF > 100$), suggesting participants were more sensitive to the reward values (i.e., more substantial updates to their means estimates). Error variance was also weakly correlated('$r_\tau = .18$, $p = .003$, $BF = 10$; without outliers: $r = .28$, $p = .002$, $BF = 21$). **b**) As with the GP model reported in the main text, we also found strong differences in exploration behavior in the BMT. We found lower estimates of the exploration bonus in the conceptual task ($Z = -5.9$, $p < .001$, $r = -.52$, $BF > 100$; without outliers: $t(119) = -7.9$, $p < .001$, $d = 1.0$, $BF > 100$). There is ambiguous evidence about correlations between tasks ($r_\tau = .16$, $p = .006$, $BF = 4.8$; without outliers:$r = .14$, $p = .141$, $BF = .59$). **c**) Also in line with the GP results, we again find an increase in random exploration in the conceptual task ($Z = -6.9$, $p < .001$, $r = -.61$, $BF > 100$; without outliers: $t(108) = 7.6$, $p < .001$, $d = 0.8$, $BF > 100$). Once more, temperature estimates were strongly correlated ($r_\tau = .34$, $p < .001$, $BF > 100$; without outliers $r = .33$, $p < .001$, $BF = 70$).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

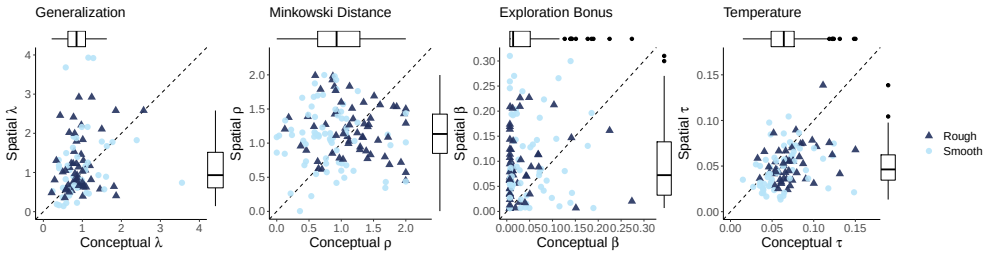

Figure S10. Shepard kernel parameters. We also considered an alternative form of the GP model. Instead of modeling generalization as a function of squared-Euclidean distance with the RBF kernel, we use the Shepard kernel described in [64], where we instead use Minkowski distance with the free parameter $\rho \in [0, 2]$. This model is identical to the GP model reported in the main text when $\rho = 2$. But when $\rho < 2$, the input dimensions transition from integral to separable representations [95]. The lack of clear differences in model parameters motivated us to only include the standard RBF kernel in the main text. **a**) We find mixed evidence for differences in generalization between tasks ($Z = -1.8$, $p = .039$, $r = -.15$, $BF = .32$; outliers removed: $t(98) = -2.8$, $p = .007$, $d = 0.4$, $BF = 4.1$). There is also marginal evidence of correlated estimates ($r_\tau = .13$, $p = .026$, $BF = 1.3$; outliers removed: $r = .21$, $p = .033$, $BF = 2.0$). **b**) There is anecdotal evidence of lower ρ estimates in the conceptual task ($Z = -2.5$, $p = .006$, $r = -.22$, $BF = 2.0$; outliers removed: $t(128) = -2.7$, $p = .008$, $d = 0.3$, $BF = 3.3$). The implication of a lower ρ in the conceptual domain is that the Gabor features were treated more independently, whereas the spatial dimensions were more integrated. However, the statistics suggest this is not a very robust effect. These estimates are also not correlated ($r_\tau = -.02$, $p = .684$, $BF = .12$; outliers removed: $r = -.04$, $p = .653$, $BF = .22$). **c**) Consistent with all the other models, we find systematically lower exploration bonuses in the conceptual task ($Z = -5.5$, $p < .001$, $r = -.49$, $BF > 100$; outliers removed: $t(121) = -6.6$, $p < .001$, $d = 0.8$, $BF > 100$). There is ambiguous evidence of a correlation across tasks ($r_\tau = .14$, $p = .021$, $BF = 1.6$; outliers removed: $r = .09$, $p = .338$, $BF = .32$). **d**) We find clear evidence of higher temperatures in the conceptual task ($Z = -6.3$, $p < .001$, $r = -.56$, $BF > 100$; outliers removed: $t(105) = 6.5$, $p < .001$, $d = 0.7$, $BF > 100$), with strong correlations across tasks ($r_\tau = .41$, $p < .001$, $BF > 100$; $r = .32$, $p < .001$, $BF = 45$)

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

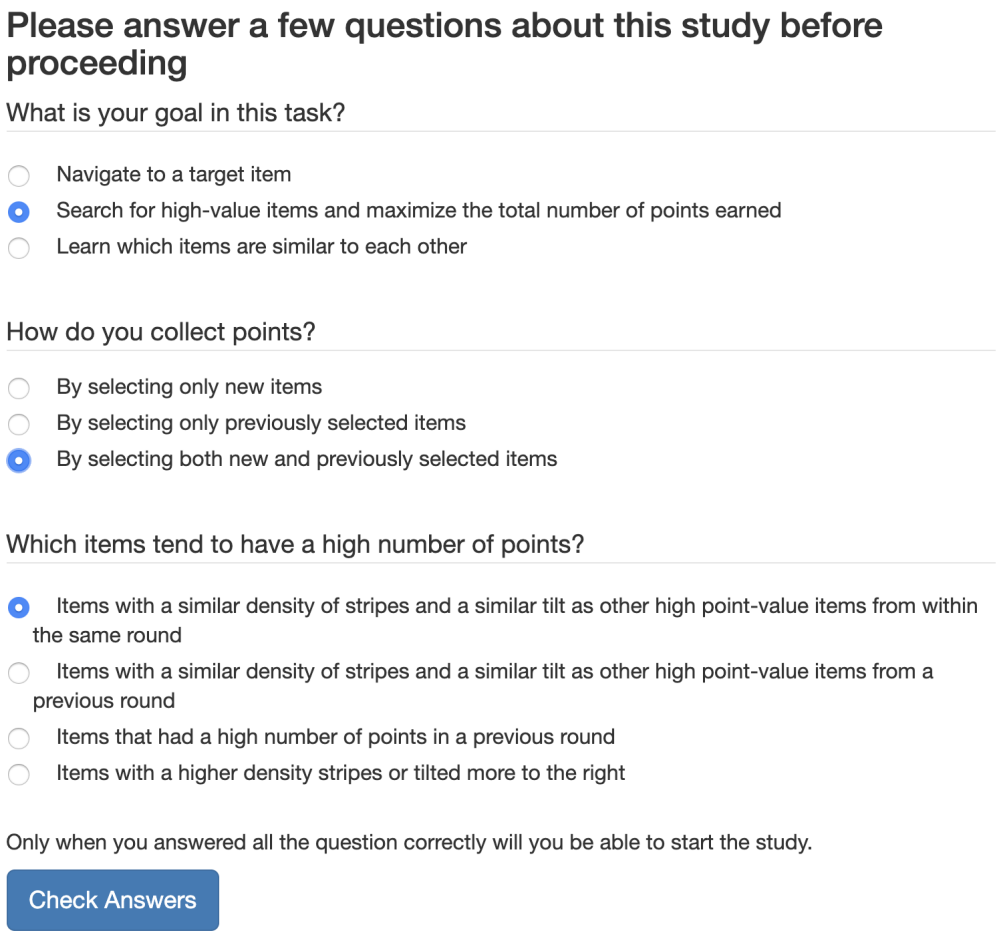

Figure S11. Comprehension questions for the conceptual task. The correct answers are highlighted.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

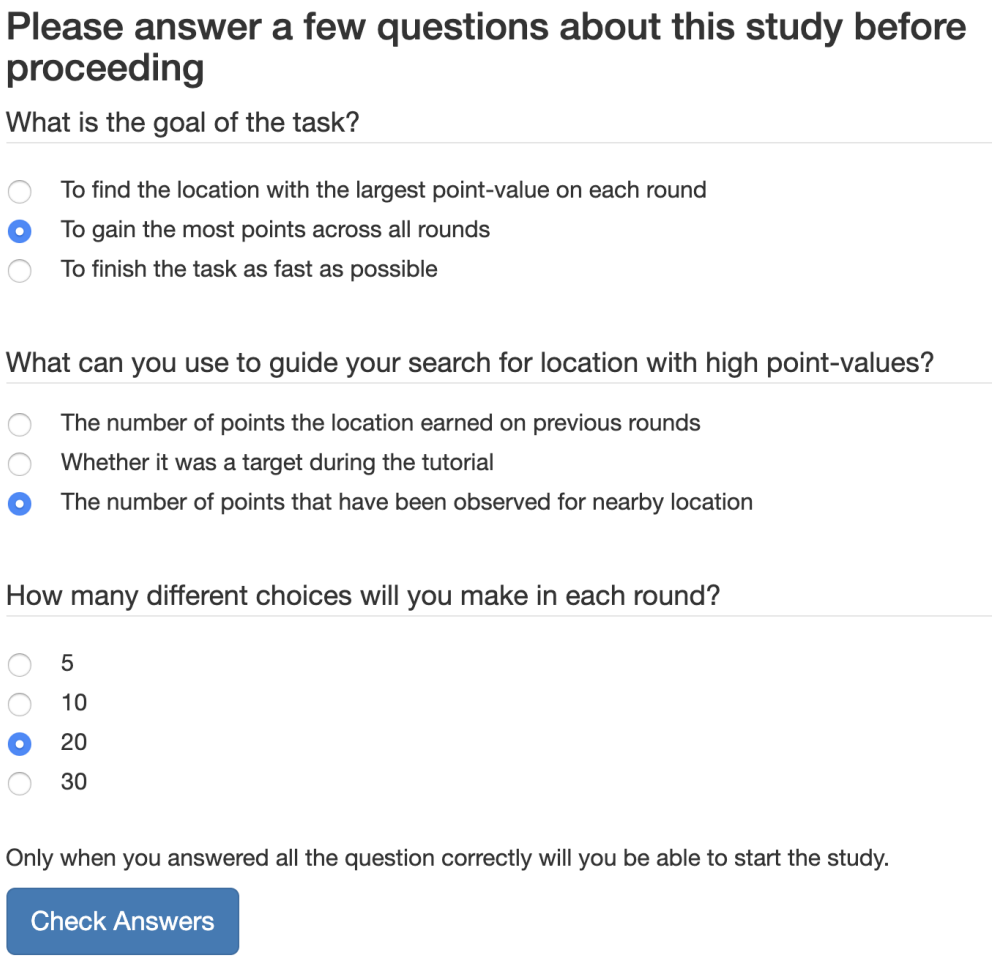

Figure S12. Comprehension questions for the spatial task. The correct answers are highlighted.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1. Mixed Effects Regression Results: Previous Reward

	Distance Between Choices		Distance from Initial Position	
Predictors	*Est.*	*95% HPD*	*Est.*	*95% HPD*
Intercept	7.31	7.07 – 7.56	4.04	3.83 – 4.25
PreviousReward	-0.06	-0.07 – -0.06	0.02	0.01 – 0.02
Spatialtask	0.57	0.46 – 0.69	-0.20	-0.36 – -0.04
PreviousReward:Spatialtask	-0.01	-0.01 – -0.01	0.00	0.00 – 0.01
Random Effects				
σ^2	0.71		0.88	
τ_{00}	7.61		8.55	
ICC	0.09		0.09	
N	129		129	
Observations	44118		441818	
Bayesian R^2	.509		.118	

Note: Both models were implemented in `brms` with default weak priors [96]. We report the posterior mean (Est.) and 95% highest posterior density (HPD) interval. σ^2 indicates the individual-level variance and τ_{00} indicates the variation between individual intercepts and the average intercept.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.21.914556; this version posted January 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S2. Mixed Effects Regression Results: Bonus round judgments

	Model Prediction		Model Uncertainty	
Predictors	*Est.*	*95% HPD*	*Est.*	*95% HPD*
Intercept	40	37.52 – 42.41	0.96	0.88 – 1.03
ParticipantJudgment	0.25	0.20 – 0.31	-0.02	-0.03 – -0.01
Spatialtask	-2.33	-4.28 – -0.33	-0.14	-0.20 – -0.08
ParticipantJudgment:Spatialtask	0.06	0.01 – 0.10	0.01	0.00 – 0.02
Random Effects				
σ^2	17.45		0.03	
τ_{00}	130.57		0.06	
ICC	0.12		0.31	
N	129		129	
Observations	2580		2580	
Bayesian R^2	.313		.332	

Note: Both models were implemented in `brms` with default weak priors [96]. We report the posterior mean (Est.) and 95% highest posterior density (HPD) interval. In the first model (Model Prediction), participant judgments in the range [1,100] are used to predict the GP posterior mean, whereas the second model (Model Uncertainty) uses confidence judgments in the range [1,11] to predict the GP posterior variance. All GP posteriors are computed based on individual participant λ-values, estimated from the corresponding bandit task. σ^2 indicates the individual-level variance and τ_{00} indicates the variation between individual intercepts and the average intercept.