Approximate Planning in Spatial Search

Marta Kryven[1,*] Suhyoun Yu[2] Max Kleiman-Weiner[1]

Tomer Ullman[3,†] Joshua Tenenbaum[1,†]

Abstract

How people plan is an active area of research in cognitive science, neuroscience, and artificial intelligence. However, there is still no agreed-on model of how people balance costs, rewards, and information when planning under uncertainty in the real world, where costs and rewards derive from perceived physical quantities. To move toward a study of planning in a more naturalistic context, we present a novel spatial Maze Search Task (MST) where the costs and rewards are physically situated as distances and locations. We used this task to evaluate and contrast multiple distinct computational models of planning. The models included optimal expected utility planning, a family of planners that approximate optimal planning, and simple myopic heuristics. We found that in contrast to myopic heuristics or the optimal planning, people's behavior is best explained by approximate planners, in which values are estimated by the interactions between perception and cognition.

[01] Department of Brain and Cognitive Sciences, MIT
[2] Department of Mechanical Engineering, MIT
[3] Department of Psychology, Harvard University
[*] Corresponding author, mkryven@mit.edu
[†] Indicates equal contribution

Author Summary

We present a computational study of spatial planning under uncertainty using a novel Maze Search Task (MST), in which people search mazes for probabilistically hidden rewards. The MST is designed to resemble real-life planning where costs and rewards are physically situated as distances and locations. We found that people's behavior is best explained by approximate planners, as opposed to both myopic heuristics or the optimal planning.

1 Introduction

People make plans every day: working out a new route, playing a game, thinking through a possible conversation. Despite the ubiquity of planning, *how* people plan is an active topic of research in many different fields. Planning involves making sequences of choices, where the possible actions that are available at each step depend on the outcome of the previous step. This process has been formalized as navigating a decision tree, which starts at an initial 'root' state, and continues until it reaches a 'leaf' that meets the goal criteria [1, 2, 3, 4, 5]. For any situation beyond trivial toy problems, the growing complexity of a branching decision tree makes planning computationally costly. And yet, people daily face situations that require planning, and handle them remarkably well.

A prime example of a daily planning task that people are quite adept at is spatial planning. But while an enormous amount of work has studied and modeled planning in different domains (as we review below), to our knowledge relatively few studies considered detailed computational models of spatial planning in naturalistic contexts, under uncertainty. In this work, we take a step toward understanding human spatial planning under uncertainty using a novel Maze Search Task (MST), designed to resemble a natural environment where the costs are distances and rewards are spatial locations. A version of MST was previously used to study how people evaluate the goodness of plans made by others [6], but a detailed computational account of how people themselves plan in the MST was not addressed. We use the MST to explore in detail a family of computational planning models that approximate expected utilities by integrating perceptual transformations, such as numerosity [7] and probability weighting [8]. We contrast these approximate planning models both with an optimal Expected Utility model, and with a family of intuitive myopic heuristics, that could in principle be used to search an environment without planning ahead.

In what follows, we briefly review recent relevant empirical work that includes computational models of human planning. Based on this literature, we outline the relevant take-aways for the models we will consider in this work, including perceptual transformations and constraints that may influence how people plan in real-world.

First, while we mentioned that relatively little of the work on planning has focused on computational models of naturalistic spatial planning, certainly research has been done on it. For example, a small study modeled route planning within a city neighborhood as optimally solved by Breadth First Search [9]. However, larger empirical studies find that people rarely take shortest routes [10, 11], indicating that people likely use cognitive approximations.

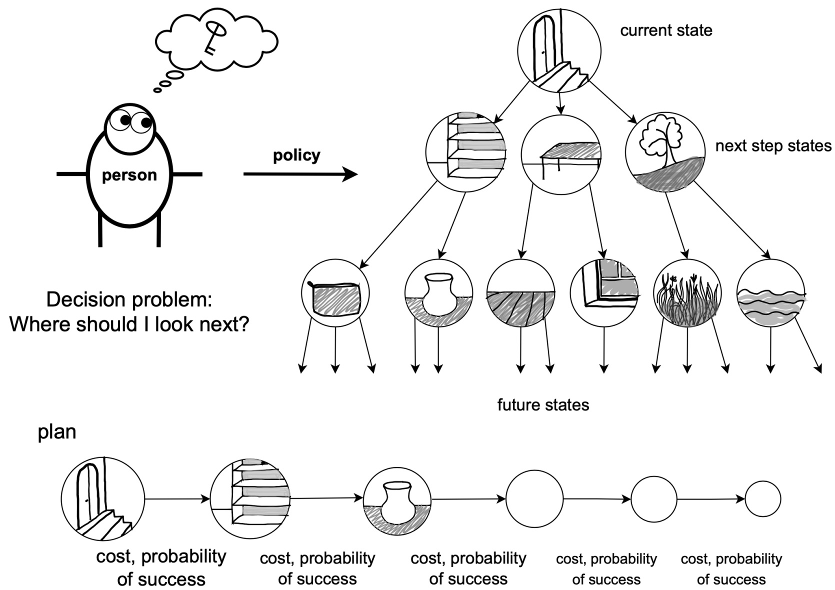


Figure 1: The decision problem facing an agent is to take actions that maximize long-run rewards. The agent can solve this problem by planning a path through a decision tree that recursively minimizes costs, while taking into account the probabilities of success. Our task captures this process in a spatial setting, by mapping costs to steps taken to make an observation, and probabilities of success to the relative size of an observed area.

Given the time constraints of in-lab studies, most prior work has focused on planning short action sequences in non-spatial situations, such as sequences of 2 or 3 actions, in rigorously designed yet simple experimental paradigms [3, 12, 5, 13, 4]. Several studies of games, such as Chess and Tic-tac-toe, have modeled planning over longer horizons in contexts with rich behavioral variation [14, 15, 16]. However, game state utilities in these tasks derive from task-specific heuristic game-board assessments that evaluate the player's position based on various features of the board, making it hard to generalize the results beyond those specific games.

Several studies of planning suggest that people manage cognitive demands by limiting their planning horizon, which has implications for models of spatial planning. For example, when people are asked to connect moving consecutive disks to maximize their total volume, they do so in a way best explained as considering all possible paths up to a certain limited depth, in contrast to simple myopic heuristics, such as moving toward the largest disk [2]. In a different popular example, the Tower of London task requires stacking disks in a certain order, using the fewest moves. People doing this task increasingly deviate from the optimal number of moves as the depth of planning required to optimally solve the task increases, which again suggested that people use depth-limited planning [17]. In yet another example, several studies examined how people learn an implied decision tree in a three-stage bandit task, which involves making sequential decisions to learn and exploit a hidden task structure. People were found to prune their implied state-space during the choice stage in response to rising cognitive demands, suggesting that they relied on a flexible depth-limited representation to control cognitive costs [3, 12, 5]. Other examples of a limited yet flexible depth of search are found in board games, such as Chess [15, 18, 16], and Four in a Row [14], where increasing expertise corresponds to greater depth in search algorithms, [15, 18, 16, 14].

The results above strongly suggest that human-like planning models should include limited yet flexible planning horizons. These can be implemented formally in several ways. A common approach limits planning depth by probabilisticaly pruning the planning tree at a each node [19, 3], which is equivalent to a discount rate (see [3] for a detailed discussion). Other implementations constrain the number of nodes explored by an algorithm, such as Monte Carlo Tree Search [14], or plan up to a certain fixed depth [2]. In this work, we implemented a limited planning horizon by discounting, which integrates naturally with our modeling approach (see below).

Beyond limited horizons, everyday planning often takes place in conditions of uncertainty, which may require gathering more information. Information gathering can include learning about costs and rewards [4], asking questions [20], or observing an unknown environment [21]. Prior work tended to model information seeking behavior by myopic heuristics: choosing observations one at a time, rather than by planning ahead [21, 20]. Though, in a recent exception, people's information gathering about three-step plans with deterministic state transitions was best explained by optimal planning [13, 4]. In our task uncertainty arises from the partially observed environment: planning a search in MST means deciding in which order to observe the environment, similar to [21].

In addition to limited planning horizons and uncertainty, people's plans rest on subjective utilities, which estimate the expected value of prospective states. Subjective utility functions can take different forms, examined in decision theories, such as Prospect Theory [22, 23, 8]. In real-world spatial planning, however, value likely derives from perceived physical quantities, such as distance, number, and area. Given this, we hypothesized that real-world planning models may need to account for perceptual transformations or psychological constraints over these quantities. For example, the perception of numerosity is known to follow a non-linear relationship between the actual and perceived number [7], sensory stimuli are generally perceived on logarithmic scale [24], and probabilities tend to be represented by a non-linear mapping that overestimates small and underestimate large probabilities. And while probability perception has been widely explored [25, 22, 23], probability models have not been integrated for the most part into models that evaluate human planning. In our modeling approach we will consider previously found transformations from physical to perceived quantities, and examine if they result in more human-like planning models.

When it comes to modeling people's spatial planning, the previous computational and empirical work on planning paints the following picture: We need to build and evaluate approximate planning models that have limited planning horizons, gather information to resolve uncertainty, and estimate subjective utilities by psychological transformations of perceived physical quantities. We also need to compare such models to simpler alternatives that do not involve planning ahead, which we broadly construe as myopic heuristics. Finally, we need a rich physically grounded environment to contrast the different models of planning as well as heuristics, in naturalistic settings. In the following sections, we detail the novel task/environment we used to probe and evaluate people's spatial planning, as well the ideal planning models, the approximate planning models, and the heuristics we compared people to.

2 Experimental Methods

2.1 Maze Search Task

The objective of a participant in the Maze Search Task (MST) is to navigate a series of partially observable, two-dimensional grid-world mazes while

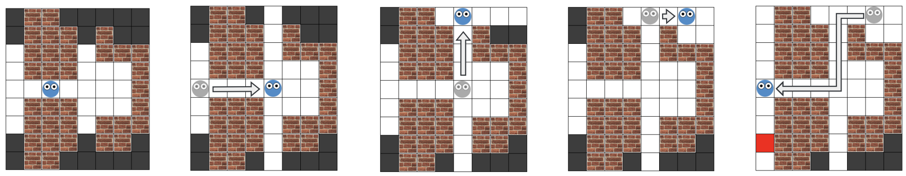


Figure 2: An example path in the Maze Search Task (MST). Black tiles are not-yet-observed areas, which hide an exit (red square). This maze has six 'rooms', groups of black tiles that are revealed all at once. Revealing tiles can be done in any order, but players are incentivized to plan their path so as to reach a hidden exit in fewer steps.

minimizing the distance traveled to reach a hidden exit. Each maze consists of walls, corridors, and rooms – clusters of hidden tiles that can be observed together. One of the tiles contains an exit, which remains hidden until the room containing it is observed. Participants are told that each of the black tiles is *equally likely* to hide the exit, and are instructed to reach the exit in as few steps as possible. The exit becomes visible as a red grid tile once its location is observed. Figure 2 shows a maze with the player's location indicated by the face avatar, and the player's path indicated by arrows. An experiment demo is available at `https://marta-kryven.github.io/mst.html`. The complete task instructions are available in Supplemental Materials, Section 5.5.

The player in the MST can move to any adjacent grid tiles that are not blocked by walls in the four cardinal directions (up, down, left, right), and reveal the black unobserved tiles by bringing them into the avatar's line of sight. Upon reaching the exit the player is moved to the next maze. If the current maze is the last trial in the experiment, then upon reaching the exit the experiment ends.

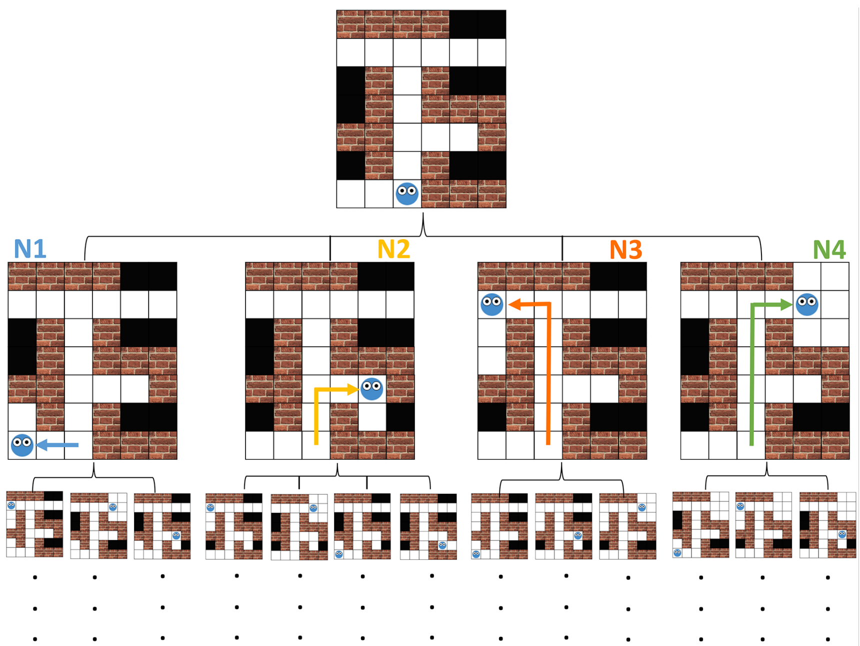


Figure 3: Decision-tree for a maze with four rooms (hidden tiles that are revealed together). The tree abstracts away from specific moves like 'up' and 'left' and considers more general actions like which area to uncover next. The root of the decision-tree corresponds to the player's starting location. The four nodes accessible from the root indicate the possible observations that can be made next, followed by the observations that can follow each of those, and so on.

2.2 Computational Models

To plan a route through a maze, we first define a decision-tree detailing all possible orderings in which the rooms in a maze can be revealed. Each node in this tree corresponds to a unique combination of a location (x, y) from which an observation is made, and the area being observed. If a room can be observed from two different locations, then this room will be represented by two nodes in the decision tree, each corresponding to a unique location from which an observation is made. An example of a maze along with a partial decision tree for this maze is shown in Fig. 3. The starting location corresponds to the root of the decision-tree. The leaves correspond to all possible unique ordering of observations through which the entire maze can be revealed.

Planning a path through a maze entails computing values of decision tree nodes V_k, and mapping the node values to probabilities of choosing them. We use a softmax mapping, a standard method of modeling expressed preferences in decision-making [3, 26, 27, 28, 29]:

$$\sigma(\mathbf{V})_k = \frac{\exp(-V_k/\tau)}{\sum_j \exp(-V_j/\tau)}. \tag{1}$$

Here, τ is the temperature parameter that controls the strength of the softmax mapping, k indexes tree nodes, and j indexes siblings of node N_k. The negative sign in front of the value ensures that shorter paths result in higher probabilities. As $\tau \to 0$, the agent will always choose the shortest expected path. As tau increases, the agent will behave in a more noisy way, and as $\tau \to \infty$ agents will chose actions at random.

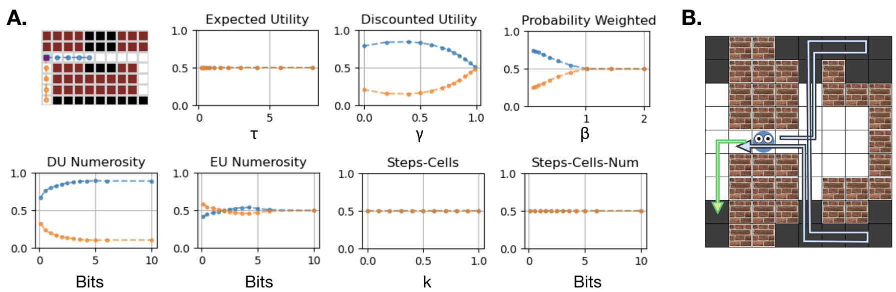


Figure 4: Models predicting different paths. **A.** A maze in which the next directions are Right and Down, along with model-parameter graphs. Each graph is a different model. The X axes show model parameters. The y-axis shows probabilities assigned by different models to the actions Right and Down. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment. The Expected Utility (EU) model is indifferent between the two directions, since the probabilities of finding the exit in each room are equal. The heuristic models (Step-Cells and Step-Cells-Num) are also indifferent, as the distances to the two rooms are the same (4 steps), as are the number of cells observed in each room (9 cells). The Discounted Utility (DU) model predicts that players should be more likely to go Right first. **B.** An example where models predict different paths, showing the most likely paths of the EU model (green) and the DU model (blue), assuming the exit is in the bottom left. The models diverge because the DU model discounts the possibility of having to backtrack. The Steps-Cells heuristic model is indifferent between initial Left and Right choices.

2.2.1 Planning ahead

We first define the optimal Expected Utility value function, which computes the shortest paths to the exit. We will use the optimal model to derive other, approximate planners.

Formally, the **Expected Utility (EU)** of a node N_i is given by the expected number of steps to the exit if this node is chosen, assuming all subsequent choices are optimal as well:

$$V_{EU}(N_i) = s_i + p_i e_i + (1 - p_i) \min_{c_j \in C(N_i)} V(c_j). \tag{2}$$

Where p_i is the probability that the exit is found at N_i. Assuming the exit is equally likely to be in any of the black tiles, p_i is the ratio of the number of tiles observed at N_i to the total number of unobserved tiles remaining in the maze. In a general case this probability could be arbitrary. Further, s_i is the number of steps to reach N_i from its parent node in the tree – that is, the number of steps between the previous observation, and the observation at N_i; e_i is the expected number of steps to the exit from N_i, if the exit is observed at N_i – in other words, the average number of steps to a cell revealed in N_i. Lastly, $C(N_i)$ is the set of children of N_i in the planning tree – that is, a set of observations that can be made next, after reaching N_i.

The Discounted Utility (DU) model modifies the EU model by discounting the values of future nodes, by a rate of $\gamma \in [0, 1]$:

$$V_{DU}(N_i) = s_i + p_i e_i + \gamma(1 - p_i) \min_{c_j \in C(N_i)} V(c_j) \tag{3}$$

We use discounting to implement a limited planning horizon. Discounting is computationally equivalent to probabilistically pruning computations at each node with a certain stopping probability [3].

The Probability Weighed Utility (PW) model modifies the EU model by transforming probabilities, using a weighting function that overestimates small probabilities and underestimates large probabilities. We will use the form $\pi(p) = exp(-|\ln(p)|^{\beta})$ [8]. While probability weighting is widely used to model probability perception in monetary gambles [25, 22, 23] it is not commonly evaluated in the context of planning.

We define the PW value function as follows:

$$V_{PW}(N_i) = s_i + \pi(p_i)e_i + \pi(1 - p_i) \min_{C_j \in C(N_i)} V(C_j) \tag{4}$$

The parameter $\beta \in [0, 1]$ controls probability weighting. When $\beta < 1$ smaller probabilities are overestimated and larger probabilities are underestimated. If $\beta = 0$ all probabilities have the same uniform value, and $\beta = 1$ is equivalent to the optimal Expected Utility in which all probabilities are equally weighted.

The Probability Weighting and Discounting (PW-DU) model combines probability weighting and discounting in a model with three free parameters—τ, γ, and β (corresponding to the softmax strength, the discount rate, and the probability weighting respectively).

$$V_{Comb}(N_i) = s_i + \pi(p_i)e_i + \gamma\pi(1 - p_i) \min_{C_j \in C(N_i)} V(C_j) \tag{5}$$

Expected Utility with Numerosity psychophysics (EU-Num) When people reason about varying amounts of hidden tiles, they could in principle count the tiles exactly. At the same time, people could instead roughly estimate the number of tiles, causing systematic deviations [30, 7]. We use a recent information-theoretic numerosity model to account for the latter possibility [7]. The information-theoretic numerosity model has one free parameter B – the number of bits processed by the perceptual system to estimate numbers. People who count the tiles can be modeled with a large B, and people who guess at a glance can be modeled with a small B (we used $B \in [0.1, 10]$).

Discounted Utility with Numerosity psychophysics (DU-Num) The DU-Num incorporates information-theoretic numerosity into the DU model along the lines above.

2.2.2 Monte Carlo Tree Search Sampling model (Sampling)

The planning models considered so far accurately represent the planning tree, but assume that the utilities of its nodes are approximated in some way. Another way to make approximate plans is to construct a partial planning tree. This process can be formalized by the Monte Carlo Tree Search (MCTS), an algorithmic framework for simulating multiple possible outcomes while keeping track of them in a tree, and choosing the best one based on the simulation results. The accuracy of the planning tree in MCTS is controlled by two free parameters – the computational budget that controls how many nodes are sampled, and exploration that controls how greedy or stochastic the process is [31]. Implementations of MCTS have been previously used to model games such as chess, go, and four-in-a-row [14]. See Supplemental Materials for our implementation of an MCTS algorithm for MST.

2.2.3 Myopic Heuristics

A large body of decision-making literature has focused on heuristic solutions to problems, where a heuristic derives from simple rules or features, such as the size of an area [2, 32, 20]. While it is impossible to enumerate all the possible heuristics that could be invented in response to a given situation, we considered 7 one-step heuristics based on prominent features of MST, to rule out simple interpretations of the task.

Steps heuristic (Steps) The value of a node is taken to be the number of steps to the node from its parent: $V(N_i) = s_i$.

Cells heuristic (Cells) The value of a node is taken to be the number of tiles observed at that node: $V(N_i) = -cells_i$. We use a negative sign because $V(N_i)$ is treated as a cost, and revealing more cells results in a *smaller* expected cost for finding the exit.

Steps-Cells heuristic (S-C) A combination of the two heuristics above. The value of a node is a combination of the steps to get to it and the revealed cells, but without planning more than one step ahead: $V(N_i) = ks_i - (1-k)cells_i$, where $k \in [0, 1]$ is a free parameter.

Adding information-theoretic numerosity to each of the three heuristics above results in three more modified heuristic models – Steps with numerosity

Steps-Num, Cells with numerosity **Cells-Num**, Steps-Cells with numerosity **Steps-Cells-Num**.

Random Lastly, we consider a random policy, in which the value of any node is the same, $V(N_i) = 1$.

The models described above can make different predictions about how a maze should be traversed. For example, the PW model can underestimate how likely a large room is to hide the exit, and overestimate how likely is the exit to be in a smaller room. The Steps heuristic can predict a preference for closer rooms, regardless of their shape or size. The Cells heuristic can explain a preference to reveal larger rooms regardless of distance. The Steps-Cells heuristic will be indifferent between choosing one of two equidistant rooms with the same number of tiles. The DU model can be insensitive to incurring a path with potentially large detours occurring later in the path. Figure 4 shows examples of such differences. See Supplemental Materials 5.2 for more details about model behavior. In other words, the space of models we consider above spans a wide range of possible behaviors, due to very different commitments they make about how one plans through a spatial environment.

2.3 Data analysis methods

The models were fitted at the individual level by considering all decisions made by an individual during the experiment, which are treated as independent of each other. A 'decision' in our case refers to choosing one of two or more child nodes, meaning choosing which one of the unobserved rooms to search next. We fit our models to individuals, and obtained out-of-sample predictions using fivefold cross-validation. We measured the *model performance* as the total test log likelihood (LL) of the model across all five test folds. This metric accounts for the flexibility of the different models without parameter counting, in contrast to AIC and BIC which can be a poor measure of flexibility [33]. Differences in this cross-validated LL (ΔLL) can be interpreted similarly to differences in AIC: $\Delta\text{LL} = 1$ is roughly equivalent to $\Delta\text{AIC} = 2$. The Supplemental Materials also show model performance analyzed using Monte-Carlo cross-validation. We analyzed *individual differences in planning* by comparing the likelihood of the best fitting planning model (one of EU, EU-Num, PW, DU, DU-Num, PW-DU, Sampling) and best fitting heuristic (one of Steps, Cells, Steps-Cells, Random, Steps-Num, Cells-Num, Steps-Cells-Num) for each individual.

We also considered a model's ability to explain *behavior aggregated across participants*. To do this, we computed the correlations between model predictions, and the probability that people will make the corresponding choice in the following way.

We first computed the probabilities of participants making each choice at the initial decision point in each maze, aggregated across individuals. Although individuals then proceed to search each maze in different ways (with each person potentially facing a unique set of choices), the initial decisions are shared by all participants in the experiment. We then correlated these probabilities with each model's predictions, where the models are parameterized with the mean parameters of the experimental population. While we consider the first choice to be the most indicative and carefully controlled measure, in Supplemental Materials we also show correlations computed using all decisions in the experiment that were visited by at least 20% of the participants.

2.4 Experimental Procedure

The experiment was conducted in a web browser, using a JavaScript and PHP interface developed by the authors. Participants first read a consent page and a short description of the Maze Search Task. Following consent, participants read a detailed description of the task, completed maze navigation practice, and answered an instruction quiz. The quiz included questions about the objectives of the task, task controls, and the line-of-sight mechanism by which the mazes are revealed. Participants could not proceed to the experiment until they submitted the correct answers to the quiz. On each trial a participant was placed at the starting position, and navigated by clicking on one of the adjacent tiles, until the exit was reached. The starting locations in each maze were predetermined, and the exit locations were randomly chosen for each maze at the time of design. After completing the experiment, each participant answered a demographic questionnaire, and provided a free-form description of search strategies they used in the experiment. Participants were paid a US minimal wage and received a performance-based bonus, configured so that on average 70% of individuals receive a bonus. The study was approved by our institutional IRB board. The same procedure was used for both experiments. Please see Supplemental Materials 5.5 for an explanation of the bonus scheme and task instructions.

3 Experiments

3.1 Experiment 1

In the first experiment we aimed to test whether people's planning follows optimal expected utility. The experiment included 40 mazes, presented in random order. Of the mazes, 30 consisted of two rooms designed to trade off steps and observations, such that the bigger of the two rooms also required more steps to reach than the smaller room. The two-room mazes correspond to simple decision trees with two leafs, similar to the experimental paradigms that involve choosing between two gambles, which are traditionally used to study subjective utilities [22, 23]. Another 10 mazes consisted of 3-4 rooms. See Supplemental Materials 5.6 for full set of mazes used in the experiment.

We recruited 120 US participants on Amazon Mechanical Turk, of which 4 were excluded for failing to answer instruction quiz correctly in two attempts. In total 116 participants (56 female, 60 male, $M(age) = 39$, $SD(age) = 12$) were included in the analysis. Rerunning the analysis with all participants included did not change the results. On average the experiment took 14 minutes to complete, with people making on average 51.5 decisions during this time.

Given the lack of comparable studies there was no immediate way to establish the requisite participant number for a correct power analysis. Pilot studies showed that 100 participants were more than sufficient to show variability between people, and so 120 participants was erring on the side of caution.

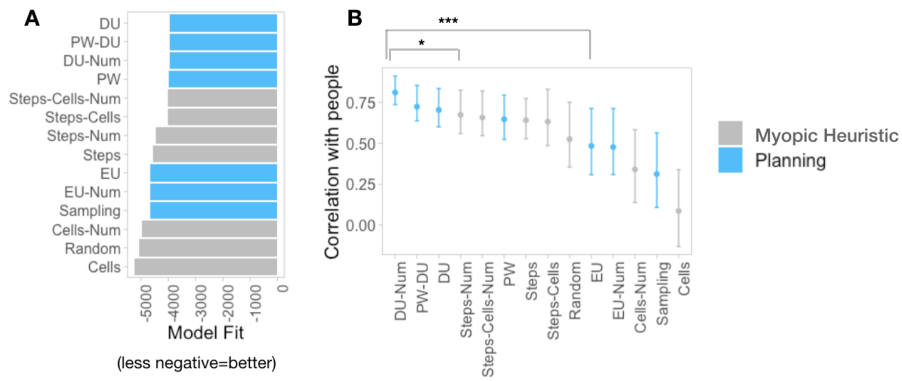


Figure 5: **Experiment 1, results**. **A.** Model performances, measured as the total log likelihood of each model across all five folds. Shorter bars indicate better fit to human behavior. **B.** Bootstrapped correlations of choice probabilities aggregated across participants with each model's predictions. Error bars indicate 95% confidence intervals.

3.1.1 Results

People do not plan according to optimal Expected Utility. Model performance is shown in Figure 5 A, with the models ordered by their likelihood. The difference in log-likelihood between EU and the best most likely planner DU is $\Delta LL = 764$. Figure 5 B shows the bootstrapped correlations between the probabilities of decisions made by people and the models. The correlation of the best-performing DU-Num model with people is $r = .81(95CI[.74, .91])$, and the correlation of EU is $r = .49(95CI[.31, .71])$. These correlations are significantly different, with bootstrapped difference between their means of [.1.5], indicating that the DU-Num model predicts the aggregate population better than the optimal EU model.

Overall, we find that while several computational models can reasonably predict human planning, all of them plan ahead in a way that deviates from the model based on optimal expected utility.

Evidence in favor of planning over heuristics. We also found that planning models are better at explaining people's behavior, compared to myopic heuristics. The difference in log-likelihood between the most likely planner DU and the most likely heuristic Steps-Cells-Num is $\Delta LL = 47$. Bootstrapped correlations between models and people's choices in Figure 5B show that the 95% CI of bootstrapped difference between correlations of DU-Num (a planner with the highest correlation) and Steps-Num (a heuristic with the highest correlation) is significant, [.01.2], suggesting that people's choices are best explained by models that plan. See Supplementary Information for more analysis, including variability between individuals.

While the difference between approximate planning models and heuristics is significant, the differences are small, due to the experimental design dominated by simple two room mazes. In the next experiment we aimed to further differentiate between planning and heuristics, by focusing maze design on decisions that elicit planning.

3.2 Experiment 2

In the second experiment, we aimed to differentiate between models that plan ahead and myopic heuristics. We followed the experimental procedure from Experiment 1 with a new set of 23 mazes containing between 4 to 10 rooms. The mazes were designed so that in the initial decision in each maze heuristics could not distinguish between the available choices, but planning models could. Figure 4 shows two examples of this design. Since each direction of travel during an initial decisions leads to a room of the same size after

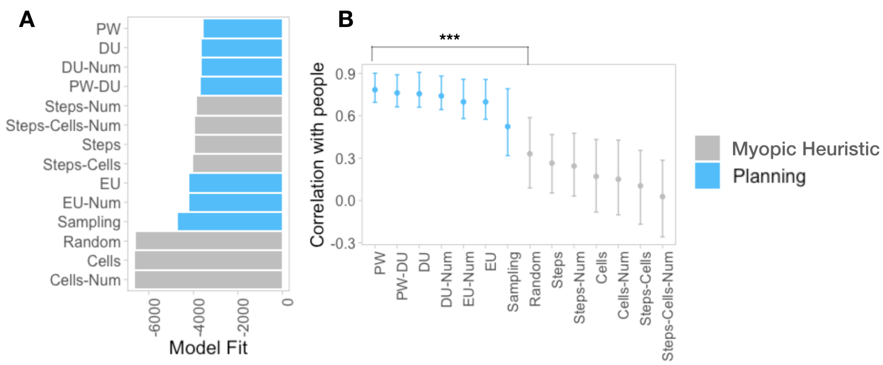


Figure 6: Experiment 2 Results. **A.** Model performances, measured as the total log likelihood of each model across all five folds. Shorter bars indicate better fit to human behavior. **B.** Bootstrapped correlations of choices aggregated across participants with each model's predictions. Error bars indicate 95% confidence intervals.

taking the same number of steps, neither direction is preferable to a heuristic. However, the areas of a maze that become accessible in subsequent decisions differ between directions, meaning that planning models prefer one of the directions over the others. The full set of mazes used in the experiment is available in Supplemental Materials 5.7.

We recruited 107 US participants on Amazon Mechanical Turk, of which 7 were excluded for failing to answer instruction quiz correctly in two attempts. In total 100 participants (35 female, 80 male, $M(age) = 33$, $SD(age) = 9$) were included in the analysis. Rerunning the analysis with all 107 participants included did not change the results. The experiment took on average 20 minutes to complete, with people making on average 70 decisions during the experiment.

3.2.1 Results

Evidence in favor of planning over heuristics. Overall, we found that planning models are better at explaining people's behavior compared to myopic heuristics. The difference in likelihoods between PW and the best-fitting Steps-Num heuristic is $\Delta LL = 262$. The 95% CI of bootstrapped difference between correlations of PW (a planner with the highest correlation) and Steps (a heuristic with the highest correlation) is [0.20.5], as shown in Figure 5B, indicating that planning models predict the aggregate population better than myopic heuristics. See Supplementary Information for more analysis, including variability between individuals.

4 Discussion

We examined how people plan under uncertainty in a spatial setting using a novel Maze Search Task (MST), which requires people to navigate partially observable mazes in search of a randomly placed exit. We evaluated a family of computational models that plan ahead, along with a family of myopic heuristics that choose the next observation one step at a time. In addition to planners parametrized with a limited planning horizon we evaluated approximate planners that incorporate perception of numerosity and probability, which were not previously evaluated in the context of planning. We found that people's decisions in the MST were best explained by models that plan, as opposed to myopic heuristics or the optimal planning, in a way that can be explained by subjective utilities arising from interactions between cognition and perceptual constraints. We do not argue that one of the planning models is definitively superior at predicting how people plan, as individual differences in the MST suggest that multiple planning approximations may be available to people at the same time (this point is similar to previously reported differences in decision-making [19, 34]).

Consistent with previous studies [3, 12, 2], we found that modifying the optimal Expected Utility by a discount rate explains human behavior better than Expected Utility alone. However, we found that discounting is neither necessary, nor superior to other forms of planning approximations in MST. Only for 30% of individuals in the second Experiment the best-fitting model included discounting (See Supplement, Figure 18). While we examine planning in decision-trees with maximal depth of 10 (corresponding to mazes with up to 10 rooms), future studies could investigate how discounting interacts with perceptual approximations as the computational complexity of the problem increases.

In our analysis, all decisions are treated as Markovian, meaning each decision is independent and regardless of whether a certain maze layout is encountered as a new maze, or as a part of a partially explored maze. While this assumption is common in computational models of planning, future studies could test whether in practice individuals make consistent choices after partially exploring a maze, compared to when an identical partially revealed maze is presented as a new trial.

The planners built in this work do not purport to describe all there is to spatial planning, or to exhaust all possible ways in which people could make choices in MST. Our modeling goal in this work was to contrast a set of naturalistic planning models motivated by prior literature against myopic heuristics motivated by literature as well as by participant feedback received during pilots. While it is possible to imagine more sophisticated heuristics, for example based on symmetries [35] or hierarchical representations [36, 37], such heuristics would serve to reduce complexity of the task rather than support myopic decision-making.

While we used analytical models for methodological rigor, future studies can investigate algorithmic models that sample a small number of paths, which could be particularly well suited for large environments. Future studies can also investigate the stability and generalization of individual planning strategies between tasks, and over time. Do people learn to plan by learning a library of task-specific strategies (e.g. Chess strategies [15]), or by learning abstract cognitive mechanisms for planning [38, 13], for example by learning to plan further ahead? Another interesting question is to examine whether people mentally represent probabilities in the same way across spatial and non-spatial contexts, such as, investigating whether probability weighting generalizes between MST and monetary gambles studied by Prospect Theory [23].

Our work takes a step toward computationally understanding and modeling an important, real-life domain of planning: how humans plan in spatial multi-step contexts. Our results have implications for the study of human cognition, as well as for building an AI that can interpret human planning to infer goals, offer assistance, and support human-AI collaboration. Beyond our theoretical and empirical contributions, we hope our MST methodology can become a sandbox for exploring a larger variety of planning environments, cognitive models and algorithms.

Acknowledgments This work is supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

References

[1] John von Neumann, Oskar Morgenstern, et al. Theory of games and economic behavior. 1972.

[2] Joseph Snider, Dongpyo Lee, Howard Poizner, and Sergei Gepshtein. Prospective optimization with limited resources. *PLoS computational biology*, 11(9):e1004501, 2015.

[3] Quentin JM Huys, Níall Lally, Paul Faulkner, Neir Eshel, Erich Seifritz, Samuel J Gershman, Peter Dayan, and Jonathan P Roiser. Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, 112(10):3098–3103, 2015.

[4] Frederick Callaway, Bas van Opheusden, Sayan Gul, Priyam Das, Paul M Krueger, Thomas L Griffiths, and Falk Lieder. Rational use of cognitive resources in human planning. *Nature Human Behaviour*, 6(8):1112–1125, 2022.

[5] Mehdi Keramati, Peter Smittenaar, Raymond J Dolan, and Peter Dayan. Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. *Proceedings of the National Academy of Sciences*, 113(45):12868–12873, 2016.

[6] Marta Kryven, Tomer D Ullman, William Cowan, and Joshua B Tenenbaum. Plans or outcomes: How do we attribute intelligence to others? *Cognitive Science*, 45(9):e13041, 2021.

[7] Samuel J Cheyette and Steven T Piantadosi. A unified account of numerosity perception. *Nature human behaviour*, 4(12):1265–1272, 2020.

[8] Drazen Prelec. The probability weighting function. *Econometrica*, pages 497–527, 1998.

[9] Amir-Homayoun Javadi, Beatrix Emo, Lorelei R Howard, Fiona E Zisch, Yichao Yu, Rebecca Knight, Joao Pinelo Silva, and Hugo J Spiers. Hippocampal and prefrontal processing of network topology to simulate the future. *Nature communications*, 8(1):1–11, 2017.

[10] Shanjiang Zhu and David Levinson. Do people use the shortest path? an empirical test of wardrop's first principle. *PloS one*, 10(8):e0134322, 2015.

[11] Christian Bongiorno, Yulun Zhou, Marta Kryven, David Theurel, Alessandro Rizzo, Paolo Santi, Joshua Tenenbaum, and Carlo Ratti. Vector-based pedestrian navigation in cities. *Nature Computational Science*, 1(10):678–685, 2021.

[12] Quentin JM Huys, Neir Eshel, Elizabeth O'Nions, Luke Sheridan, Peter Dayan, and Jonathan P Roiser. Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, 8(3):e1002410, 2012.

[13] Yash Raj Jain, Frederick Callaway, and Falk Lieder. Measuring how people learn how to plan. In *CogSci*, pages 1956–1962, 2019.

[14] Bas van Opheusden, Ionatan Kuperwajs, Gianni Galbiati, Zahy Bnaya, Yunqi Li, and Wei Ji Ma. Expertise increases planning depth in human gameplay. *Nature*, pages 1000–1005, 2023.

[15] Diogo R Ferreira. The impact of the search depth on chess playing strength. *ICGA journal*, 36(2):67–80, 2013.

[16] Adriaan D De Groot. Thought and choice in chess. In *Thought and Choice in Chess*. De Gruyter Mouton, 2008.

[17] Josef M Unterrainer, Benjamin Rahm, Christoph P Kaller, Rainer Leonhart, K Quiske, K Hoppe-Seyler, C Meier, C Müller, and Ulrike Halsband. Planning abilities and the tower of london: is this task measuring a discrete cognitive

function? *Journal of clinical and experimental neuropsychology*, 26(6):846–856, 2004.

[18] Dennis H Holding. *The psychology of chess skill.* Routledge, 1985.

[19] Yash Raj Jain, Frederick Callaway, Thomas L Griffiths, Peter Dayan, Ruiqi He, Paul M Krueger, and Falk Lieder. A computational process-tracing method for measuring people's planning strategies and how they change over time. *Behavior Research Methods*, 55(4):2037–2079, 2023.

[20] Björn Meder, Jonathan D Nelson, Matt Jones, and Azzurra Ruggeri. Stepwise versus globally optimal search in children and adults. *Cognition*, 191:103965, 2019.

[21] Jiri Najemnik and Wilson S Geisler. Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–391, 2005.

[22] Kai Ruggeri, Sonia Alí, Mari Louise Berge, Giulia Bertoldo, Ludvig D Bjørndal, Anna Cortijos-Bernabeu, Clair Davison, Emir Demić, Celia Esteban-Serna, Maja Friedemann, et al. Replicating patterns of prospect theory for decision under risk. *Nature human behaviour*, 4(6):622–633, 2020.

[23] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323, 1992.

[24] Gustav Theodor Fechner. *Elements of Psychophysics.* Howes, D H; Boring, E G (eds.), 1860.

[25] Joshua C Peterson, David D Bourgin, Mayank Agrawal, Daniel Reichman, and Thomas L Griffiths. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science*, 372(6547):1209–1214, 2021.

[26] Nathaniel D Daw, Samuel J Gershman, Ben Seymour, Peter Dayan, and Raymond J Dolan. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215, 2011.

[27] Maurice Allais. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, pages 503–546, 1953.

[28] R Duncan Luce. *Individual choice behavior: A theoretical analysis.* Courier Corporation, 2012.

[29] Kenway Louie, Lauren E Grattan, and Paul W Glimcher. Reward value-based gain control: divisive normalization in parietal cortex. *Journal of Neuroscience*, 31(29):10627–10639, 2011.

[30] William Stanley Jevons. *The theory of political economy.* Macmillan and Company, 1879.

[31] Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, 4(1):1–43, 2012.

[32] Jeremy N Bailenson, Michael S Shum, and David H Uttal. The initial segment strategy: A heuristic for route selection. *Memory & Cognition*, 28(2):306–318, 2000.

[33] Steven T Piantadosi. One parameter is always enough. *AIP Advances*, 8(9):095118, 2018.

[34] Joshua Peterson, Marina Mancoridis, and Tom Griffiths. To each their own theory: Exploring the limits of individual differences in decisions under risk. In

Proceedings of the Annual Meeting of the Cognitive Science Society, volume 45, 2023.

[35] Sugandha Sharma, Aidan Curtis, Marta Kryven, Josh Tenenbaum, and Ila Fiete. Map induction: Compositional spatial submap learning for efficient exploration in novel environments. *arXiv preprint arXiv:2110.12301*, 2021.

[36] Momchil S Tomov, Samyukta Yagati, Agni Kumar, Wanqian Yang, and Samuel J Gershman. Discovery of hierarchical representations for efficient planning. *PLoS computational biology*, 16(4):e1007594, 2020.

[37] Carlos G Correa, Mark K Ho, Frederick Callaway, Nathaniel D Daw, and Thomas L Griffiths. Humans decompose tasks by trading off utility and computational cost. *PLOS Computational Biology*, 19(6):e1011087, 2023.

[38] Marcos Economides, Zeb Kurth-Nelson, Annika Lübbert, Marc Guitart-Masip, and Raymond J Dolan. Model-based reasoning in humans becomes automatic with training. *PLoS computational biology*, 11(9):e1004463, 2015.

[39] Mark K Ho, David Abel, Carlos G Correa, Michael L Littman, Jonathan D Cohen, and Thomas L Griffiths. People construct simplified mental representations to plan. *Nature*, 606(7912):129–136, 2022.

5 Supplemental Materials

5.1 Planning state-space

We based our computational approach on choosing a path through a decision tree that maps states to specific observations within a maze. This design is supported by an empirical evidence that people take direct routes between observations, even through the grid-world layout of MST in principle allows players to take indirect routes, or even indefinitely move between any adjacent empty tiles without making any observations. People's tendency to take direct routes between observations strongly suggests that people use efficient problem representations, in line with previous work [39]. Our decision tree model is further supported by the empirical distribution of human decision times in different types of tiles inside a maze, as shown in Figure 7. The figure shows that people move quickly when travelling between observations (in Corridors) and take longer to make a move whenever new hidden tiles are revealed (Decision). The longest decision time occurrs at the initial starting location (Start), where people may study the map and plan their path before moving.

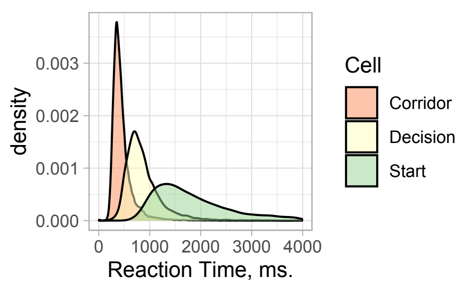


Figure 7: Distribution of decision times in milliseconds as humans click on maze tiles, plotted by location type: Start - at the start of the trial; Corridor - moving between observations; Decision - at observation locations.

5.2 Planning Models

Probability Weighted Utility

We used probability weighting of the form $p = exp(-1(-\log(p))^{\beta})$, the shape this function takes for different β is shown in Fig.8. The original Prospect Theory sets $\beta \in [0, 1]$. Here we also consider $\beta \in [0, 2]$ where values of $\beta \in (1, 2]$ imply overweighting large probabilities.

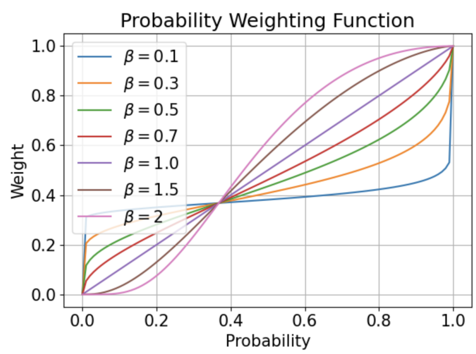


Figure 8: The probability weighting function for different β.

Numerosity perception in Maze Search Task

The information-theoretic numerosity model assumes that the systematic deviation in the perceived number is due to processing limited bits of information, where the prior distribution over observed numbers favors small quantities. Figure9 shows how the number of tiles perceived by people is predicted to change with the number of bits of information processed.

Using the perceived number of tiles to estimate probabilities of finding exit in a room likewise results in subjective probability. The subjective probability function for different B is shown in Fig.10.

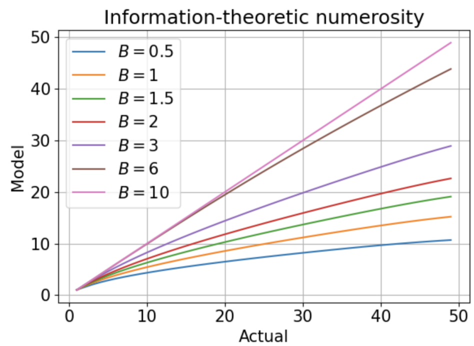


Figure 9: Number of tiles perceived under the model plotted against the actual number.

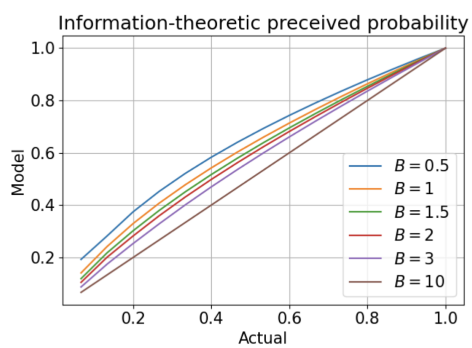


Figure 10: Perceived probability predicted by information-theoretic numerosity model plotted against the actual probability, assuming the exit is equally likley to be in any hidden tile.

Monte-Carlo Tree Search - Sampling Model

Algorithm 1: Monte Carlo Tree Search for MST

Data: Tree to determine best child of node N

Initialize: Decision node N, *Budget*

$k \leftarrow 1$ **while** $k \leq$ *Budget* **do**

    $n \leftarrow N$

    while *node n is not a leaf node* **do**

        $n \leftarrow max_{c \in C(n)} ubc(c)$

    end

    if *node n has been visited* **then**

        add children nodes of node n

        $c \leftarrow$ random child of c

    end

    while *$p \sim$ Uniform(0,1) $>$ P(exit found at c)* **do**

        $c \leftarrow$ random child of c

    end

    value of node $n \leftarrow$ value of node $n + c$

    number of visits to node $n \leftarrow 1 +$ number of visits to node n

    $k \leftarrow k + 1$

end

The Sampling model approximates the EU model with increasing accuracy as the budget parameter is increased. Figure 11 shows the probabilities of actions predicted by the EU and Sampling model, correlation $r = 0.95$, assuming budget and exploration parameters are chosen to maximize the correlation, and EU is parameterized with $\tau = 1$. The optimal exploration parameter depends on the budget, with larger budgets typically requiring higher exploration parameter to achieve best approximation.

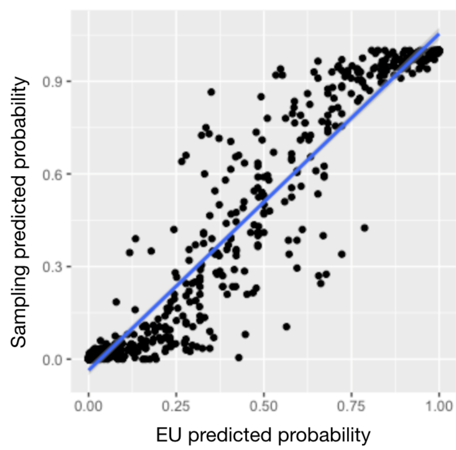


Figure 11: Correlation between action probabilities predicted by EU and Sampling.

5.3 Experiment 1, Additional results

Fitting models with Monte-Carlo Cross validation

Monte-Carlo cross-validation is a method of fitting models over multiple bootstrapped iterations (we run 100) which allows us to obtain confidence intervals on each individual's LLs. Figure 12 A. shows the mean LL per decision as model fit, with 95% CI over participants. Figure 12 B. illustrates variability between individuals. For each indivudual we first determine the best fitting planner and the best-fitting heuristic, and plot the LL of best heuristic and best planning model for each individual with 95% CI. The labels "Heuristic", "Planning", and "Not defined" are assigned based on whether the 95% CI of the LL for best-fitting planner and heuristic overlap. Note that this method may overestimate the number of "Not defined" individuals, as we bootstrap the 95% CI of the mean, not the 95% CI of the *difference* between means. Figure 12 C. summarizes the number of individuals presented in each category in Figure 12 B. in a bar plot. The distribution of best fitting models across individuals is shown in Figure 14.

Computing correlations

Correlation analysis presented in the main text focuses on decisions visited by all participants, so that each data-point used in computing correlation is based on the same population of people. In Figure 13 we show an alternative analysis, aimed to maximize the number of data-points used to compute correlation. Figure 13 shows correlations computed using decisions that were visited by at least 20% of participants.

Here, the correlation of the best-performing DU-Num model with peo-

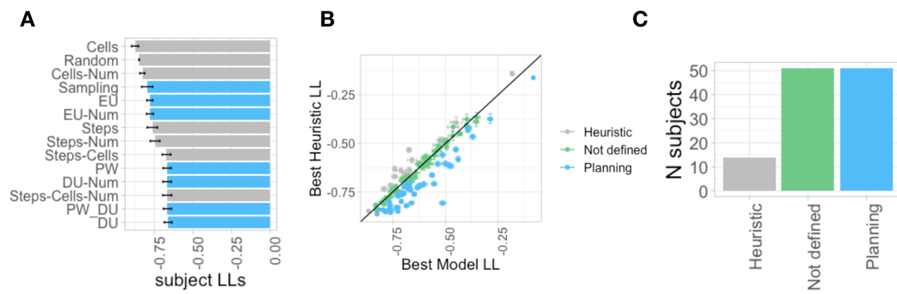


Figure 12: Experiment 1. Models fitted with Monte-Carlo cross-validation. **A.** Mean LL per decision with 95% CI, averaged over participants. **B.** Mean LL per decision for an individual's best-fitting planning model and best-fitting heuristic. Each dot represents an individual, error bars are 95%CI. Participants for whom CIs of heuristic and planning model do not overlap are labeled as "Heuristic" or "Planning". **C.** The number of individuals in each of the categories in panel B

ple is $r = .89(95CI[.86, .92])$, and the correlation of the optimal EU is $r = .72(95CI[.64, .82])$. These correlations are significantly different, with bootstrapped difference between their means of [.07.3], indicating that DU-Num predicts the aggregate population behavior better than the optimal EU model. The 95 CI of bootstrapped difference between correlations of DU-Num (a planner with the highest correlation) and Steps-Cells (a heuristic with the highest correlation) [.02.14], suggesting that the DU-Num model predicts the aggregate population better than the myopic Steps-Cells heuristic. These results remained consistent with conclusions presented in the main text as we re-ran this analysis for percentages $\in [10, 50]\%$.

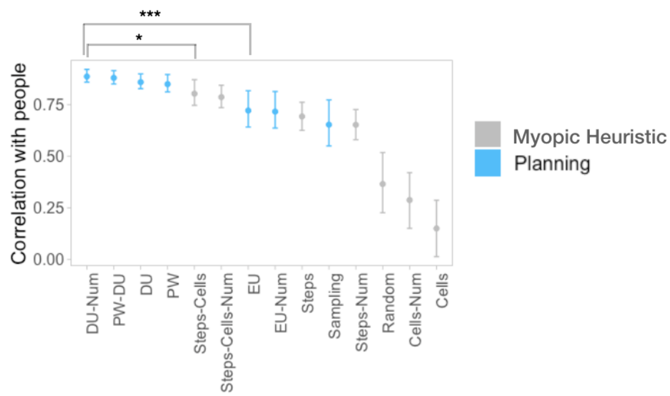


Figure 13: Experiment 1. Bootstrapped correlations of models' predictions with choice probabilities aggregated across the experimental population. The analysis includes all decisions visited by at least 20% of participants. Error bars indicate 95% confidence intervals.

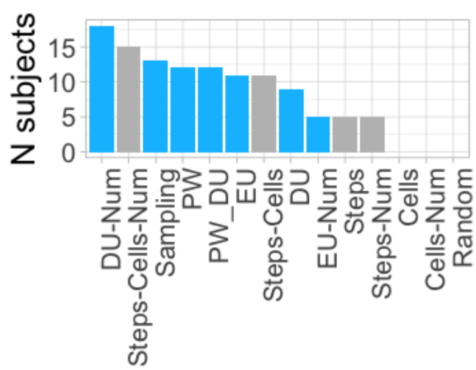


Figure 14: Experiment 1. The distribution of best-fitting models across individuals

Distribution of fitted parameters, Experiment 1

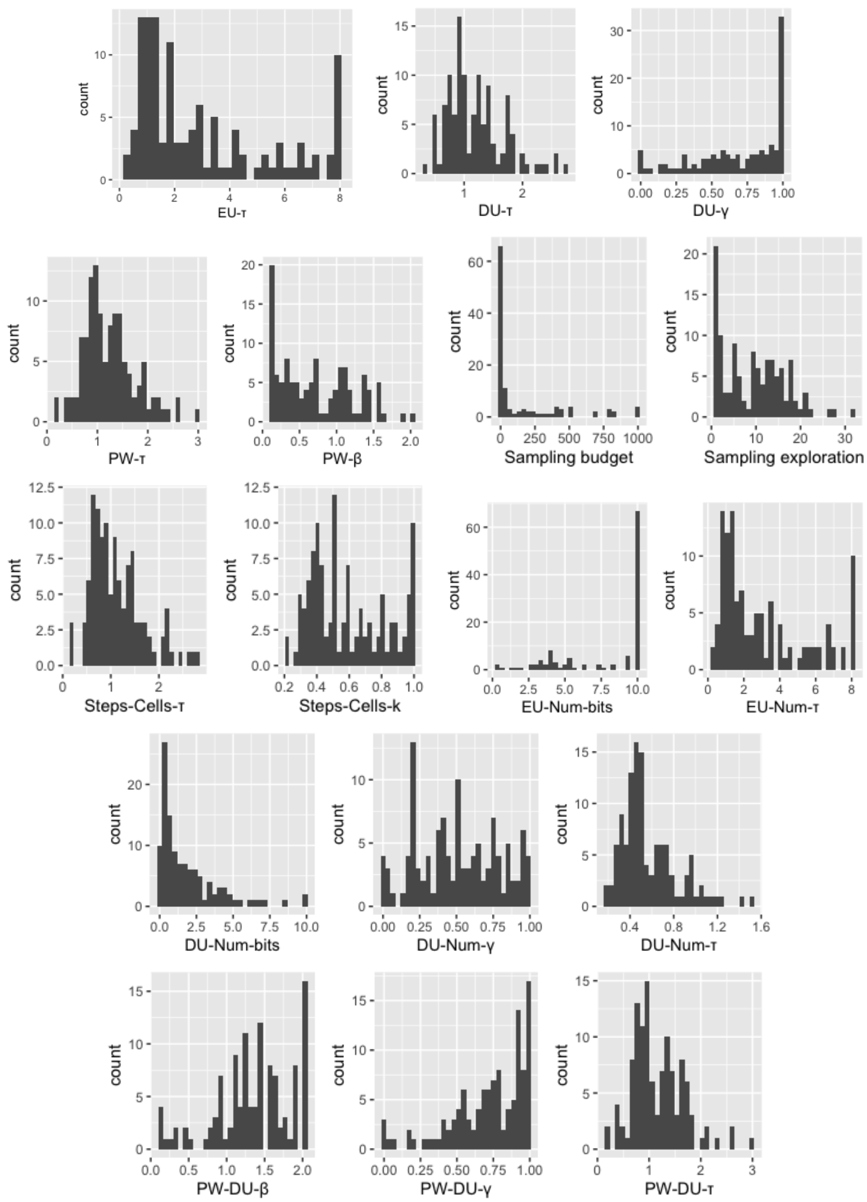


Figure 15: Experiment 1. Distribution of parameters fitted at individual level

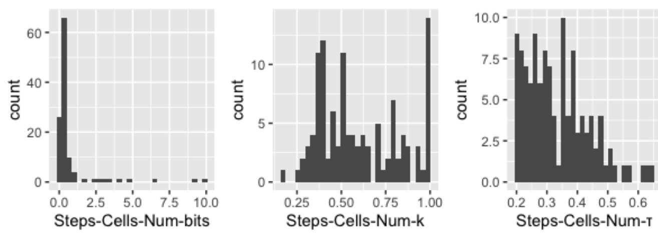


Figure 16: Experiment 1. Distribution of parameters fitted at individual level

5.4 Experiment 2, Additional results

Fitting models with Monte-Carlo Cross validation

We fitted models to individuals using Monte-Carlo cross-validation to illustrates variability between individuals. Figure 17 shows mean LLs of each model per move (panel A) and mean LLs per move of each individual's best fitting planner and heuristic. The labeling of individuals as "Planning" or "Not Defined" is based on non-overlapping bootstrapped 95 CI of best-fitted heuristic and planner (not the bootstrapped 95 CI for difference between means). The distribution of best fitting models across individuals is shown in Figure 18.

Computing correlations

Correlation analysis presented in the main text focuses on decisions visited by all participants, so that each data-point used in computing correlation is based on the same population of people. In Figure 19 we show an alternative analysis, aimed to maximize the number of data-points used to compute correlation, by using decisions that were visited by at least 20% of participants.

Here, the correlation of the best-performing PW-DU model with people is $r = .96(95CI[.95, .97])$, and the correlation of the optimal EU is

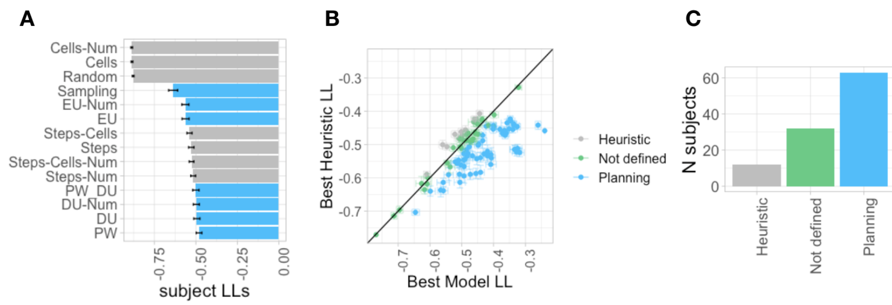


Figure 17: Experiment 2. Models fitted with Monte-Carlo cross-validation. **A.** Mean LL per decision with 95% CI, averaged over participants. **B.** Mean LL per decision for an individual's best-fitting planning model and best-fitting heuristic. Each dot represents an individual, error bars are 95%CI. Participants for whom CIs of heuristic and planning model do not overlap are labeled as "Heuristic" or "Planning". **C.** The number of individuals in each of the categories in panel B

$r = .82(95CI[.76, .86])$. These correlations are significantly different, with bootstrapped difference between their means of [.1.2], indicating that PW-DU predicts the aggregate population behavior better than the optimal EU model. The 95% CI of bootstrapped difference between correlations of PW-DU (a planner with the highest correlation) and Steps (a heuristic with the highest correlation) was [0.020.05], suggesting that the DU-Num model predicts the aggregate population better than the myopic Steps heuristic. These results remained consistent with conclusions presented in the main text as we re-ran this analysis for percentages $\in [10, 50]\%$.

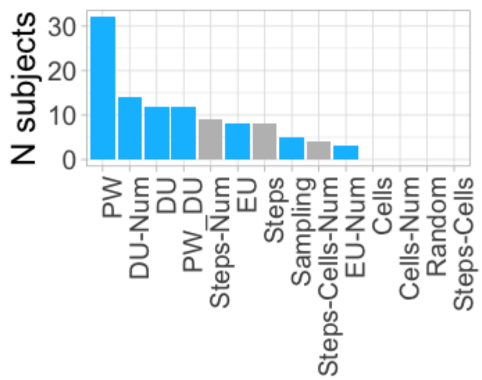


Figure 18: Experiment 2. The distribution of best-fitting planners across individuals

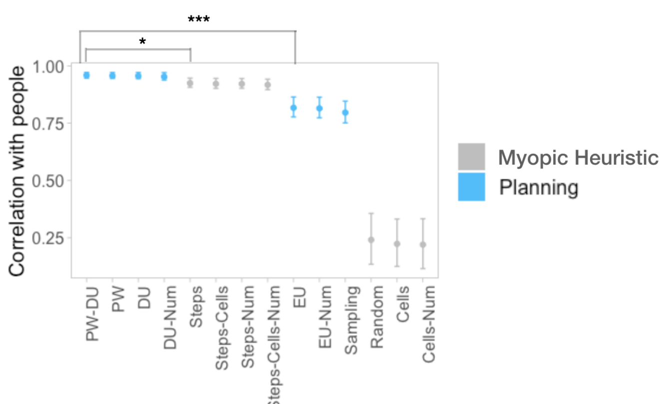


Figure 19: Experiment 2. Bootstrapped correlations of models' predictions with choice probabilities aggregated across the experimental population. The analysis includes all decisions visited by at least 20% of participants. Error bars indicate 95% confidence intervals.

Distribution of fitted parameters, Experiment 2

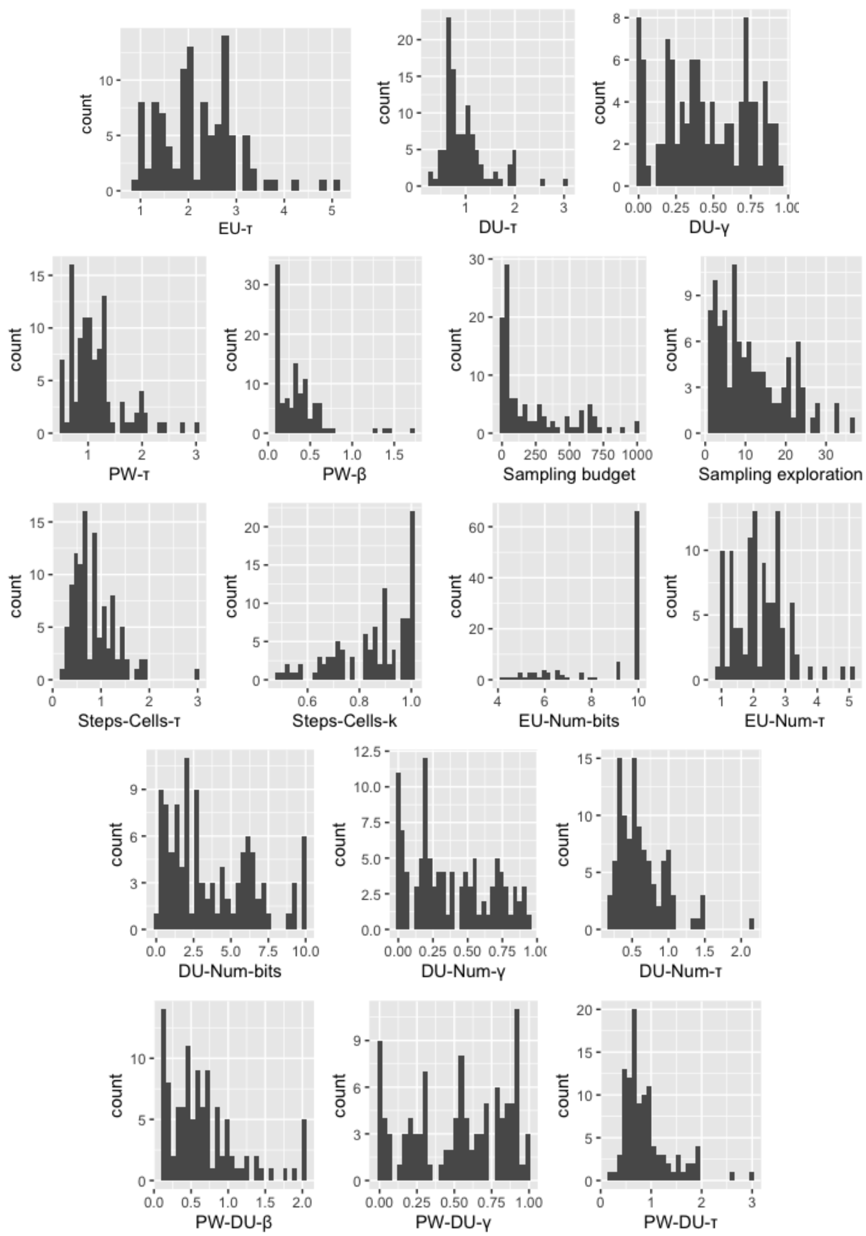


Figure 20: Experiment 2. Distribution of parameters fitted at individual level

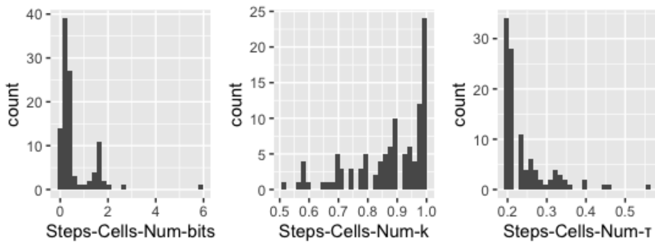


Figure 21: Experiment 2. Distribution of parameters fitted at individual level

5.5 Maze Search Task

The MST task was extensively piloted to ensure the clarity of instructions and a sufficient amount of practice to make the task intuitive to humans. A version of MST has been used to study how do people evaluate the goodness of plans made by others [6], however human planning in MST has not yet been studied by detailed computational modeling, which is the goal of current work. An online versoin of MST can be accessed at `https://marta-kryven.github.io/mst.html`.

Task Instructions:

Screen 1 – Instructions

Welcome to our study!

IMPORTANT

This study runs best in Firefox, on a desktop/laptop.

The study will NOT run on Safari, or a mobile device.

In this study you will look for an exit in a maze.

After this task, you will be asked to provide demographic information.

The study is expected to take about 20 minutes.

Thanks for participating!

(For brevity, we omit the informed consent statement at the end of this page)

button: [I AGREE]

Screen 2 – Instructions

INSTRUCTIONS (PLEASE READ CAREFULLY)

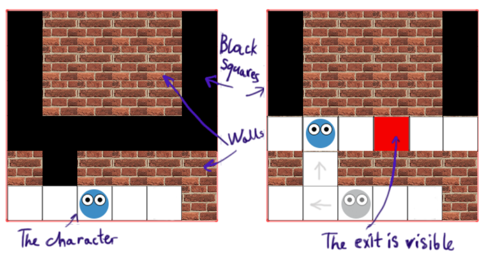
Your task is to exit the maze by reaching the red square in as few steps as possible.

You can move one square at a time by clicking on the white squares next to your character.

You cannot see through the walls. The squares you cannot see yet are black.

The exit is equally likely to be behind any of the black squares.

A maze looks like this:



YOU CAN GET A BONUS!

Planning your path wisely will pay off.

The better you plan your path, the fewer steps you'll take and the more bonus you can earn.

A bonus of $3 will be awarded if you're in the top 20%, OR

A bonus of $2 will be awarded if you're better than average, OR

A bonus of $1 will be awarded if you're better than the bottom 30%.

button: [Let's practice!]

Screen 3 – Practice mazes

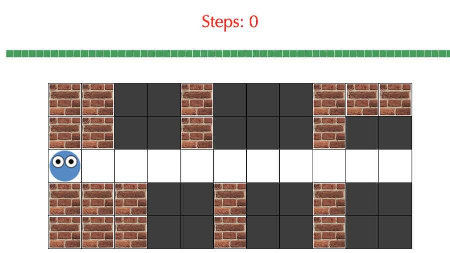
Practice Maze X of Y

Let's look at this map. There are some black squares, a brick wall, and your character.

There is ONE exit in this maze. This exit could be behind any one of the black cells.

You can move your blue character by clicking one of adjacent white cells.

Please find the exit in as few steps as possible.



Screen 4 – Instructions Quiz

Great, you have finished Practice!

Please answer the quiz questions below to move on.

Question 1: My task is to ..

- visit every square in the maze
- see how lucky I am
- solve the mazes in as few steps as possible
- click as fast as possible

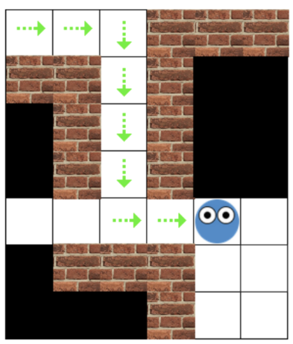
Question 2: Exits are always placed ...

- in the bottom left corner
- anywhere in one of the black cells
- in the first place I search
- in the top right corner

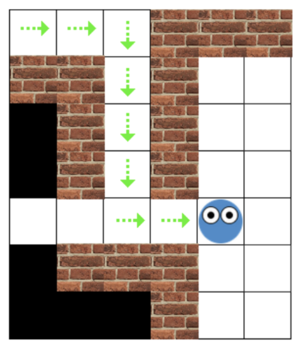
Question 3: Which image correctly shows parts of the maze the character has not seen yet (black squares)?

button: [Submit]

○ Image A ○ Image B



A.

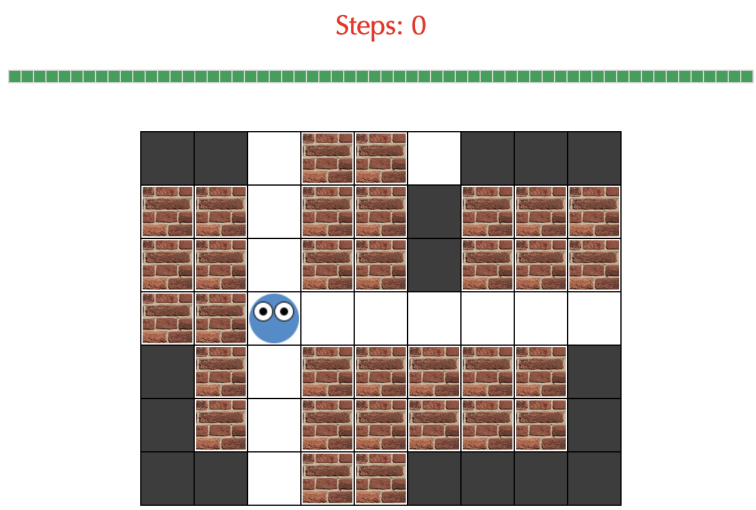


B.

Screen 5 – Experiment

Maze X of Y

Please find the exit in as few steps as possible.



Screen 6

Thank you!

How did you make your decisions about which way to go?

Text input: [...]

button: [Submit]

Screen 7 - Demographics

Your age: [...]

Your gender: [...]

OPTIONAL: Please leave any comments about the study here, we welcome any feedback.

[...]

button: [Submit]

5.6 Mazes Used Experiment 1

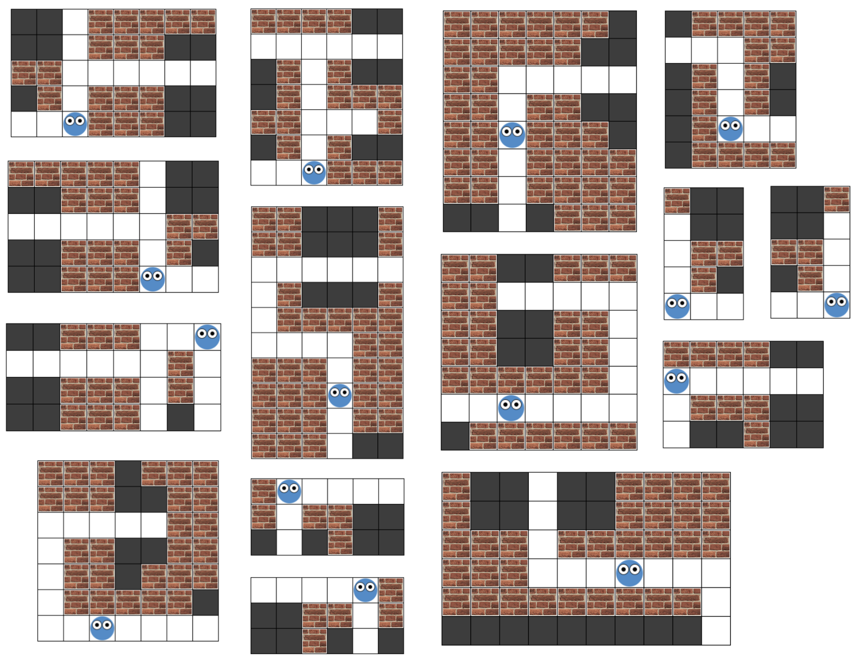


Figure 22: Mazes used in Experiment 1. The mazes were presented in a randomized order. The exit location was chosen randomly at the time of experiment design.

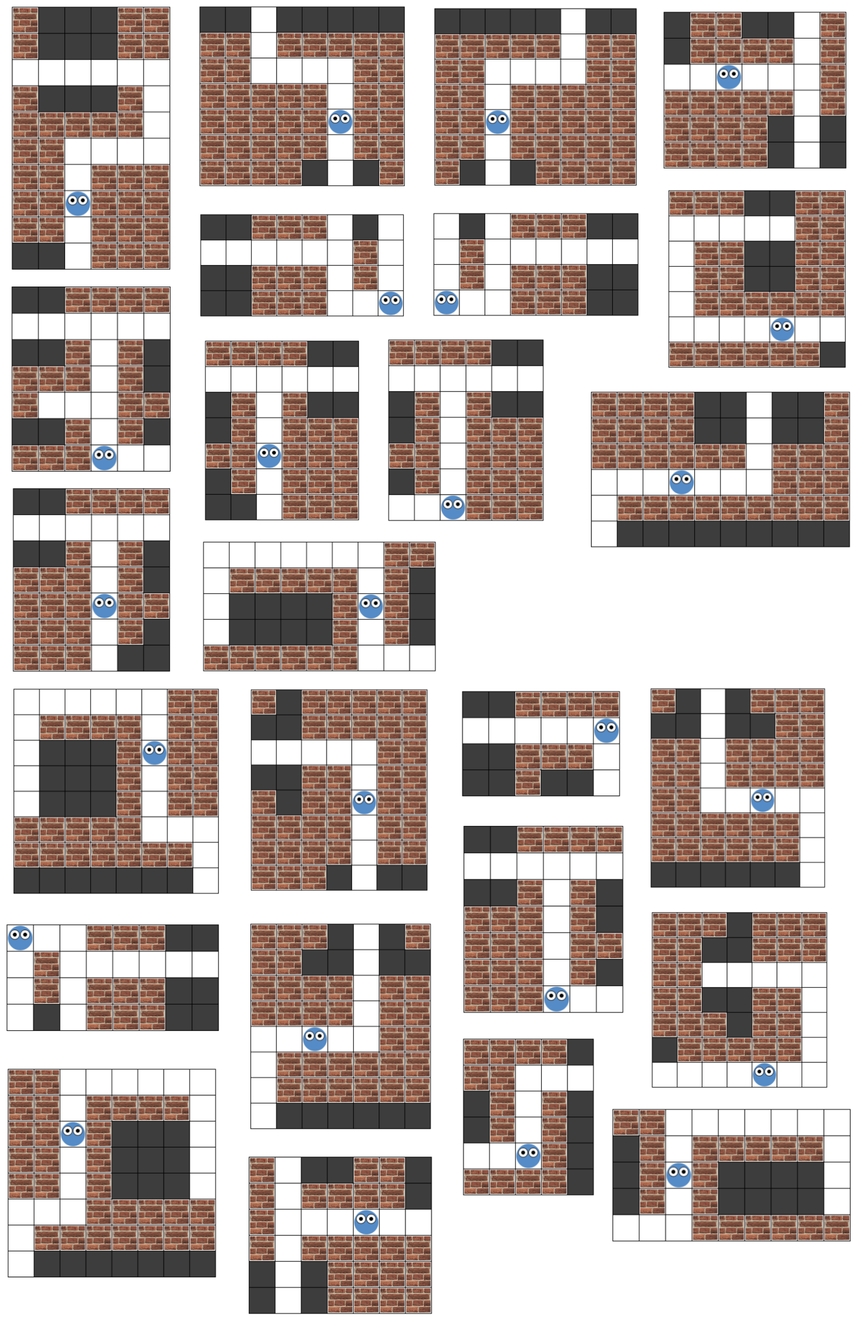


Figure 23: Mazes used in Experiment 1. The mazes were presented in a randomized order. The exit location was chosen randomly at the time of experiment design.

5.7 Mazes Used Experiment 2

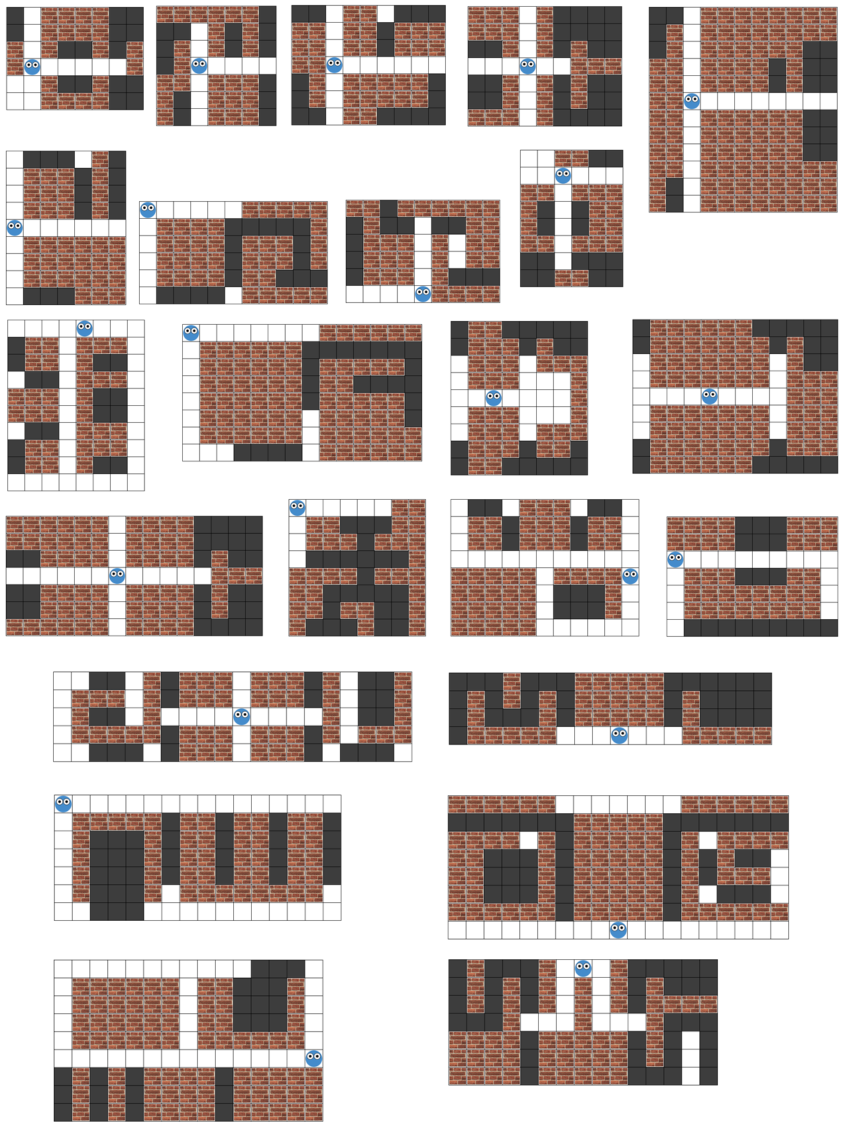


Figure 24: Mazes used in Experiment 2. The mazes were presented in a randomized order. The exit location was chosen randomly at the time of experiment design.