

¹ Approximate Planning in Spatial Search

² Marta Kryven^{1,*} Suhyoun Yu² Max Kleiman-Weiner¹
³ Tomer Ullman^{3,†} Joshua Tenenbaum^{1,†}

⁴ **Abstract**

⁵ How people plan is an active area of research in cognitive science,
⁶ neuroscience, and artificial intelligence. However, tasks traditionally
⁷ used to study planning in the laboratory tend to be constrained to
⁸ artificial environments, such as Chess and bandit problems. To date
⁹ there is still no agreed-on model of how people plan in realistic contexts,
¹⁰ such as navigation and search, where values intuitively derive from
¹¹ interactions between perception and cognition. To address this gap
¹² and move towards a more naturalistic study of planning, we present a
¹³ novel spatial Maze Search Task (MST) where the costs and rewards
¹⁴ are physically situated as distances and locations. We used this task in
¹⁵ two behavioural experiments to evaluate and contrast multiple distinct
¹⁶ computational models of planning, including optimal expected utility
¹⁷ planning, a family of planners that approximate optimal planning, and
¹⁸ myopic heuristics inspired by studies of information search. We found
¹⁹ that in contrast to myopic heuristics or the optimal planning, people's

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

20 behavior is best explained by approximate planners with limited plan-
21 ning horizon, in which values are estimated by the interactions between
22 perception and cognition. This result makes a novel theoretical contri-
23 bution in showing that limited planning horizon generalizes to spatial
24 planning, and demonstrates the value of our multi-model approach for
25 understanding cognition.

26 Author Summary

27 We present a computational study of spatial planning under uncertainty
28 using a novel Maze Search Task (MST), in which people search mazes for
29 probabilistically hidden rewards. The MST is designed to resemble real-life
30 planning where costs and rewards are physically situated as distances and
31 locations. We found that in this context people’s behavior is best explained
32 by approximate planners with limited planning horizon, as opposed to both
33 myopic heuristics or the optimal planning, showing that limited planning
34 horizon can generalize to spatial planning tasks.

35 1 Introduction

36 People make plans every day: working out a new route, playing a game,
37 thinking through a possible conversation. Despite the ubiquity of planning,
38 *how* people plan is an active topic of research in many different fields. Plan-
39 ning involves making sequences of choices, where the possible actions that
40 are available at each step depend on the outcome of the previous step. This
41 process has been formalized as navigating a decision tree, which starts at

42 an initial ‘root’ state, and continues until it reaches a ‘leaf’ that meets the
43 goal criteria [1, 2, 3, 4, 5]. For any situation beyond trivial toy problems, the
44 growing complexity of a branching decision tree makes planning computa-
45 tionally costly. And yet, people daily face situations that require planning,
46 and handle them remarkably well.

47 A prime example of a daily planning task that people are quite adept at
48 is spatial planning. But while an enormous amount of work has studied and
49 modeled planning in different domains (as we review below), to our knowledge
50 relatively few studies considered detailed computational models of spatial
51 planning in naturalistic contexts, under uncertainty. In this work, we take a
52 step toward understanding human spatial planning under uncertainty using a
53 novel Maze Search Task (MST), designed to resemble a natural environment
54 where the costs are distances and rewards are spatial locations. A version
55 of MST was previously used to study how people evaluate the goodness
56 of plans made by others [6], but a detailed computational account of how
57 people themselves plan in the MST was not addressed. We use the MST to
58 explore a family of cognitively-inspired computational planning models, that
59 approximate expected utilities by integrating perceptual transformations,
60 such as numerosity [7] and probability weighting [8]. We contrast these models
61 both with an optimal Expected Utility model, and with a family of intuitive
62 myopic heuristics, that could in principle be used to search an environment
63 without planning ahead. In what follows, we briefly review recent relevant
64 empirical work that attempted to model human planning. Based on this
65 literature, we outline the relevant take-aways for the models we will consider
66 in this work, including perceptual transformations and constraints that may

67 influence how people plan in real-world.

68 First, while we mentioned that relatively little of the work on planning has
69 focused on computational models of naturalistic spatial planning, certainly
70 research has been done on it. For example, a small study modeled route
71 planning within a city neighborhood as optimally solved by Breadth First
72 Search [9]. However, larger empirical studies find that people rarely take
73 shortest routes [10, 11], indicating that people likely plan using cognitive
74 approximations when they navigate. Cognitive constraints that cause such
75 approximations, or the mental computations implementing them, are not yet
76 addressed in prior work.

77 Given the time constraints of in-lab studies, most prior work has focused
78 on planning short action sequences in non-spatial situations, such as sequences
79 of 2 or 3 actions, in rigorously designed yet simple experimental paradigms
80 [3, 12, 5, 13, 4]. Several studies of games, such as Chess and Tic-tac-toe,
81 have modeled planning over longer horizons in contexts with rich behavioral
82 variation [14, 15, 16, 17]. However, game state utilities in these tasks derive
83 from task-specific heuristic game-board assessments that evaluate the player's
84 position based on various features of the board, making it hard to generalize
85 the results beyond those specific games.

86 Several studies of planning suggest that people manage cognitive demands
87 by limiting their planning horizon, which has implications for models of
88 spatial planning. For example, when people are asked to connect moving
89 consecutive disks to maximize their total volume, they do so in a way best
90 explained as considering all possible paths up to a certain limited depth,
91 in contrast to simple myopic heuristics, such as moving toward the largest

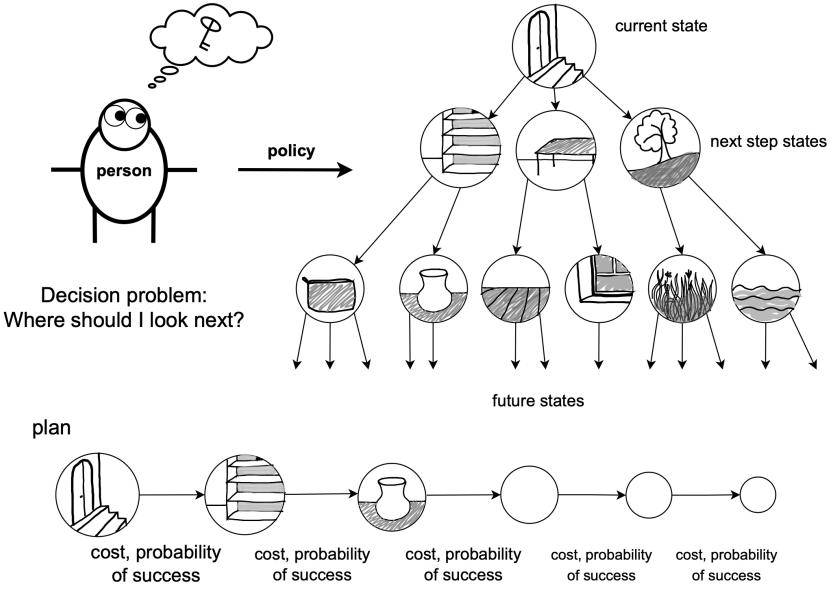


Figure 1: The decision problem facing an agent is to take actions that maximize long-run rewards. The agent can solve this problem by planning a path through a decision tree that recursively minimizes costs, while taking into account the probabilities of success. Our task captures this process in a spatial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

92 disk [2]. In a different popular example, the Tower of London task requires
 93 stacking disks in a certain order, using the fewest moves. People doing this
 94 task increasingly deviate from the optimal number of moves as the depth of
 95 planning required to optimally solve the task increases, which again suggested
 96 that people use depth-limited planning [18]. In yet another example, several
 97 studies examined how people learn an implied decision tree in a three-stage
 98 bandit task, which involves making sequential decisions to learn and exploit a
 99 hidden task structure. People were found to prune their implied state-space
 100 during the choice stage in response to rising cognitive demands, suggesting

101 that they relied on a flexible depth-limited representation to control cognitive
102 costs [3, 12, 5]. Other examples of a limited depth of search are found in board
103 games, such as Chess [15, 19, 16], and Four in a Row [14], where increasing
104 expertise corresponds to greater depth in search algorithms [15, 19, 16, 14].

105 Given that several studies model limited planning horizon in various non-
106 spatial tasks, prior literature provides strong evidence that limited planning
107 horizon does occur under certain scenarios, and may generalize to realistic
108 spatial contexts as well. Based on prior work, limited planning horizons can
109 be implemented formally in several ways. Studies have limited planning depth
110 by probabilistically pruning the planning tree at each node, approximated
111 by a discount rate applied to future rewards [12, 3], limiting the number of
112 nodes explored by an algorithm, such as Monte Carlo Tree Search [14, 20],
113 or planning up to a fixed depth [2]. Probabilistic tree pruning has been
114 implemented as a discount rate, by assuming a mean-field approximation
115 where at each step the future is weighted by the probability that it will be
116 encountered (see Methods, [12]). Similarly, reinforcement learning routinely
117 relies on discounting of future states to keep state-values from growing to
118 infinity in tasks where horizon can be potentially unlimited [21], although
119 this approach does not guarantee an optimal solution [22]. In this work, we
120 implemented a limited planning horizon by discounting, which integrates
121 naturally with our modeling approach (see below).

122 Beyond limited horizons, everyday planning often takes place in conditions
123 of uncertainty, which may require gathering more information. In our task,
124 uncertainty arises from the partial observability of the environment, as
125 planning a search in MST means deciding in which order to observe a given

126 maze. In the context of observing an unknown environment, people were
127 previously found to use myopic strategies to choose the next observation [23].
128 Prior work in non-visual domains likewise tends to model information-seeking
129 by myopic heuristics: choosing observations one at a time, rather than
130 by planning ahead [24], with an extensive body of literature showing that
131 people but tend to use cost-insensitive information search heuristics - that
132 is consider the value of information, but not its cost [25, 26]. Though, in a
133 recent exception, gathering information about costs and rewards in decision-
134 trees of depths two and tree with deterministic state transitions was best
135 explained by a family of models that go beyond single-step heuristics [13, 4, 27].
136 In [13, 4, 27] people saw a tree-like state-space structure, and had an
137 opportunity to click on specific nodes to reveal their value for a given cost
138 before navigating from the root to one of the leafs. The study found that
139 during the initial information-seeking phase people optimized their total
140 expected reward against the costs of gathering information, in contrast to
141 heuristically selecting the next node (e.g. the closest node). They also found
142 that people systematically used a variety of near-optimal planning strategies
143 to reduce uncertainty and seek information relevant to achieving their goal,
144 when information comes at different costs [27].

145 In addition to limited planning horizons and uncertainty, people's plans
146 rest on intuitive subjective utilities, which estimate the expected value of
147 prospective states. Subjective utility functions can take different forms,
148 examined in decision theories, such as Prospect Theory [28, 29, 8]. In
149 real-world spatial planning, however, value derives from perceived physical
150 quantities, such as number, and area. Given this, we hypothesized that real-

151 world planning models may need to account for perceptual transformations
152 over these quantities. In particular, the perception of numerosity is known to
153 follow a non-linear relationship between the actual and perceived number [7],
154 and probabilities tend to be perceived in a non-linear way that overestimates
155 small and underestimate large probabilities. While probability perception
156 has been widely explored [30, 28, 29], cognitive probability models are not
157 yet integrated into studies of human planning. In our modeling approach
158 we will consider previously found perceptual transformations of number and
159 probability, and examine if they result in more human-like planning models.

160 Overall, previous computational and empirical work on human planning
161 suggests that we need to build and evaluate planning models that have
162 limited planning horizons and estimate intuitive utilities by using perceptual
163 transformations. We also need to compare such models to simpler alternatives
164 that do not involve planning ahead, which we broadly construe as myopic
165 heuristics. Finally, we need a rich physically grounded environment to contrast
166 the different models of planning in naturalistic settings. In the following
167 sections, we detail the novel task we used to study people’s spatial planning,
168 as well formally define planning models we compared people to.

169 **2 Experimental Methods**

170 **2.1 Maze Search Task**

171 The objective of a participant in the Maze Search Task (MST) is to navigate
172 a series of partially observable, two-dimensional grid-world mazes while
173 minimizing the distance traveled to reach a hidden exit. Each maze consists

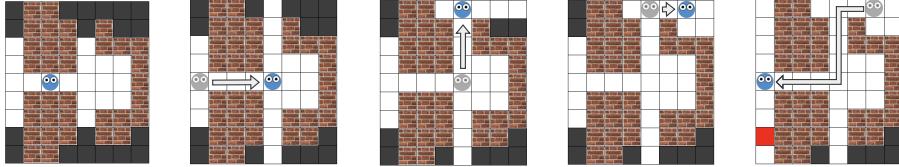


Figure 2: An example path in the Maze Search Task (MST). Black tiles are not-yet-observed areas, which hide an exit (red square). This maze has six ‘rooms’, groups of black tiles that are revealed all at once. Revealing tiles can be done in any order, but players are incentivized to plan their path so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

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

183 The player in the MST can move to any adjacent grid tiles that are not
 184 blocked by walls in the four cardinal directions (up, down, left, right), and
 185 reveal the black unobserved tiles by bringing them into the avatar’s line
 186 of sight. To compute the agent’s isovist in the context of a gridworld, we
 187 consider a cell to be visible from the agent’s location if its entire area can
 188 be seen from anywhere within the agent’s current cell, taking into account
 189 any obstacles. For example, the diagonal cell at node N2 in Figure 3 is not
 190 visible, as its upper-right corner can not be seen from anywhere inside the

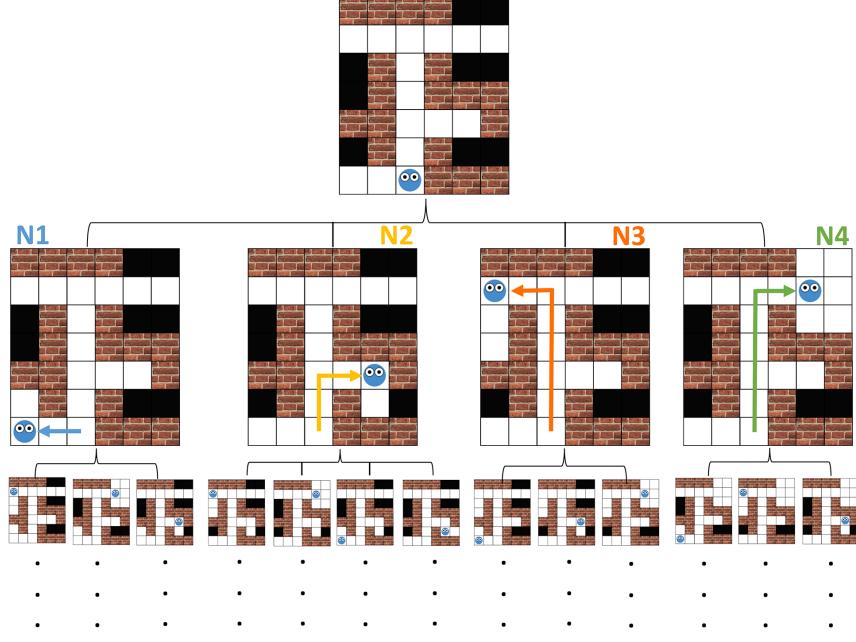


Figure 3: Decision-tree for a maze with four rooms (hidden tiles that are revealed together). The tree abstracts away from specific moves like 'up' and 'left' and considers more general actions like which area to uncover next. The root of the decision-tree corresponds to the player's starting location. 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.

191 agent's cell. Upon reaching the exit the player is moved to the next maze. If
 192 the current maze is the last trial in the experiment, then upon reaching the
 193 exit the experiment ends.

194 2.2 Computational Models

195 To plan a route through a maze, we first define a decision-tree detailing all
 196 possible orderings in which the rooms in a maze can be revealed. Each node in
 197 this tree corresponds to a unique combination of a location (x, y) from which

198 an observation is made, and the list of all observations so far, including the
 199 newly observed area. If a room can be observed from two different locations,
 200 then this room will be represented by two nodes in the decision tree, each
 201 corresponding to a unique location from which an observation is made. An
 202 example of a maze along with a partial decision tree for this maze is shown
 203 in Fig. 3. The starting location corresponds to the root of the decision-tree.
 204 The leaves correspond to all possible unique ordering of observations through
 205 which the entire maze can be revealed.

206 Planning a path through a maze entails computing costs `values` of decision
 207 tree nodes C_k , and mapping the node costs `values` to probabilities of choosing
 208 them. We use a softmax mapping, a standard method of modeling expressed
 209 preferences in decision-making [3, 31, 32, 33, 34]:

$$\sigma(\mathbf{C})_k = \frac{\exp(-C_k/\tau)}{\sum_j \exp(-C_j/\tau)}. \quad (1)$$

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

216 2.2.1 Planning ahead

217 We first define the optimal Expected Utility cost `value` function, minimizing
 218 which computes the shortest paths to the exit. We will use the optimal model

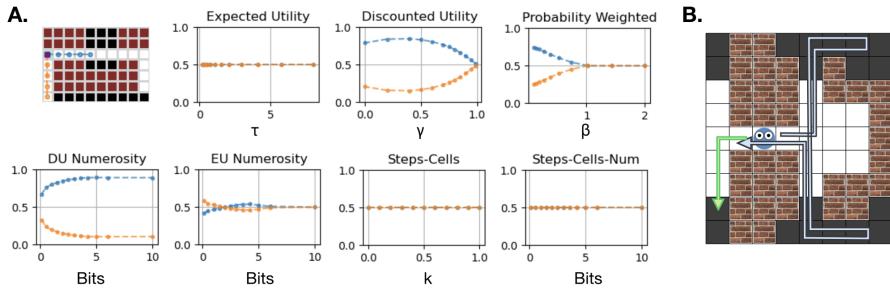


Figure 4: Models predicting different paths. **A.** In this maze the two directions (Right and Down) have a different likelihood of being chosen by the different models, in a way that depends on parameters of the models. Each plot shows the probability of choosing each direction by ~~A maze in which 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), and the Probability Weighting (PW) model predict 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. Notice that for all path choices, the initial decision opens an identical numbers of cells and takes an identical number steps, meaning that at the initial decision heuristics will be indifferent between taking blue and green paths.

219 to derive other, approximate planners.

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

$$C_{EU}(N_i) = s_i + p_i e_i + (1 - p_i) \min_{c_j \in Child(N_i)} C_{EU}(c_j). \quad (2)$$

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

233 The **Discounted Utility (DU)** model modifies the EU model by dis-
234 counting the costs values of future nodes, by a rate of $\gamma \in [0, 1]$, to implement
235 a limited planning horizon:

$$C_{DU}(N_i) = s_i + p_i e_i + \gamma(1 - p_i) \min_{c_j \in Child(N_i)} C_{DU}(c_j) \quad (3)$$

236 The **Probability Weighed Utility (PW)** model modifies the EU
237 model by transforming probabilities, using a weighting function of the form
238 $\pi(p) = \exp(-|\ln(p)|^\beta)$ [8]. While probability weighting is widely used to

239 model probability perception in monetary gambles [30, 28, 29] it is not
240 commonly evaluated in the context of planning.

241 We define the PW costvalue function as follows:

$$C_{PW}(N_i) = s_i + \pi(p_i)e_i + \pi(1 - p_i) \min_{c_j \in Child(N_i)} C_{PW}(c_j) \quad (4)$$

242 This probability weighting function has an effect of overestimating small
243 probabilities and underestimate large probabilities when $0 <= \beta < 1$, and
244 has the opposite effect when $\beta > 1$. If $\beta = 0$ all probabilities have the same
245 uniform value, and $\beta = 1$ is equivalent to the optimal Expected Utility in
246 which all probabilities are equally weighted.

247 **The Probability Weighting and Discounting (PW-DU)** model com-
248 bines probability weighting and discounting in a model with three free
249 parameters— τ , γ , and β (corresponding to the softmax temperature, the
250 discount rate, and the probability weighting respectively).

$$C_{pw-du}(N_i) = s_i + \pi(p_i)e_i + \gamma\pi(1 - p_i) \min_{c_j \in Child(N_i)} C_{pw-du}(c_j) \quad (5)$$

251 **Expected Utility with Numerosity psychophysics (EU-Num)**

252 When people reason about varying amounts of hidden tiles, they could
253 in principle count the tiles exactly. At the same time, people could instead
254 roughly estimate the number of tiles, causing systematic deviations [35, 7].
255 We use a recent information-theoretic numerosity model to account for the
256 latter possibility [7]. The information-theoretic numerosity model has one free

257 parameter, bit threshold B – the number of bits processed by the perceptual
258 system to estimate numbers, such that people who count the tiles can be
259 modeled with a large B , and people who guess at a glance can be modeled
260 with a small B (we used $B \in [0.1, 10]$).

261 To fit bit threshold B to people, we generate a table of mappings between
262 observed quantities $n \in [1, 80]$ (an empirical upper bound on the number of
263 unobserved cells in a maze) and an expectation of a subjective estimate q ,
264 assuming different levels of $B \in [0.1, 10]$. The upper bound on B is empirically
265 chosen as sufficient to perceive 15 cells exactly, which we deem sufficient, given
266 that human subitizing range is shown to be limited to about 6 [36]. The lower
267 bound is an empirically chosen minimum with a subitizing range below 2. We
268 use this table to transform numbers of tiles used when computing node costs
269 in the models parameterized with numerosity, substituting q for n given each
270 level of B . This transformation affects the model’s estimates of the number of
271 steps between rooms, room sizes (the number of tiles that would be revealed
272 by an action), and the total number of remaining tiles. Perceptual number
273 transformation also has an indirect effect on probability perception, as it
274 affects both the number of cells in a room and the total number of cells in a
275 maze, in a non-linear way that depends on B . See Supplemental Materials 5.2
276 for more details. To estimate the expectation, we computing the probability
277 distribution $Q(k|n, B)$ that an observed numerosity n is represented by q
278 using equation (5) from [7]. We set a prior over numbers as probability of
279 how often a numerosity n is encountered, $P(n) \propto 1/n^2$ according to Zipf law,
280 following the approach used in [7].

281 **Discounted Utility with Numerosity psychophysics (DU-Num)**

282 The DU-Num incorporates information-theoretic numerosity into the DU
283 model along the lines above.

284 **2.2.2 Monte Carlo Tree Search Sampling model (Sampling)**

285 The planning models considered so far accurately represent the planning
286 tree, but assume that the utilities of its nodes are approximated in some way.
287 Another way to make approximate plans is to construct a partial planning
288 tree. This process can be formalized by the Monte Carlo Tree Search (MCTS),
289 an algorithmic framework for simulating multiple possible outcomes while
290 keeping track of them in a tree, and choosing the best one based on the
291 simulation results. The accuracy of the planning tree in MCTS is controlled
292 by two free parameters – the computational budget that controls how many
293 nodes are sampled, and exploration that controls how greedy or stochastic the
294 process is [37]. Implementations of MCTS have been previously used to model
295 games such as chess, go, and four-in-a-row [14]. Here we implement MSTC
296 to approximate the EU model, with it does with increasing accuracy as the
297 budget parameter is increased. See Supplemental Materials for details. We
298 point the interested reader to Supplementary Section 5.2.4 for an alternative
299 formulation of the Maze Search problem using a classic reinforcement learning
300 notation for Partially Observable Markov Decision processes in a grid-based
301 state-space.

302 **2.2.3 Myopic Heuristics**

303 A large body of decision-making literature has focused on heuristic solutions
304 to problems, where a heuristic derives from simple rules or features, such as

305 the size of an area [2, 38, 24]. While it is impossible to enumerate all the
306 possible heuristics that could be invented in response to a given situation, we
307 considered 7 one-step heuristics based on prominent features of MST, to rule
308 out simple interpretations of the task.

309 **Cells heuristic (Cells)** The costvalue of a node is taken to be the
310 number of tiles observed at that node: $C(N_i) = -cells_i$. We use a negative
311 sign because revealing more cells results in a *smaller* expected cost for finding
312 the exit. This approach greedily minimizes the entropy over goal locations,
313 similar to previously studied cost-insensitive information search heuristics,
314 e.g. [25, 26]. .

315

316 **Steps heuristic (Steps)** The costvalue of a node is taken to be the
317 number of steps to the node from its parent: $C(N_i) = s_i$. This approach
318 treats the cost of an observation as reduced to its immediate cost, disregarding
319 future states to which this observation might lead.

320

321 **Steps-Cells heuristic (S-C)** A combination of the two heuristics
322 above. The costvalue of a node is a combination of the steps to get to
323 it and the revealed cells, but without planning more than one step ahead:
324 $C(N_i) = ks_i - (1-k)cells_i$, where $k \in [0, 1]$ is a free parameter. This heuristic
325 balances value of information against its costs, similarly to one of the models
326 explored in [25].

327

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

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

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

334 **2.2.4 Different paths preferred by different models.**

335 The models described above can make different predictions about how a
336 maze should be traversed. A simple example is shown in Figure 4, and in
337 Figures 15 - 20 in the Supplementary Materials. At the basic level, the Steps
338 heuristic predicts a preference for closer rooms, regardless of their shape or
339 size. Importantly, the Steps heuristic makes this choice one step at a time,
340 so that the path it takes to search an entire maze does not minimize *overall*
341 steps. The Cells heuristic can explain a preference to open larger rooms, one
342 step at a time, regardless of distance and regardless of which rooms will be
343 made accessible by this choice. The Steps-Cells heuristic allows for combining
344 the Steps and Cells within a single rule. Importantly, this heuristic will still
345 be indifferent between choosing one of two equidistant rooms with the same
346 number of tiles, even when one of them leads to a revealing a bigger space
347 on the next step.

348 The DU model can prioritize visiting closer rooms, and be insensitive to
349 incurring a path with potentially large detours occurring later in the path.
350 Figure 4 B. shows an examples where models EU and DU take different
351 routes, because DU discounts the possibility of having to backtrack in the
352 future. Figure 4 A shows a maze with a looping corridor, where neither the
353 optimal EU model, nor heuristics can differentiate between the two directions,

354 as both reveal rooms of an equal size that are equidistant from the observer.
355 In this example, the majority of people search the rectangular room first. The
356 DU model accurately predicts this preference, as it distinguishes between the
357 different shapes of the two rooms – upon finding the exit in the rectangular-
358 shaped room we can expect it to be on average closer, compared to average
359 distance to a black cell in the stick-shaped room. Given this, for the DU
360 model it makes sense to search the rectangular room sooner. The PW models
361 also captures this preference, but for a different reason - as it prefers to
362 travel clockwise through the loop. See also Supplementary Figure15 for an
363 illustration where PW and DU models behave differently.

364 While probability weighting principle in model PW has been previously
365 explored to model how people choose between gambles, extending it to
366 sequential choices in the spatial search context has unique implications. First,
367 for $\beta < 1$ PW makes room of different size seem more alike, so that, for all
368 cases where the optimal model EU takes a longer path toward a bigger room,
369 there exists a $\beta < 1$ that reverses this preference. In this sense, both PW
370 and DU are more likely to choose closer rooms, but PW does this through
371 relaxing disparity in room sizes, while DU cares less about a future in which
372 it has to search a further room, if that future can be put off.

373 Second, values of $\beta > 1$ makes room of different sizes appear more different,
374 such that the seeker pays exaggerated attention to room size. Boundary
375 values of $\beta = 0.1$ will guide the model to behave like a variant of the Steps
376 heuristic with look-ahead - that is, minimizing the length of the overall path
377 to *reveal the entire maze* instead of minimizing the length of path *to reach
378 the exit*. Boundary value of $\beta - > \infty$ will behave like the Cells heuristic with

379 look-ahead. See Supplemental Section 5.2 and Supplemental Figures 15 -
380 20 for more details, including examples of how model PW can in principle
381 traverses a large maze in two different ways depending on its parameters,
382 and do so by a trajectory distinctly different from the optimal EU model. In
383 other words, the space of models we consider above spans a wide range of
384 possible behaviors, due to very different commitments they make about how
385 one plans through a spatial environment.

386 2.3 Data analysis methods

387 **Model fitting.** The models were fitted at the individual level by considering
388 all decisions made by an individual during the experiment, which are treated
389 as independent of each other. A ‘decision’ in our case refers to choosing one of
390 two or more child nodes, meaning choosing which one of the unobserved rooms
391 to search next. We fit our models to individuals, and obtained out-of-sample
392 predictions using fivefold cross-validation. The model fitting procedure sets
393 a prior on softmax temperature τ inversely proportional to its magnitude,
394 but assumes a uniform prior over all other parameters (e.g. γ, B, β, k). We
395 measured the *model performance* as the total test log likelihood (LL) of the
396 model across all five test folds. This metric accounts for the flexibility of
397 the different models without parameter counting, in contrast to AIC and
398 BIC which can be a poor measure of flexibility [39]. Differences in this
399 cross-validated LL (ΔLL) can be interpreted similarly to differences in AIC:
400 $\Delta LL = 1$ is roughly equivalent to $\Delta AIC = 2$. The Supplemental Materials
401 also show model performance analyzed using Monte-Carlo cross-validation.
402 We analyzed *individual differences in planning* by comparing the likelihood

403 of the best fitting planning model (one of EU, EU-Num, PW, DU, DU-Num,
404 PW-DU, Sampling) and best fitting heuristic (one of Steps, Cells, Steps-Cells,
405 Random, Steps-Num, Cells-Num, Steps-Cells-Num) for each individual.

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

410 We first computed the probabilities of participants making each choice
411 at the initial decision point in each maze, aggregated across individuals.
412 Although individuals then proceed to search each maze in different ways
413 (with each person potentially facing a unique set of choices), the initial
414 decisions are shared by all participants in the experiment. We then correlated
415 these probabilities with each model's predictions, where the models are
416 parameterized with the mean parameters of the experimental population.
417 While we consider the first choice to be the most indicative and carefully
418 controlled measure, in Supplemental Materials we also show correlations
419 computed using all decisions in the experiment that were visited by at least
420 20% of the participants.

421 **Decision time analysis.** There are two types of decision times in Maze
422 Search:

- 423 (1) **Initial decision time** to make the first move at the start of the trial,
424 during which people form a mental representation of a maze, and plan a
425 search within a certain planning horizon (or choose where to observe next
426 using a step-wise heuristic) – the time at the root of the decision tree.
- 427 (2) **Subsequent decision time**, corresponding to inner nodes of the decision

428 tree where people either make a pre-planned move, or decide where to observe
429 next using a step-wise heuristic. We analyze both types of decision times to
430 assess the relationship between the model parameters fitted to people and
431 the extent of mental computations, as evidenced by decision times.

432 The simplest interpretation of decision times is that they reflect precision
433 of computation, so that longer Initial decision times should indicate more
434 accurate planning. Additionally, more planning may lead to Shorter sub-
435 sequent decision times, as people making a per-planned move should need
436 less time compared to people who compute a step-wise heuristic. Without
437 making specific assumption about implementation, precision of computation
438 is measured by the fitted softmax temperature τ of the optimal EU model,
439 (a smaller τ is more optimal). To test whether decision times reflect precision
440 of computation, we fit a linear regression model predicting decision times by
441 fitted softmax temperature in the EU model. Since a smaller τ corresponds to
442 closer approximation of expected utility, by our hypothesis we would expect
443 to see a negative slope α for Initial decision times, and a positive slope for
444 Subsequent decision times.

445 Decision times may also reflect people's planning horizon, as measured
446 by fitted discounting parameter γ , in which case a higher γ (more planning,
447 less discounting) fitted to people should corresponds to longer Initial and
448 shorter Subsequent decision times. To test this hypothesis, we compute linear
449 regression predicting decision times from discounting parameters fitted to
450 people in models DU, DU-Num, and PW-DU. Since a higher γ corresponds to
451 more planning and less discounting, by our hypothesis we should see a positive
452 slope at γ for Initial decision times, and a negative slope for Subsequent

⁴⁵³ decision times.

⁴⁵⁴ 2.4 Experimental Procedure

⁴⁵⁵ The experiment was conducted in a web browser, using a JavaScript and PHP
⁴⁵⁶ interface developed by the authors. Participants first read a consent page and
⁴⁵⁷ a short description of the Maze Search Task. Following consent, participants
⁴⁵⁸ read a detailed description of the task, completed maze navigation practice,
⁴⁵⁹ and answered an instruction quiz. The quiz included questions about the
⁴⁶⁰ objectives of the task, task controls, and the line-of-sight mechanism by which
⁴⁶¹ the mazes are revealed. Participants could not proceed to the experiment until
⁴⁶² they submitted the correct answers to the quiz. On each trial a participant
⁴⁶³ was placed at the starting position, and navigated by clicking on one of the
⁴⁶⁴ adjacent tiles, until the exit was reached. The starting locations in each maze
⁴⁶⁵ were predetermined, and the exit locations were randomly chosen for each
⁴⁶⁶ maze at the time of design. After completing the experiment, each participant
⁴⁶⁷ answered a demographic questionnaire, and provided a free-form description
⁴⁶⁸ of search strategies they used in the experiment. Participants were paid a US
⁴⁶⁹ minimal wage and received a performance-based bonus, configured so that on
⁴⁷⁰ average 70% of individuals receive a bonus. The study was approved by our
⁴⁷¹ institutional IRB board. The same procedure was used for both experiments.
⁴⁷² Please see Supplemental Materials 5.5 for an explanation of the bonus scheme
⁴⁷³ and task instructions.

⁴⁷⁴ **3 Experiments**

⁴⁷⁵ **3.1 Experiment 1**

⁴⁷⁶ In the first experiment we aimed to test whether people’s planning follows
⁴⁷⁷ optimal expected utility. The experiment included 40 mazes, presented in
⁴⁷⁸ random order. Of the mazes, 23 were two-room mazes corresponding to
⁴⁷⁹ simple decision trees with two leafs. The two-choice mazes were designed to
⁴⁸⁰ be similar to the experimental paradigms that involve choosing between two
⁴⁸¹ gambles, traditionally used to study subjective utilities [28, 29]. They were
⁴⁸² designed so that the bigger of the two rooms also requires more steps to reach
⁴⁸³ than the smaller room, to bring out differences between human choices, and
⁴⁸⁴ the EU model. In terms of Expected Utility of these binary choice mazes,
⁴⁸⁵ they were designed so that:

- ⁴⁸⁶ (1) The Expected Utility is not predicted by differences in distance or size
⁴⁸⁷ alone, and
- ⁴⁸⁸ (2) The Expected Utility of the bigger and further room is larger or equal to
⁴⁸⁹ that of the closer and smaller room.

⁴⁹⁰ This was done because in the pilot experiments we noticed that people prefer
⁴⁹¹ closer rooms more often than the EU model. See also See Supplementary
⁴⁹² Figure21. Another 4 mazes consisted of 2 rooms, but included a looped
⁴⁹³ corridor, meaning that the decision tree was not strictly binary. The remaining
⁴⁹⁴ mazes consisted of 3-4 rooms. See Supplemental Materials 5.6 for full set of
⁴⁹⁵ mazes used in the experiment.

⁴⁹⁶ We recruited 120 US participants on Amazon Mechanical Turk, of which 4
⁴⁹⁷ were excluded for failing to answer instruction quiz correctly in two attempts.

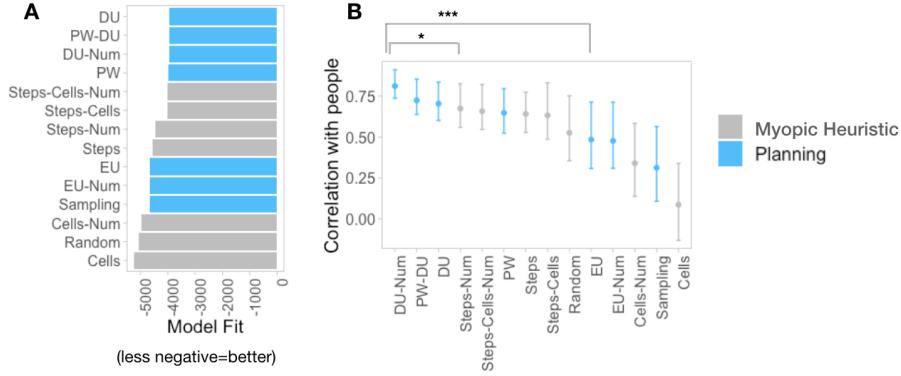


Figure 5: Experiment 1, results. **A.** Model performances, measured as the total log likelihood of each model across all five folds. Shorter bars indicate better fit to human behavior. **B.** Bootstrapped correlations of choice probabilities aggregated across participants with each model's predictions.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

498 In total 116 participants (56 female, 60 male, $M(\text{age}) = 39$, $SD(\text{age}) = 12$)
 499 were included in the analysis. Rerunning the analysis with all participants
 500 included did not change the results. On average the experiment took 14
 501 minutes to complete, with people making on average 51.5 decisions during
 502 this time.

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

508 **3.1.1 Results**

509 **People do not plan according to optimal Expected Utility.** Model
510 performance is shown in Figure 5 A, with the models ordered by their
511 likelihood. The difference in log-likelihood between EU and the best most
512 likely planner DU is $\Delta LL = 764$. Figure 5 B shows the bootstrapped
513 correlations between the probabilities of decisions made by people and the
514 models. Models were parametrised with the mean parameter values fitted to
515 human distribution, shown in Supplementaty Materials. The correlation of
516 the best-performing DU-Num model with people is $r = .81$, ($95CI[.74 - .91]$),
517 and the correlation of EU is $r = .49$, ($95CI[.31 - .71]$). These correlations are
518 significantly different, with bootstrapped difference between their means of
519 [.1 – .5], indicating that the DU-Num model predicts the aggregate population
520 better than the optimal EU model.

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

524 **Evidence in favor of planning over heuristics.** We also found that
525 planning models are better at explaining people’s behavior, compared to
526 myopic heuristics. The difference in log-likelihood between the most likely
527 planner DU and the most likely heuristic Steps-Cells-Num is $\Delta LL = 47$.
528 Bootstrapped correlations between models and people’s choices in Figure 5B
529 show that the 95% CI of bootstrapped difference between correlations of DU-
530 Num (a planner with the highest correlation) and Steps-Num (a heuristic with
531 the highest correlation) is significant, [.01 – .2], suggesting that people’s choices

532 are best explained by models that plan. See Supplementary Information for
533 more analysis, including variability between individuals.

534 **Depth-limited planning.** Below we show a test-case of an environment
535 where depth limited planning can explain human choice preferences, but
536 other models can not. Figures 13 and 14 shows a maze from Experiment 1,
537 with paths predicted by different models. The DU model prefers searching
538 room ‘1’ first, while the other models are indifferent between ‘1’ and ‘2’. This
539 environment has occurred in Experiment 1 twice (see Supplementary Materials
540 for the full list of mazes), in the original form shown in Supplementary Figures
541 13 and 14, and as its mirror symmetric version. In both presentations, the
542 proportion of people who searched the smaller room first was significantly
543 greater than chance ($0.71, \chi^2 = 19.6, df = 1, p < .0001$ and $0.68, \chi^2 =$
544 $14.3, df = 1, p < .0001$). The DU model is able to distinguish between
545 the ‘long’ room, in which in the worst case scenario the exit may be still
546 6 steps away when the room is opened, and the ‘compact’ room, in which
547 the exit will be at most 2 steps away. The EU and PW models can not
548 distinguish between these two directions, as the exit is equally likely to be in
549 either room. Distance and size based step-wise heuristics do not distinguish
550 between these two directions, as both rooms require 5 steps to reach, and
551 both rooms contain 6 unrevealed tiles.

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

557 **3.2 Experiment 2**

558 In the second experiment, we aimed to differentiate between models that
559 plan ahead and myopic heuristics. We followed the experimental procedure
560 from Experiment 1 with a new set of 23 mazes containing between 4 to 10
561 rooms. The mazes were designed so that in the initial decision in each maze
562 heuristics could not distinguish between the available choices, but planning
563 models could. Supplementary Figure 5.7 shows the full set of mazes used
564 in the experiment, illustrating this design. Since each direction of travel
565 during an initial decisions leads to a room of the same size after taking the
566 same number of steps, neither direction is preferable to a heuristic. However,
567 the areas of a maze that become accessible in subsequent decisions differ
568 between directions, meaning that planning models prefer one of the directions
569 over the others. [The full set of mazes used in the experiment is available in](#)
570 [Supplemental Materials 5.7.](#)

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

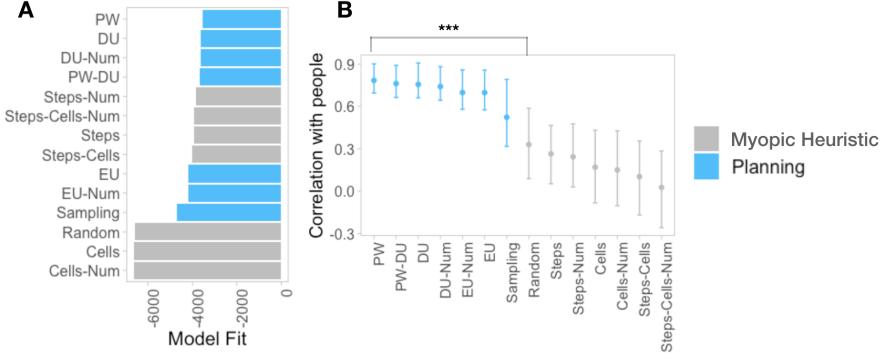


Figure 6: Experiment 2 Results. **A.** Model performances, measured as the total log likelihood of each model across all five folds. Shorter bars indicate better fit to human behavior. **B.** Bootstrapped correlations of choices aggregated across participants with each model’s predictions. Error bars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

578 3.2.1 Results

579 **Evidence in favor of planning over heuristics.** Overall, we found that
 580 planning models are better at explaining people’s behavior compared to
 581 myopic heuristics. The difference in likelihoods between PW and the best-
 582 fitting Steps-Num heuristic is $\Delta LL = 262$. The 95% CI of bootstrapped
 583 difference between correlations of PW (a planner with the highest correlation)
 584 and Steps (a heuristic with the highest correlation) is [0.2 – 0.5], as shown in
 585 Figure 5B, indicating that planning models predict the aggregate population
 586 better than myopic heuristics. See Supplementary Information for more
 587 analysis, including variability between individuals and the distributions of
 588 model parameter values fitted to individuals.

589 **Depth-limited planning.** Below we show that our modeling approach
 590 can predict human decision times. First, we test whether decision times

591 reflect precision of computation, by fitting a linear regression model $time_i =$
 592 $\alpha_0 + \alpha_1 \tau_i + e$. Here α_0 is the Intercept measured in milliseconds, i indexes
 593 people, $\tau_i \in [0, 10]$ is the temperature parameter fitted to people in the EU
 594 model, and $time_i$ is decision time. Since a smaller τ corresponds to closer
 595 approximation of expected utility, for our hypothesis to hold, we expect to
 596 see a negative slope at τ_i for Initial decision times, and a positive slope for
 597 Subsequent decision times. We found the slopes to be significant in line with
 598 this expectation, indicating that τ reflects the extent of planning in maze
 599 search (see Table1).

Times	p	F-statistic	α_0	Slope τ
Initial	< .0001	F(1,2322)=32	3515	-255
Subsequent	< .0001	F(1,6045)=29	905	55

Table 1: Linear regressions predicting decision times by parameters fitted by model EU

600 Next, we test whether human decision times can be explained by limited
 601 planning depth, as measured by the discounting parameter $\gamma \in [0, 1]$ fitted to
 602 people. We compute linear regression of the form $time_i = \alpha_0 + \alpha_1 \gamma_i + \alpha_2 \tau_i + e$,
 603 where γ_i and τ_i are fitted to people in model DU. Since a higher γ corresponds
 604 to more planning and less discounting, by our hypothesis we expect to see
 605 a positive slope at γ_i for Initial decision times, and a negative slope for
 606 Subsequent decision times. We found the regression slopes to be significant
 607 and in line with this expectation, as shown in Table 2, providing further
 608 evidence in support of decision times explained by limited planning depth.

609 A similar regression models for parameters fitted to people in model
 610 DU-Num further support our conclusion. For model DU-Num, regression

Times	p	F-statistic	α_0	Slope τ	Slope γ
Initial	.0001	F(2,2321)=9	2880	-163 $p = .03$	530 $p < .0001$
Subs.	< .0001	F(2,6044)=10.4	1069	16 $p = .3$	-134 $p = .001$

Table 2: Linear regressions predicting decision times by parameters fitted by model DU.

model takes the form $time_i = \alpha_0 + \alpha_1\gamma_i + \alpha_2B_i + \alpha_3\tau_i + e$, where $B_i \in [0, 10]$ is the threshold of bits processed when estimating the numbers of tiles (larger B - more accurate). The results of the fitted regression model in Table 3 show slope coefficients for γ to be significant, supporting the hypothesis that decision times depend on planning depth. We also observe a positive slope associated with the information-theoretic numerosity parameter B , meaning that higher value of B is associated with longer deliberation times in both types of decisions. This result provides further evidence in support of our modeling approach, as it associates longer decision times with more accurate number perception.

Times	p	F-statistic	α_0	Slope B	Slope γ	Slope τ
Initial	< .0001	F(3,2320)=20	2612	88 $p < .0001$	607 $p < .0001$	-340, $p = .02$
Subs.	< .0001	F(3,6043)= 39	951	26.6 $p < .0001$	-192 $p < .0001$	76 $p = .03$

Table 3: Linear regressions predicting decision times by parameters fitted by model DU-Num

In the PW-DU model probabilities are perceived most accurately when $\beta = 1$, while any deviation from this value leads to less exact perception of probabilities, meaning that the relationship between β and probability perception is non-linear. Therefore, one way to compute a regression model

625 for parameters of model PW-DU is by defining a new parameter $b \in [0, 1]$ as
 626 the distance of $\beta \in [0, 2]$ to the optimal value of 1 (smaller b – more optimal).
 627 The results of fitting the regression model $time_i = \alpha_0 + \alpha_1 \gamma_i + \alpha_2 b_i + \alpha_3 \tau_i + e$,
 628 are shown in Table 4, showing slope coefficients for γ to be significant in line
 629 with our hypothesis.

Times	p	F-statistic	α_0	Slope b	Slope γ	Slope τ
Initial	.001	F(3,2320)=5.23	2587	467 $p = .001$	337 $p = .005$	-57, $p = .5$
Subs.	< .0001	F(3,6043)=11.15	985	111 $p=.0006$	-92 $p = .0006$	32 $p = .07$

Table 4: Linear regressions predicting decision times by parameters fitted by model PW-DU

630 In summary, these results suggest that (1) discounting is positively cor-
 631 related with Initial decision times, meaning that longer deliberation results
 632 in a longer planning horizon, (2) discounting is negatively correlated with
 633 Subsequent decision times, meaning that a longer initial deliberation results
 634 in fewer decisions made online and (3) discounting explains decision times
 635 beyond measuring noise in utility computation.

636 4 Discussion

637 We examined how people plan under uncertainty in a spatial setting using a
 638 novel Maze Search Task (MST), which requires people to navigate partially
 639 observable mazes in search of a randomly placed exit. We designed and
 640 evaluated a family of computational models that plan ahead, along with a
 641 family of myopic heuristics that choose the next observation one step at a
 642 time. In addition to limited planning horizon, which has been explored in

643 previous work although not in spatial planning tasks, our planning models
644 incorporate number and probability perception, which were not previously
645 evaluated in the context of planning.

646 We found that people’s decisions in the MST were best explained by
647 models that plan, as opposed to myopic heuristics. Moreover, in contrast
648 to the optimal planning, people were best explained by our cognitively-
649 inspired planners in which intuitive utilities arise from interactions between
650 cognition and perception. We do not argue that one of the planning models is
651 definitively superior at predicting how people plan, as individual differences in
652 the MST suggest that multiple planning strategies may be available to people
653 at the same time (this point is similar to previously reported differences
654 in decision-making [20, 40]). However, we find strong evidence in favor of
655 limited planning horizon being one of the central principles that influence
656 how people plan to search. Consistent with previous studies of non-spatial
657 tasks [3, 12, 2], we found that modifying the optimal Expected Utility by
658 limited planning horizon explains human behavior better than Expected
659 Utility alone in terms of likelihood and correlation analysis. We also show
660 that longer initial deliberation times at the start of the trial, and shorter
661 decision times during navigation, are associated with a longer planning horizon
662 fitted to individuals.

663 Our results demonstrate that planning strategies vary significantly across
664 individuals, and illustrate the strengths of our computational modeling ap-
665 proach that evaluate multiple models for understanding diversity of behavior
666 in cognitive tasks. This result adds to a growing trend in cognitive science
667 literature where large families of models are used to explore the diversity of

behaviors[27, 28]. In this regard, increasing expressiveness of the hypotheses class may be of practical value for developing theories in the cognitive domain. While we examine planning in decision-trees with maximal depth of 10 (corresponding to mazes with up to 10 rooms and an order of 2^{10} states), future studies could investigate how discounting interacts with perceptual approximations in even larger problems as the computational complexity of the problem increases. In particular, several studies show that practice can increase planning horizon in games such as Chess and Four-in-a-row, in terms of the number of nodes explored by a model [14, 15], which suggests that human are so good at planning in realistic spatial tasks, such as navigation, at least partly due to extensive practice. This rises the question of whether practice would similarly improve performance in MST, in terms of planning depth measured as discount rate, and to what extent such an improvement could transfer to other tasks. Additionally, studies have shown that performance in high-level cognitive tasks, such as tasks that require learning, is relatively immune to financial incentive, in contrast to tasks that require a fixed strategy [41]. This result rises the question of how motivation might impact performance in Maze Search, suggesting that increasing motivation could lead to faster reaction times in people who use heuristics, but is unlikely to impact the depth to which people plan. Further, affective states, such as mood and arousal, are known to influence performance across a variety of tasks [42], although the effects of affect of planning are not yet addressed. A computational account of how affective states impact planning in MST could guide design of mental state inference algorithms, and advance our understanding of cognitive deficits in people with mental health disorders. .

693 The planners built in this work do not purport to describe all there is
694 to spatial planning, or to exhaust all possible ways in which people could
695 make choices in MST. Our modeling goal in this work was to contrast a set
696 of naturalistic planning models motivated by prior literature against myopic
697 heuristics motivated by literature as well as by participant feedback received
698 during pilots. While it is possible to imagine more sophisticated heuristics,
699 for example based on symmetries [43] or hierarchical representations [44, 45],
700 such heuristics would serve to reduce complexity of the task rather than
701 support myopic decision-making.

702 While we used analytical models for methodological rigor, future studies
703 can investigate algorithmic models that sample a small number of paths,
704 which could be particularly well suited for large environments. Future studies
705 can also investigate the stability and generalization of individual planning
706 strategies between tasks, and over time. Do people learn to plan by learning
707 a library of task-specific strategies (e.g. Chess strategies [15]), or by learning
708 abstract cognitive mechanisms for planning [46, 13], for example by learning to
709 plan further ahead? Another interesting question is to examine whether people
710 mentally represent probabilities in the same way across spatial and non-spatial
711 contexts, such as, investigating whether probability weighting generalizes
712 between MST and monetary gambles studied by Prospect Theory [29].

713 Lastly, we note several limitations of our modelling approach. In our
714 analysis, all decisions are treated as Markovian, meaning that each decision
715 is independent of whether a certain maze state is encountered as a new maze
716 (as root of a decision tree), or as a part of a partially explored maze (one of
717 the sub-trees). While this assumption is common in computational models

718 of planning, future studies could test whether in practice individuals make
719 consistent choices after partially exploring a maze, compared to when an
720 identical partially revealed maze is presented as a new trial. This consistency
721 (or the lack of it) could indicate to what extent people plan a search policy, or
722 engage in online updating. Further, while our results suggest that probability
723 weighting plays a role in spatial planning, we note an overlap in the scope of
724 behaviors that can be modeled by discounting and by probability weighting,
725 motivating further study of cognitive mechanisms of probability perception
726 in spatial domains.

727 Our work takes a step toward computationally understanding and model-
728 ing an important, real-life domain of planning: how humans plan in spatial
729 multi-step contexts. Our results have implications for the study of human
730 cognition, as well as for building an AI that can interpret human planning to
731 infer goals, offer assistance, and support human-AI collaboration. Our work
732 makes a novel theoretical contribution, by demonstrating evidence of limited
733 planning horizon in spatial planning tasks, and illustrates the value of our
734 multi-model approach for understanding cognition. Beyond our theoretical
735 and empirical contributions, we hope our MST methodology can become a
736 sandbox for exploring a larger variety of planning environments, cognitive
737 models and algorithms.

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

740 **References**

- 741 [1] John von Neumann, Oskar Morgenstern, et al. Theory of games and economic
742 behavior. New York: John Wiley & Sons, 1972.
- 743 [2] Joseph Snider, Dongpyo Lee, Howard Poizner, and Sergei Gepshtain. Prospec-
744 tive optimization with limited resources. PLoS computational biology,
745 11(9):e1004501, 2015.
- 746 [3] Quentin JM Huys, Níall Lally, Paul Faulkner, Neir Eshel, Erich Seifritz,
747 Samuel J Gershman, Peter Dayan, and Jonathan P Roiser. Interplay of
748 approximate planning strategies. Proceedings of the National Academy of
749 Sciences, 112(10):3098–3103, 2015.
- 750 [4] Frederick Callaway, Bas van Opheusden, Sayan Gul, Priyam Das, Paul M
751 Krueger, Thomas L Griffiths, and Falk Lieder. Rational use of cognitive
752 resources in human planning. Nature Human Behaviour, 6(8):1112–1125, 2022.
- 753 [5] Mehdi Keramati, Peter Smittenaar, Raymond J Dolan, and Peter Dayan.
754 Adaptive integration of habits into depth-limited planning defines a habitual-
755 goal-directed spectrum. Proceedings of the National Academy of Sciences,
756 113(45):12868–12873, 2016.
- 757 [6] Marta Kryven, Tomer D Ullman, William Cowan, and Joshua B Tenenbaum.
758 Plans or outcomes: How do we attribute intelligence to others? Cognitive
759 Science, 45(9):e13041, 2021.
- 760 [7] Samuel J Cheyette and Steven T Piantadosi. A unified account of numerosity
761 perception. Nature human behaviour, 4(12):1265–1272, 2020.
- 762 [8] Drazen Prelec. The probability weighting function. Econometrica, pages
763 497–527, 1998.
- 764 [9] Amir-Homayoun Javadi, Beatrix Emo, Lorelei R Howard, Fiona E Zisch, Yichao
765 Yu, Rebecca Knight, Joao Pinelo Silva, and Hugo J Spiers. Hippocampal and

- 766 prefrontal processing of network topology to simulate the future. Nature
767 communications, 8(1):1–11, 2017.
- 768 [10] Shanjiang Zhu and David Levinson. Do people use the shortest path? an
769 empirical test of wardrop’s first principle. PLoS one, 10(8):e0134322, 2015.
- 770 [11] Christian Bongiorno, Yulun Zhou, Marta Kryven, David Theurel, Alessan-
771 dro Rizzo, Paolo Santi, Joshua Tenenbaum, and Carlo Ratti. Vector-based
772 pedestrian navigation in cities. Nature Computational Science, 1(10):678–685,
773 2021.
- 774 [12] Quentin JM Huys, Neir Eshel, Elizabeth O’Nions, Luke Sheridan, Peter Dayan,
775 and Jonathan P Roiser. Bonsai trees in your head: how the pavlovian system
776 sculpts goal-directed choices by pruning decision trees. PLoS computational
777 biology, 8(3):e1002410, 2012.
- 778 [13] Yash Raj Jain, Frederick Callaway, and Falk Lieder. Measuring how people
779 learn how to plan. In CogSci, pages 1956–1962, 2019.
- 780 [14] Bas van Opheusden, Ionatan Kuperwajs, Gianni Galbiati, Zahy Bnaya, Yunqi
781 Li, and Wei Ji Ma. Expertise increases planning depth in human gameplay.
782 Nature, pages 1000–1005, 2023.
- 783 [15] Diogo R Ferreira. The impact of the search depth on chess playing strength.
784 ICGA journal, 36(2):67–80, 2013.
- 785 [16] Adriaan D De Groot. Thought and choice in chess. In Thought and Choice in
786 Chess. De Gruyter Mouton, 2008.
- 787 [17] Fernand Gobet and Herbert A Simon. Templates in chess memory: A mecha-
788 nism for recalling several boards. Cognitive psychology, 31(1):1–40, 1996.
- 789 [18] Josef M Unterrainer, Benjamin Rahm, Christoph P Kaller, Rainer Leonhart,
790 K Quiske, K Hoppe-Seyler, C Meier, C Müller, and Ulrike Halsband. Planning

- 791 abilities and the tower of london: is this task measuring a discrete cognitive
792 function? Journal of clinical and experimental neuropsychology, 26(6):846–856,
793 2004.
- 794 [19] Dennis H Holding. The psychology of chess skill. Routledge, 1985.
- 795 [20] Yash Raj Jain, Frederick Callaway, Thomas L Griffiths, Peter Dayan, Ruiqi He,
796 Paul M Krueger, and Falk Lieder. A computational process-tracing method
797 for measuring people’s planning strategies and how they change over time.
798 Behavior Research Methods, 55(4):2037–2079, 2023.
- 799 [21] Richard S Sutton and Andrew G Barto. Reinforcement learning: An introduc-
800 tion. Robotica, 17(2):229–235, 1999.
- 801 [22] Abhishek Naik, Roshan Shariff, Niko Yasui, Hengshuai Yao, and Richard S
802 Sutton. Discounted reinforcement learning is not an optimization problem.
803 arXiv preprint arXiv:1910.02140, 2019.
- 804 [23] Jiri Najemnik and Wilson S Geisler. Optimal eye movement strategies in visual
805 search. Nature, 434(7031):387–391, 2005.
- 806 [24] Björn Meder, Jonathan D Nelson, Matt Jones, and Azzurra Ruggeri. Stepwise
807 versus globally optimal search in children and adults. Cognition, 191:103965,
808 2019.
- 809 [25] Doug Markant and Todd Gureckis. Does the utility of information influence
810 sampling behavior? In Proceedings of the annual meeting of the cognitive
811 science society, volume 34, 2012.
- 812 [26] Jonathan D Nelson. Finding useful questions: on bayesian diagnosticity,
813 probability, impact, and information gain. Psychological review, 112(4):979,
814 2005.

- 815 [27] F Callaway, B Van Opheusden, S Gul, P Das, P Krueger, F Lieder, and
816 T Griffiths. Human planning as optimal information seeking. Manuscript in
817 preparation, 2021.
- 818 [28] Kai Ruggeri, Sonia Alf, Mari Louise Berge, Giulia Bertoldo, Ludvig D Bjørndal,
819 Anna Cortijos-Bernabeu, Clair Davison, Emir Demić, Celia Esteban-Serna,
820 Maja Friedemann, et al. Replicating patterns of prospect theory for decision
821 under risk. Nature human behaviour, 4(6):622–633, 2020.
- 822 [29] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative
823 representation of uncertainty. Journal of Risk and uncertainty, 5(4):297–323,
824 1992.
- 825 [30] Joshua C Peterson, David D Bourgin, Mayank Agrawal, Daniel Reichman,
826 and Thomas L Griffiths. Using large-scale experiments and machine learning
827 to discover theories of human decision-making. Science, 372(6547):1209–1214,
828 2021.
- 829 [31] Nathaniel D Daw, Samuel J Gershman, Ben Seymour, Peter Dayan, and
830 Raymond J Dolan. Model-based influences on humans’ choices and striatal
831 prediction errors. Neuron, 69(6):1204–1215, 2011.
- 832 [32] Maurice Allais. Le comportement de l’homme rationnel devant le risque:
833 critique des postulats et axiomes de l’école américaine. Econometrica: Journal
834 of the Econometric Society, pages 503–546, 1953.
- 835 [33] R Duncan Luce. Individual choice behavior: A theoretical analysis. Courier
836 Corporation, 2012.
- 837 [34] Kenway Louie, Lauren E Grattan, and Paul W Glimcher. Reward value-based
838 gain control: divisive normalization in parietal cortex. Journal of Neuroscience,
839 31(29):10627–10639, 2011.

- 840 [35] William Stanley Jevons. The theory of political economy. Macmillan and
841 Company, 1879.
- 842 [36] Susannah K Revkin, Manuela Piazza, Véronique Izard, Laurent Cohen, and
843 Stanislas Dehaene. Does subitizing reflect numerical estimation? Psychological
844 science, 19(6):607–614, 2008.
- 845 [37] Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas,
846 Peter I Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon
847 Samothrakis, and Simon Colton. A survey of monte carlo tree search methods.
848 IEEE Transactions on Computational Intelligence and AI in games, 4(1):1–43,
849 2012.
- 850 [38] Jeremy N Bailenson, Michael S Shum, and David H Uttal. The initial segment
851 strategy: A heuristic for route selection. Memory & Cognition, 28(2):306–318,
852 2000.
- 853 [39] Steven T Piantadosi. One parameter is always enough. AIP Advances,
854 8(9):095118, 2018.
- 855 [40] Joshua Peterson, Marina Mancoridis, and Tom Griffiths. To each their own
856 theory: Exploring the limits of individual differences in decisions under risk. In
857 Proceedings of the Annual Meeting of the Cognitive Science Society, volume 45,
858 2023.
- 859 [41] Pamela J Osborn Popp, Ben Newell, Daniel Bartels, and Todd Gureckis. Can
860 cognitive discovery be incentivized with money? PsyArXiv, 2023.
- 861 [42] Roey Schurr, Daniel Reznik, Hanna Hillman, Rahul Bhui, and Samuel J
862 Gershman. Dynamic computational phenotyping of human cognition. Nature
863 Human Behaviour, pages 1–15, 2024.

- 864 [43] Sugandha Sharma, Aidan Curtis, Marta Kryven, Josh Tenenbaum, and Ila
865 Fiete. Map induction: Compositional spatial submap learning for efficient
866 exploration in novel environments. [arXiv preprint arXiv:2110.12301](#), 2021.
- 867 [44] Momchil S Tomov, Samyukta Yagati, Agni Kumar, Wanqian Yang, and
868 Samuel J Gershman. Discovery of hierarchical representations for efficient
869 planning. [PLoS computational biology](#), 16(4):e1007594, 2020.
- 870 [45] Carlos G Correa, Mark K Ho, Frederick Callaway, Nathaniel D Daw, and
871 Thomas L Griffiths. Humans decompose tasks by trading off utility and
872 computational cost. [PLOS Computational Biology](#), 19(6):e1011087, 2023.
- 873 [46] Marcos Economides, Zeb Kurth-Nelson, Annika Lübbert, Marc Guitart-Masip,
874 and Raymond J Dolan. Model-based reasoning in humans becomes automatic
875 with training. [PLoS computational biology](#), 11(9):e1004463, 2015.
- 876 [47] Chris Baker, Rebecca Saxe, and Joshua Tenenbaum. Bayesian theory of mind:
877 Modeling joint belief-desire attribution. In [Proceedings of the annual meeting](#)
878 [of the cognitive science society](#), volume 33, 2011.
- 879 [48] Marta Kryven, Tomer D Ullman, William Cowan, and Josh Tenenbaum. Out-
880 come or strategy? a bayesian model of intelligence attribution. In [CogSci](#),
881 2016.

882 **5 Supplemental Materials**

883 **5.1 Planning state-space**

884 We based our computational approach on choosing a path through a decision
885 tree that maps states to specific observations within a maze. This design is
886 supported by an empirical evidence that people take direct routes between
887 observations (see Figure 8), even through the grid-world layout of MST in
888 principle allows players to take indirect routes, or even indefinitely move
889 between any adjacent empty tiles without making any observations. People’s
890 tendency to take direct routes between observations strongly suggests that
891 people use efficient problem representations, in line with previous work
892 [38]. Our decision tree model is further supported by the empirical distribution
893 of human decision times in different types of tiles inside a maze, as shown in
894 Figure 7. The figure shows that people move quickly when travelling between
895 observations (in Corridors) and take longer to make a move whenever new
896 hidden tiles are revealed (Decision). The longest decision time occurs at the
897 initial starting location (Start), where people may study the map and plan
898 their path before moving. Motor errors, that is, overshooting the end of a
899 step sequence that connects observations by a step, and immediately going
900 back, occur on about 1% of corridor moves (see Figure 8).

901 Given this empirical evidence, we do not consimodel-free assumptions
902 such as treating each grid cell as a state and using RL to solve the task,
903 since we feel that a richer state-space afforded by model-free representations
904 would not add explanatory power toward modeling humans. In the context
905 of Maze Search, clicking on consecutive corridor cell elicits a variable amount

906 of effort that is perceived as cost, rather than as a state transition. In terms
 907 of optimally solving the task, our model performs as well as any model-free
 908 approach, since we construct decision that include all possible optimal paths
 909 between observation locations.

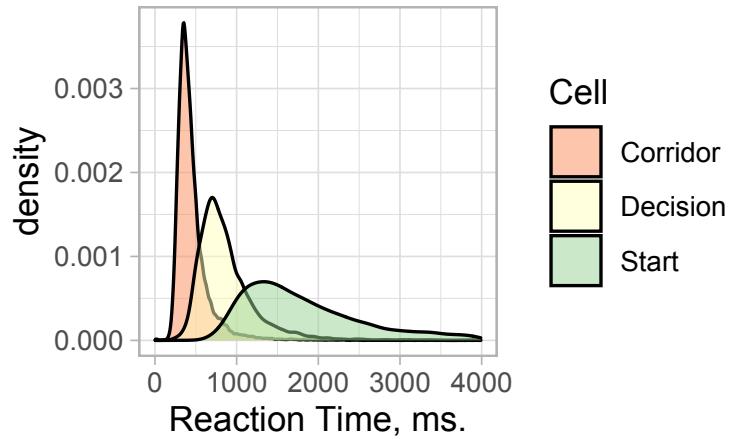


Figure 7: Distribution of decision times in Experiment 1, measured in milliseconds, as humans click on maze tiles, plotted by location type: Start - initial decision time to make the first move at the start of the trial, during which people form a mental representation of a maze and plan a search within a certain planning horizon (or choose where to observe next using a step-wise heuristic); Corridor - moving between observations; Decision - subsequent decision time, corresponding to inner nodes of the decision tree, where people either make a pre-planned move, or decide where to observe next using a step-wise heuristic. For clarity, only times under 4000ms are shown, as the longest decision times in distributions' tails take minutes.

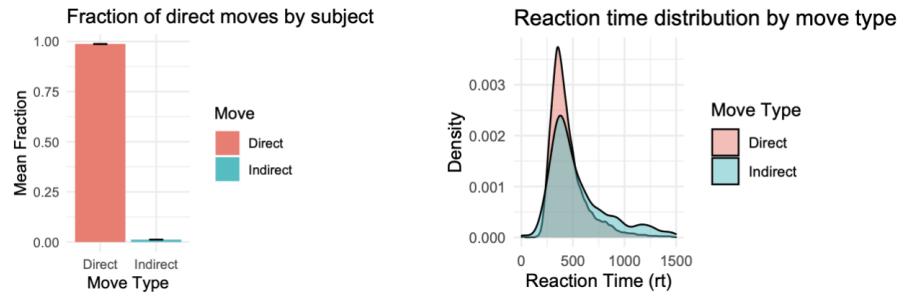


Figure 8: Distribution of decision times in corridors while moving toward the next observation by an optimal route (Direct) and while deviating from the optimal path (Indirect). Of the corridor steps, 99% lie on optimal routes to room-revealing states, and the remaining 1% are motor errors.

910 **5.2 Planning Models**

911 **5.2.1 Probability Weighted Utility**

912 We used probability weighting of the form $p = \exp(-1(-\log(p))^\beta)$, the shape
913 this function takes for different β is shown in Fig.9. The original Prospect
914 Theory sets $\beta \in [0, 1]$. Here we also consider $\beta \in [0, 2]$ where values of
915 $\beta \in (1, 2]$ imply overweighting large probabilities.

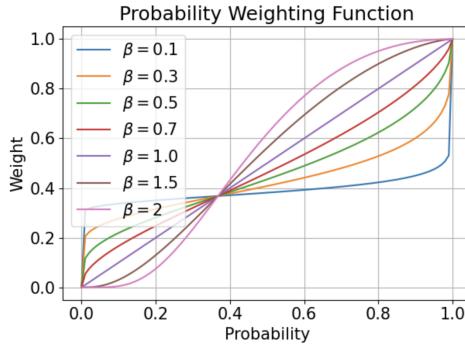
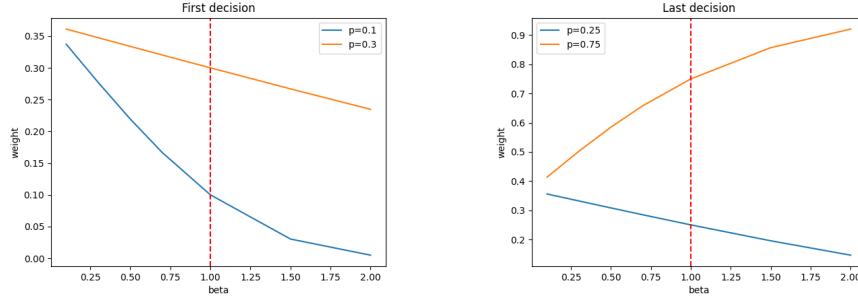


Figure 9: The probability weighting function for different β .

916 Probability Weighted with $\beta < 1$ makes room of different size appear
917 to be more alike, and with $\beta > 1$ makes room of different size appear to
918 be more different. For example, suppose that a maze with $10n$ unobserved
919 cells includes rooms of sizes n and $3n$, and suppose that a person searches
920 this maze exhaustively, so that the first decision (there are still a number
921 unobserved rooms in the maze), and the last decision (only two unobserved
922 rooms remain) both requires choosing between rooms of size n and $3n$. In
923 the first decision the probabilities associated with the rooms are $(0.1, 0.3)$.
924 In the last decision they are $(0.25, 0.75)$. Then, different values of β will
925 transform these probabilities as follows:

926

927

5.2.2 Numerosity perception in Maze Search Task

928 The information-theoretic numerosity model assumes that the systematic
 929 deviation in the perceived number is due to processing limited bits of in-
 930 formation, where the prior distribution over observed numbers favors small
 931 quantities. Figure 10 shows how the number of tiles perceived by people is
 932 predicted to change with the number of bits of information processed.

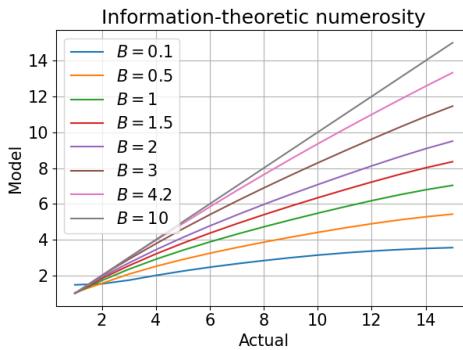


Figure 10: Number of tiles perceived under the model plotted against the actual number.

933 A indirect effect of numerosity model, is that using the perceived numbers
 934 of tiles within the value models also distorts probability perception. This
 935 happens because probabilities are computed as ratios of revealed to remaining

tiles. For example, Figure 11 shows subjective probabilities estimated using different B , assuming a total of 15 unobserved tiles, and various sizes of rooms. The effect of this distortion is different from the effect of the PW model, as modeling Numerosity slightly inflates the small and middle probabilities, while leaving large probabilities unchanged. However, the range of this distortion is also fairly small, compared to effects that can be modeled by the PW function.

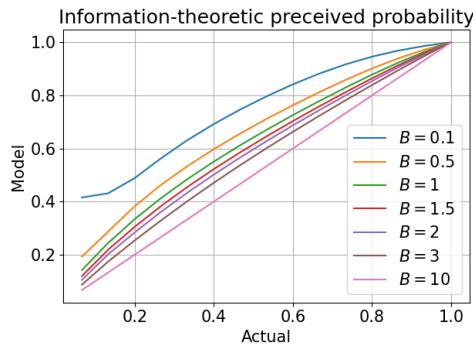


Figure 11: Perceived probability predicted by information-theoretic numerosity model plotted against the actual probability, assuming the exit is equally likely to be in any hidden tile.

943 5.2.3 Monte-Carlo Tree Search - Sampling Model

Algorithm 1: Monte Carlo Tree Search for MST

Data: Tree to determine best child of node N

```

1 Initialize: Decision node  $N$ ,  $Budget$ 
2  $k \leftarrow 1$  while  $k \leq Budget$  do
3    $n \leftarrow N$ 
4   while node  $n$  is not a leaf node do
5     |  $n \leftarrow \max_{c \in C(n)} ubc(c)$ 
6   end
7   if node  $n$  has been visited then
8     | add children nodes of node  $n$ 
9     |  $c \leftarrow$  random child of  $c$ 
10  end
11  while  $p \sim \text{Uniform}(0,1) > P(\text{exit found at } c)$  do
12    |  $c \leftarrow$  random child of  $c$ 
13  end
14  value of node  $n \leftarrow$  value of node  $n + c$ 
15  number of visits to node  $n \leftarrow 1 +$  number of visits to
     node  $n$ 
16   $k \leftarrow k + 1$ 
17 end

```

944 The Sampling model approximates the EU model with increasing accuracy
 945 as the budget parameter is increased. Figure 12 shows the probabilities of
 946 actions predicted by the EU and Sampling model, correlation $r = 0.95$,
 947 assuming budget and exploration parameters are chosen to maximize the
 948 correlation, and EU is parameterized with $\tau = 1$. The optimal exploration
 949 parameter depends on the budget, with larger budgets typically requiring
 950 higher exploration parameter to achieve best approximation.

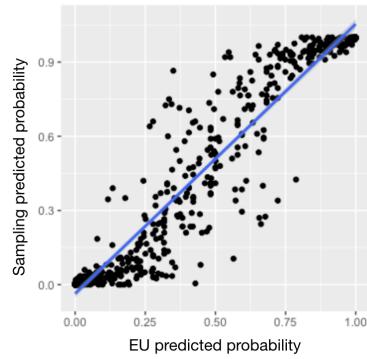


Figure 12: Correlation between action probabilities predicted by EU and Sampling.

951

952 **5.2.4 Defining the problem using the POMDP formalism**

953 While we do not solve Maze Search by Reinforcement Learning, it is possible
954 to define the problem using a POMDP formalism, as we show below. This
955 formulation is closely related to formulations of [47] and [48].

956 The agent occupies a discrete state space X of cells in a 2D grid, where
957 each cell can contain a visible wall, visible floor, unobserved floor (which could
958 hide the Exit), or visible exit. The environment state Y is the set of possible
959 assignments of exit to the unobserved floor cells. Possible actions include
960 North, South, East, and West. Valid actions yield the intended transition
961 with probability 1 and do nothing otherwise; invalid actions (e.g., moving
962 into walls) have no effect on the state.

963 The agent has a 360 degree vision, and can see any cells not occluded by
964 walls. The algorithm that determines cell visibility uses ray-casting: a cell is
965 visible, if the rays cast from the agent's location to any part of the cell do not
966 intersect walls. In other words, agent's visual observations are represented
967 by the discrete grid isovist from the agent's location: a set of previously not
968 visible cells, such that every point in a given cell is visible from any point in
969 the cell where the agent is located. The observation distribution $P(o|x, y)$
970 encodes which environments in Y are consistent with the contents of the
971 isovist from location x . Bayesian belief updating at time t is a deterministic
972 function of the prior belief b_{t-1} , the observation o_t , and the world state
973 $\langle x_t, \mathbf{y} \rangle$. The agent's updated degree of belief in environment y satisfies
974 $b_t(y) \propto P(o_t|x_t, \mathbf{y})b_{t-1}(\mathbf{y})$.

975 The agent's reward function $R(x, y, a)$ encodes the subjective utility the
 976 agent derives from taking action a from the state $\langle x_t, \mathbf{y} \rangle$, where rewards
 977 result from reaching the exit and are the same in each maze, and the costs
 978 are incurred by taking actions (e.g. for the optimal agent, an action that
 979 moves one step incurs a cost of 1). The agent's POMDP is defined by the
 980 state space, the action space, the world dynamics, the observation model,
 981 and the reward function. The agent's policy is stochastic, given by the
 982 softmax of the lookahead state-action value function Q_{LA} : $P(a|b, x, y) \propto$
 983 $\exp(\tau Q_{LA}(b, x, y, a))$. The τ parameter establishes the degree of determinism
 984 with which the agent executes its policy, capturing the intuition that agents
 985 tend to but do not always follow the optimal policy. Here LA can be
 986 represented by different cost models that may include additional parameters.
 987 The value function computes the expected utility of an action in terms of
 988 how many steps away the agent is from the expected location of the exit as a
 989 result of taking this action:

$$990 \quad Q(a, \langle x_t, \mathbf{y} \rangle) = \sum_X P(Y)_t R(x, y, a) + \gamma \max_{a_i \in A} \{Q(a_i, \langle x_{t+1}, \mathbf{y} \rangle)\}$$

991 where γ is a discount factor, similar to discounting used in our original
 992 formulation, but conceptually different as it will scale the paths between
 993 observations proportionally to their length. Finally, the reward function can
 994 be defined to depend on the size of the observed areas and distances between
 995 them, assuming a step cost function $step(x, y, a)$ and a reward for observing
 996 a given area given by a function $cells(x, y, a)$

$$997 \quad R_{EU}(x, y, a) = cells(x, y, a) - step(x, y, a)$$

998 **Examples of value functions behaving differently . Figures 13-20**
 999 show mazes where different models take different paths through a maze. In
 1000 Figure 15 the optimal EU model makes the initial decision to move toward
 1001 the bigger room, taking 5 steps. In contrast, model PW with a small β sees
 1002 rooms of different size as similar, meaning that this model goes to one of the
 1003 closer rooms, taking 4 steps.

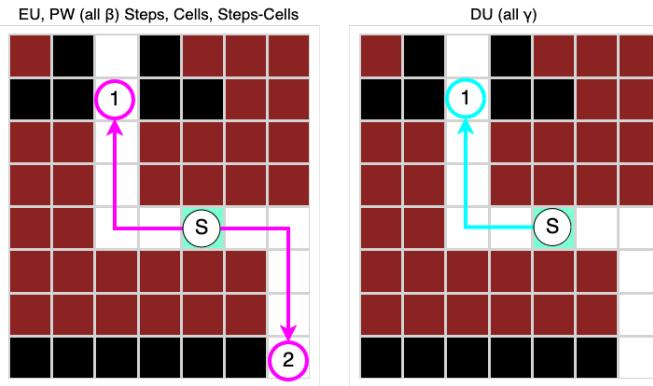


Figure 13: A maze from Experiment 1, with paths predicted by different models. The DU model prefers searching room '1' first, while the other models are indifferent between '1' and '2'.

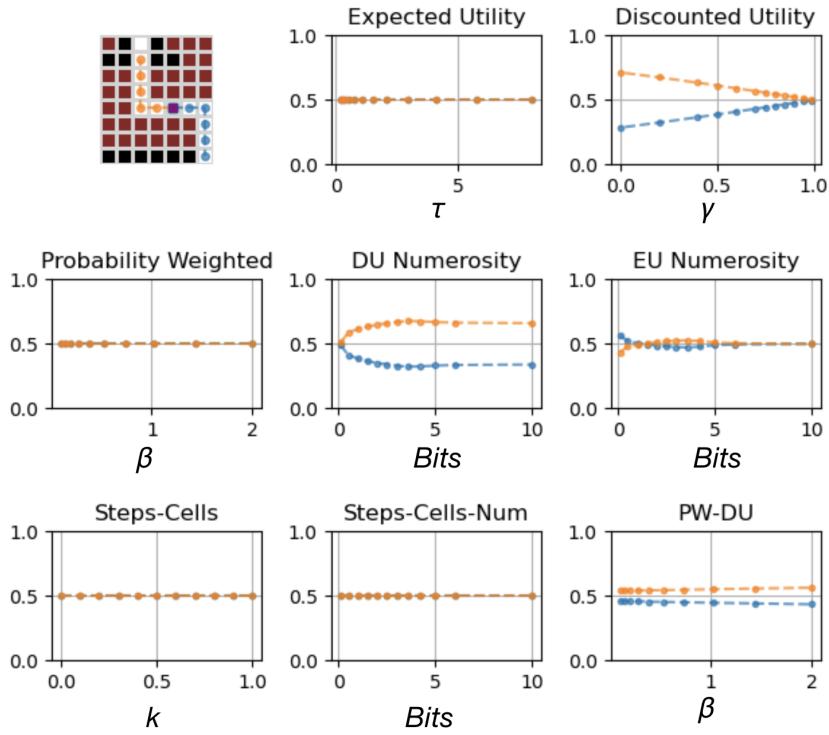


Figure 14: Models predicting different paths in maze from Figure 13. Only models parameterized with discounting can capture human preference for going up. The X axes show model parameters. The y-axis shows probabilities assigned by different models to path directions. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment.

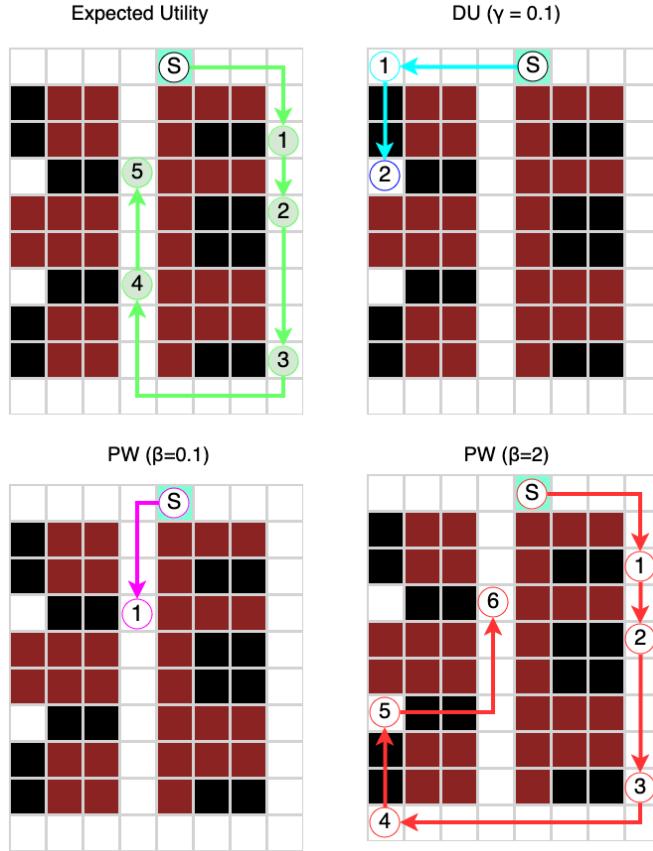


Figure 15: An example of a maze from Experiment 2, in which different models predict different path through a maze.

1004 The difference between the DU and the PW models in this example comes
 1005 from a boundary case, where the DU model becomes indifferent between two
 1006 equidistant rooms when $\gamma - > 0$, as illustrated in Figure 16.
 1007 The difference between the WU and the PW ($\beta = 2$) models in this
 1008 example comes from a downstream decision where the EU model prefers to
 1009 go Up as it prioritizes a shorter path to the next observation, while the PW
 1010 model is indifferent between Up and Left directions, as illustrated in Figure

1011 17.

1012 Figure 18 shows another example of a maze where four models take
1013 different routes, with probabilities assigned to actions by different models
1014 during specific decision shown in Figures 19 and 20.

1015 Figure 20 illustrates an interesting property of the DU model, where
1016 given a small γ the model prefers *a longer path toward a smaller room*. This
1017 happens because the four cells that the DU chooses to open are relatively
1018 closer to the agent – if the exit is found within those four cells, then it will
1019 be on average 2 steps away. In contrast, if the exit is found within the seven
1020 cells on top of the maze (this choice is preferred by all other models), it will
1021 be on average 4 cells away. In this sense, the DU model can behave similarly
1022 to the Steps heuristic, while also being sensitive to the shape of the revealed
1023 rooms.

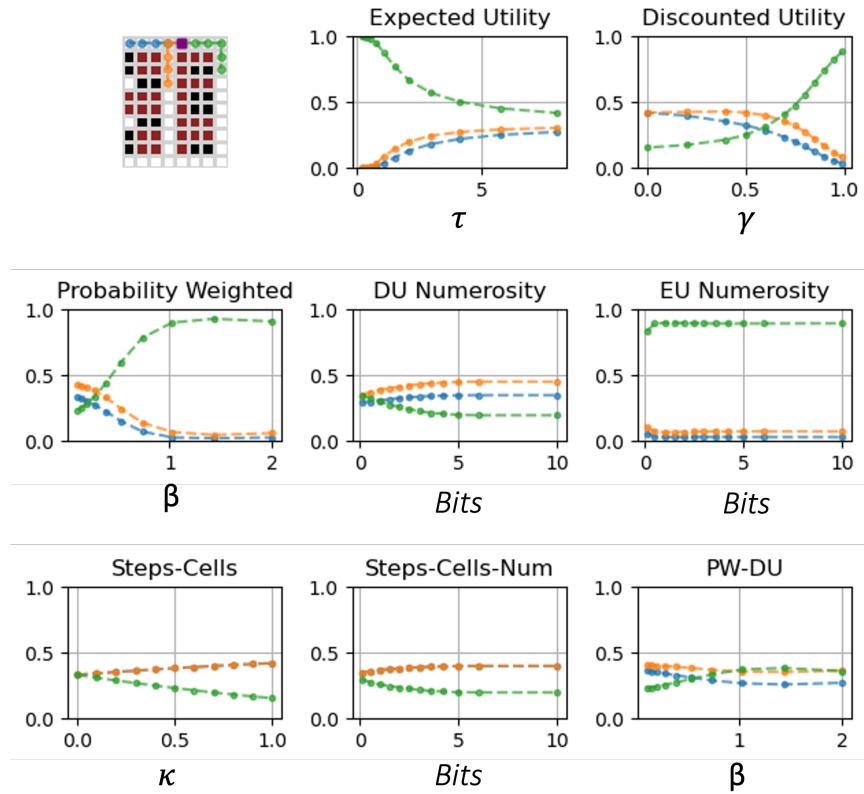


Figure 16: Different directions have a different likelihood of being chosen by the different models, based on parameters of the models. Each graph shows the likelihood of choosing each direction by a different model. The X axes show model parameters. The y-axis shows probabilities assigned by models to the actions. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment.

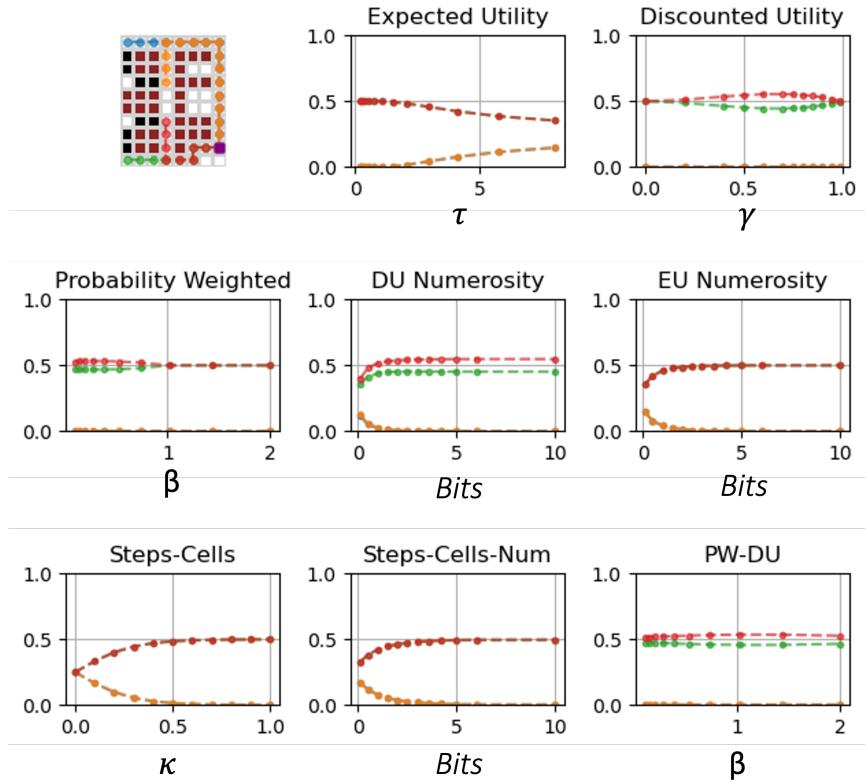


Figure 17: Different directions have a different likelihood of being chosen by the different models, based on parameters of the models. Each graph shows the likelihood of choosing each direction by a different model. The X axes show model parameters. The y-axis shows probabilities assigned by models to the actions. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment.

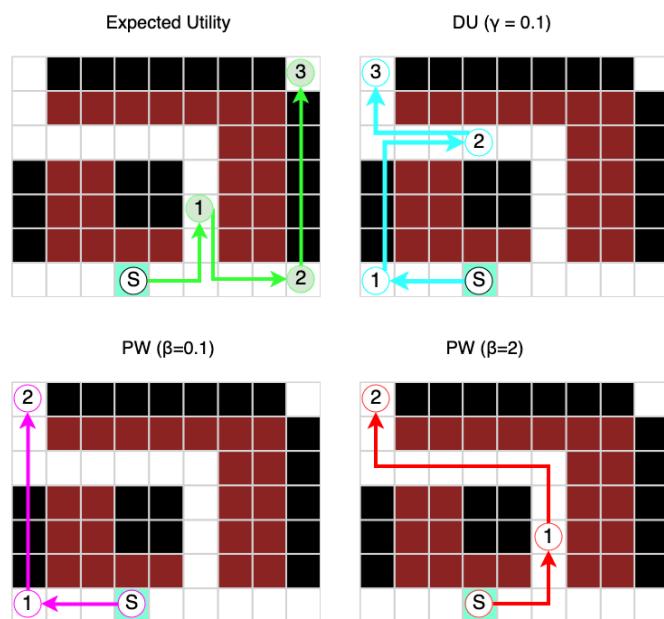


Figure 18: An example of a maze from pilot experiments, in which different models predict different path through a maze.

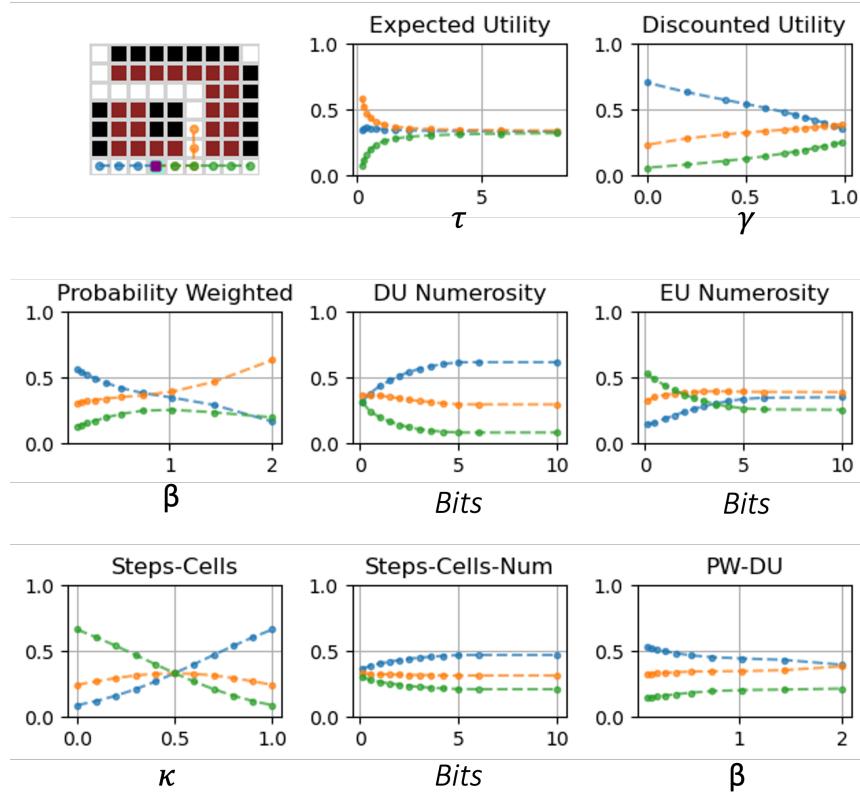


Figure 19: Likelihood of different directions being chosen by the different models, based on model parameters. This figure shows the initial decision, seen by all models. The X axes show model parameters. The y-axis shows probabilities assigned by different models to path directions. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment.

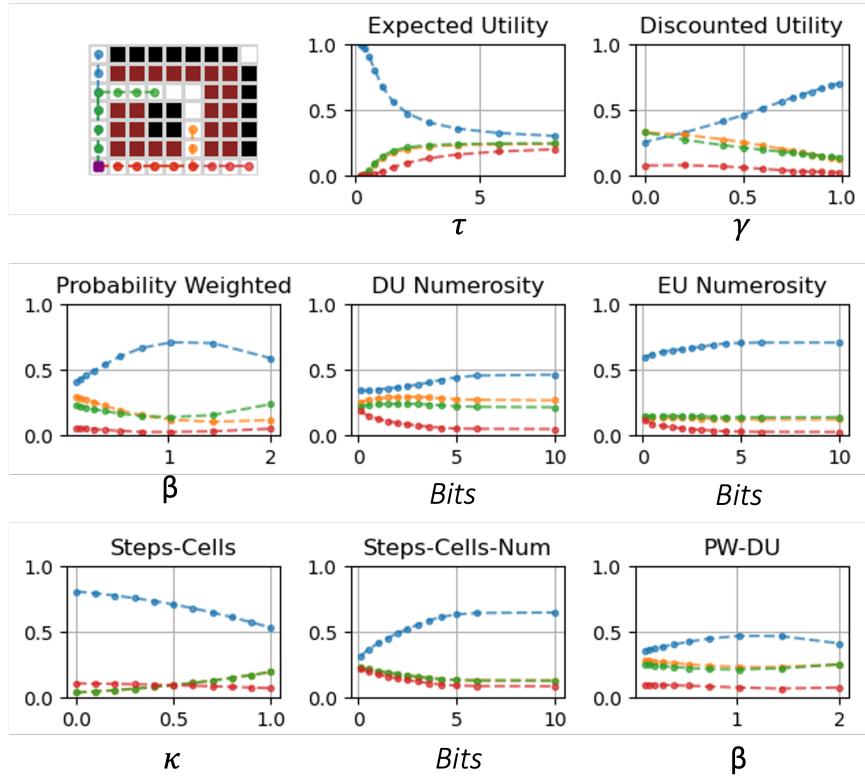


Figure 20: Likelihood of different directions being chosen by the different models, based on model parameters. This figure shows a second a decision in the path. Only models PW ($\beta = 0.1$) and DU ($\gamma = 0.1$) reach this decision. The X axes show model parameters. The y-axis shows probabilities assigned by different models to path directions. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment.

1024 **5.3 Experiment 1, Additional results**

1025 **Binary choice mazes** Experiment 1 includes 23 mazes with strictly binary
1026 trees (mazes with two observations, and no loops). These mazes were designed
1027 to contrast room distance and room size, such that bigger rooms require more
1028 steps to reach them. In terms of Expected Utility of these choices, the mazes
1029 are designed so that

1030 (1) expected value can not be predicted from distance or size differences
1031 between rooms

1032 (2) The Expected Utility of the bigger and further room is larger or equal to
1033 that of the closer and smaller room.

1034 The second point is important, as people tend to prefer closer rooms more
1035 often than the EU model (see, for example, Figure 22 showing EU and
1036 human preferences for the 23 binary mazes from Experiment 1). The plot
1037 below is generated with a minimal decision temperature, to illustrate how the
1038 preferences of the EU model relate to differences in room distance and size.

1039 To illustrate how step and size differences are not sufficient to explain
1040 expected value, consider the two examples below. In both examples, the
1041 differences between size and distance are 5 cells and 5 steps, but EU prefers
1042 a larger room in one case, and is indifferent between the two directions in
1043 the other case in the other case.

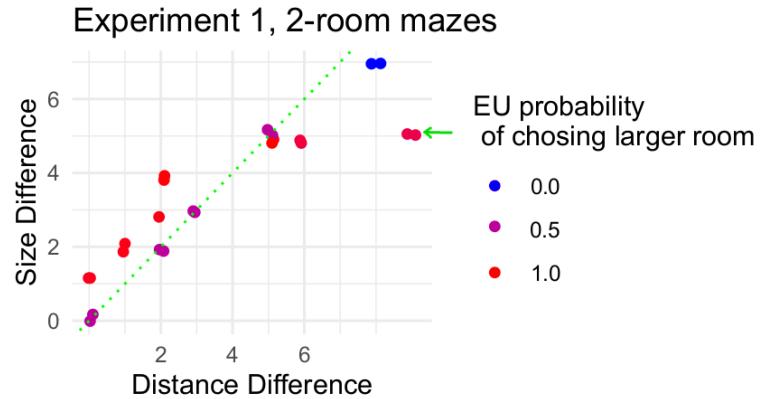


Figure 21: Expected Utility of binary choices in Experiment1. The figure is generated with a minimal decision temperature, to illustrate the preferences of the EU model. The points are jittered for readability. The arrow points to a maze with the largest difference in distance between rooms, additionally shown in Figure 23

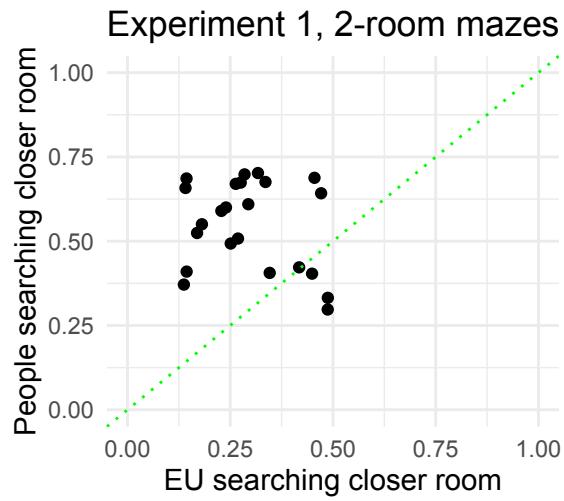


Figure 22: EU and human preferences for the 23 binary mazes from Experiment 1. Compared to EU, people are more likely to search the closer room first.

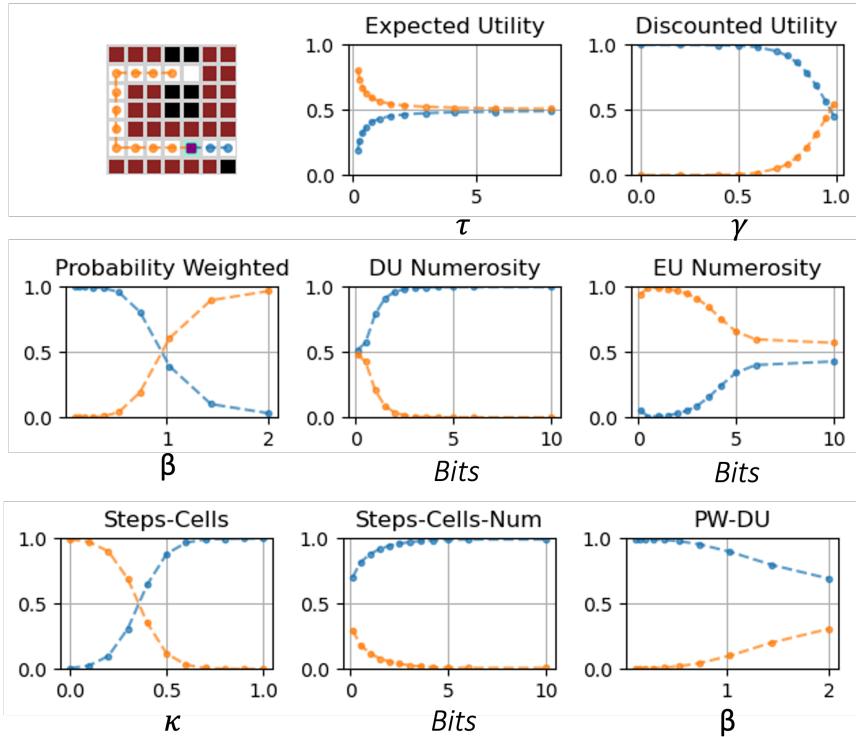


Figure 23: A maze from Experiment 1 where the EU model prefers to search the larger room first, but 0.67 of participants preferred to search the smaller room first. Different plots show probabilities of choosing either path by different models. The X axes show model parameters. The y-axis shows probabilities assigned by different models to path directions. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment.

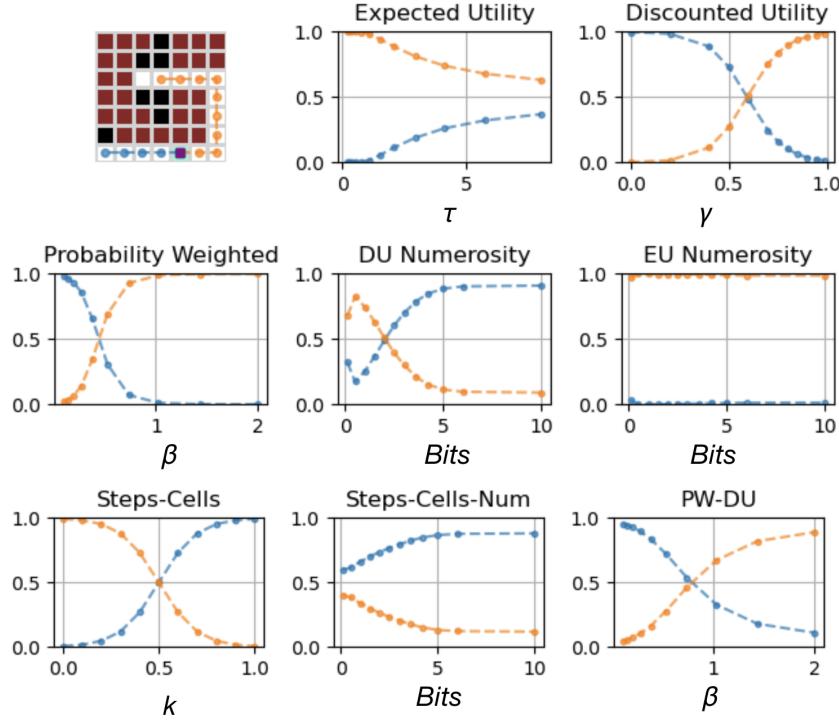


Figure 24: A maze from Experiment 1 where the EU model prefers to search the larger room first, in agreement with 0.58 of participants. Different plots show probabilities of choosing either path by different models. The X axes show model parameters. The y-axis shows probabilities assigned by different models to path directions. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment.

1044 In the figure below, EU prefers to search the room with 6 cells, vs the
 1045 room with 1 cell (difference of 5 cells). The distances to these rooms are
 1046 4 and 9 steps (difference of 5 steps). Here, 0.58 of participants explore the
 1047 larger room, in agreement with EU:

1048 In the figure below, expected utility of both directions is the same. The
1049 rooms contain 2 and 7 cells respectively (difference of 5 cells). The distances
1050 to these rooms are 2 and 2 steps (difference of 5 steps). Here, only 0.4 of
1051 participants explore the larger room.

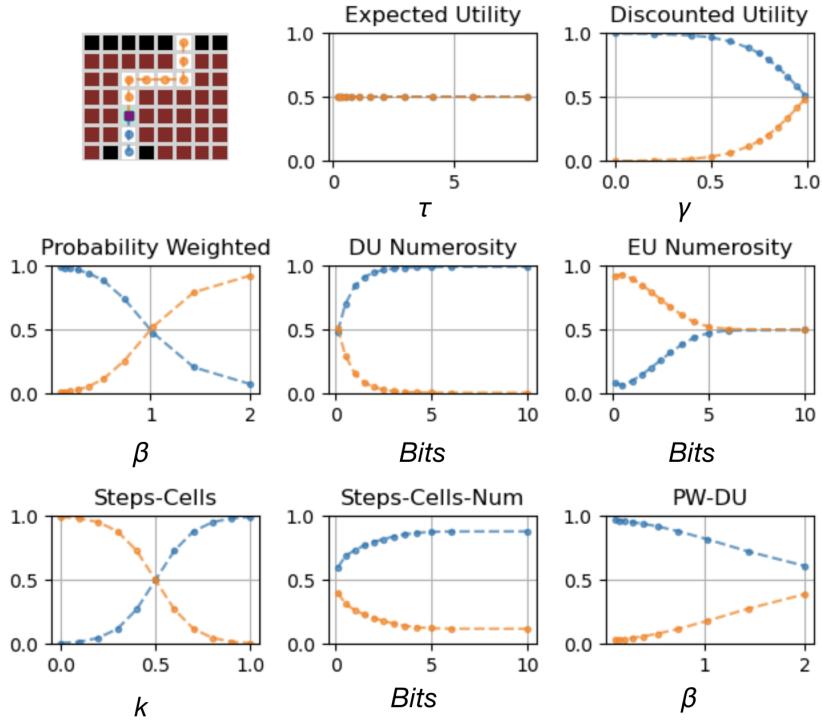


Figure 25: A maze from Experiment 1 where the expected utility of both directions is the same, but 0.6 of participants have a preference for searching the smaller room first. Different plots show probabilities of choosing either path by different models. The X axes show model parameters. The y-axis shows probabilities assigned by different models to path directions. For models with several parameters, the values of parameters not shown are fixed to the participant mean in the experiment.

1052 **Fitting models with Monte-Carlo Cross validation**

1053 Monte-Carlo cross-validation is a method of fitting models over multiple
1054 bootstrapped iterations (we run 100) which allows us to obtain confidence
1055 intervals on each individual's LLs. Figure 26 A. shows the mean LL per
1056 decision as model fit, with 95% CI over participants. Figure 26 B. illustrates
1057 variability between individuals. For each individual we first determine the
1058 best fitting planner and the best-fitting heuristic, and plot the LL of best
1059 heuristic and best planning model for each individual with 95% CI. The labels
1060 "Heuristic", "Planning", and "Not defined" are assigned based on whether
1061 the 95% CI of the LL for best-fitting planner and heuristic overlap. Note that
1062 this method may overestimate the number of "Not defined" individuals, as we
1063 bootstrap the 95% CI of the mean, not the 95% CI of the *difference* between
1064 means. Figure 26 C. summarizes the number of individuals presented in each
1065 category in Figure 26 B. in a bar plot. The distribution of best fitting models
1066 across individuals is shown in Figure 28.

1067 **Computing correlations**

1068 Correlation analysis presented in the main text focuses on decisions visited
1069 by all participants, so that each data-point used in computing correlation is
1070 based on the same population of people. In Figure 27 we show an alternative
1071 analysis, aimed to maximize the number of data-points used to compute
1072 correlation. Figure 27 shows correlations computed using decisions that were
1073 visited by at least 20% of participants.

1074 Here, the correlation of the best-performing DU-Num model with peo-
1075 ple is $r = .89(95CI[.86, .92])$, and the correlation of the optimal EU is
1076 $r = .72(95CI[.64, .82])$. These correlations are significantly different, with

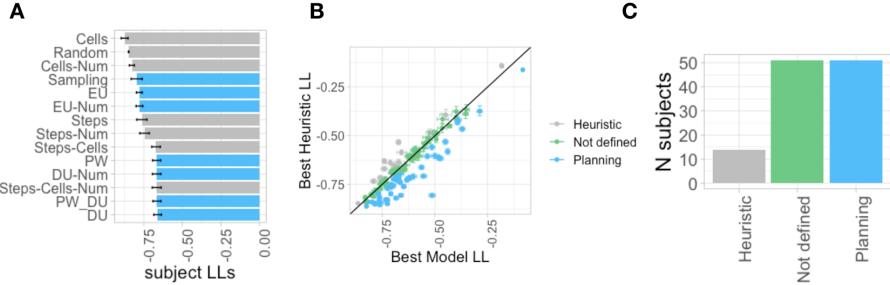


Figure 26: Experiment 1. Models fitted with Monte-Carlo cross-validation.

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

1077 bootstrapped difference between their means of [.07.3], indicating that DU-
 1078 Num predicts the aggregate population behavior better than the optimal EU
 1079 model. The 95 CI of bootstrapped difference between correlations of DU-Num
 1080 (a planner with the highest correlation) and Steps-Cells (a heuristic with the
 1081 highest correlation) [.02.14], suggesting that the DU-Num model predicts the
 1082 aggregate population better than the myopic Steps-Cells heuristic. These
 1083 results remained consistent with conclusions presented in the main text as
 1084 we re-ran this analysis for percentages $\in [10, 50]\%$.

1085 Below, we show additional analysis in which we drop the softmax parameter,
 1086 and fit the remaining parameters of each model as whichever combination
 1087 of parameters agrees with the highest number of agent’s choices. Then, for
 1088 each model, given its best fitting parameters, we count how many times the
 1089 agent’s decision corresponds to the action this model suggests is an optimal
 1090 choice. The results of this analysis agree with the results of our likelihood-

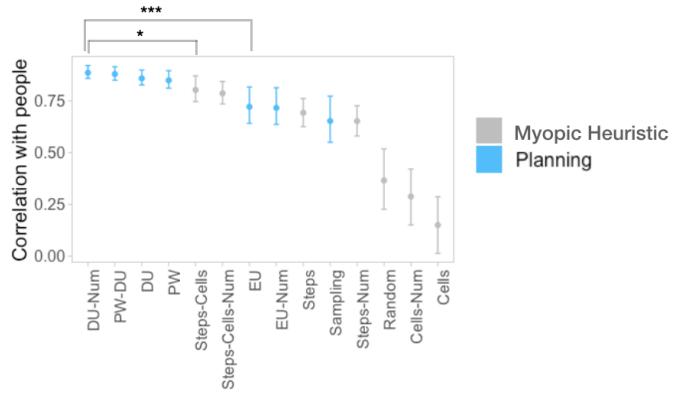


Figure 27: Experiment 1. Bootstrapped correlations of models' predictions with choice probabilities aggregated across the experimental population. The analysis includes all decisions visited by at least 20% of participants. Error bars indicate 95% confidence intervals.

¹⁰⁹¹ based analysis, showing that our cognitively-inspired planners with a limited
¹⁰⁹² horizon explain behavior better than the EU model.

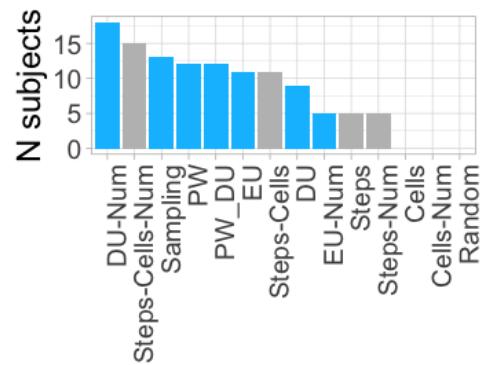


Figure 28: Experiment 1. The distribution of best-fitting models across individuals

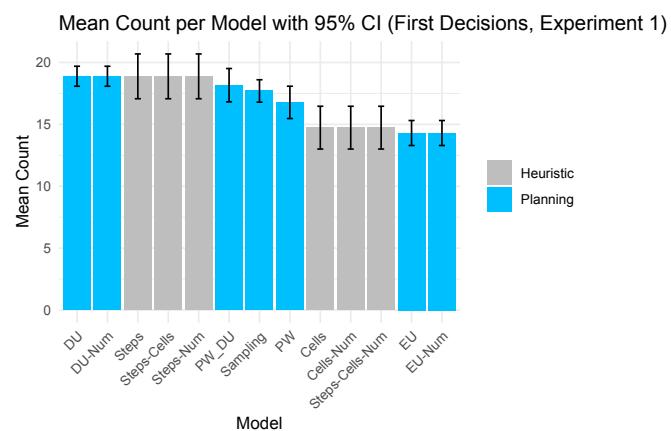


Figure 29: Experiment 1, first decisions. The counts indicate the number of an individual's decisions that were optimal under a model, given that model's best fitting parameters. Error bars indicate 95% confidence intervals across people.

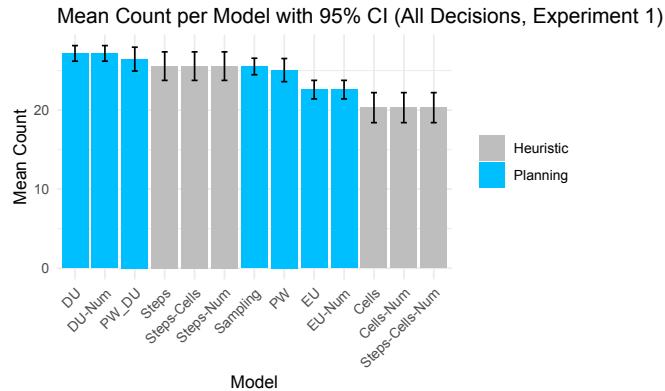


Figure 30: Experiment 1, all decisions. The counts indicate the number of an individual’s decisions that were optimal under a model, given that model’s best fitting parameters. Error bars indicate 95% confidence intervals across people.

Table 5: Parameter Estimates by Model, Experiment 1

model	tau	gamma	beta	k	bits	budget	c
Cells	6.27			0.0			
Cells-Num	3.64			0.0	0.18		
DU	1.2	0.7	1.0				
DU-Num	0.56	0.51	1.0			1.9	
EU	3.0	1.0	1.0				
EU-Num	3.0	1.0	1.0			7.8	
PW	1.24	1.0	0.71				
PW_DU	1.17	0.72	1.33				
Random	1.0						
Sampling						163.58	9.69
Steps	5.02			1.0			
Steps-Cells	1.14			0.59			
Steps-Cells-Num	0.33			0.61	0.74		
Steps-Num	1.91			1.0	0.38		

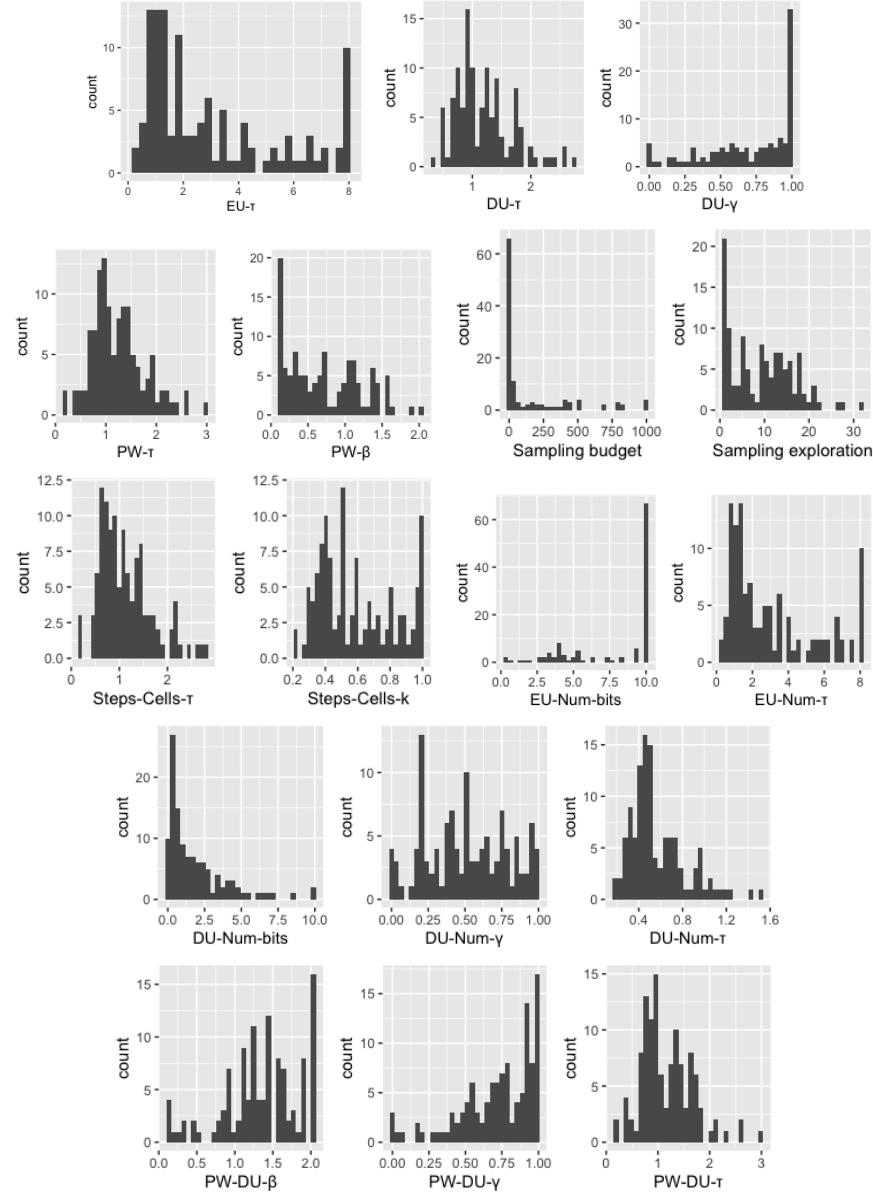


Figure 31: Experiment 1. Distribution of parameters fitted at individual level

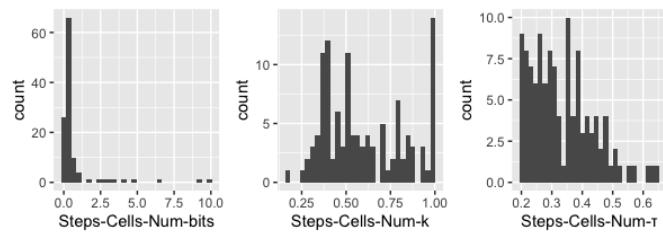


Figure 32: Experiment 1. Distribution of parameters fitted at individual level

1093 **5.4 Experiment 2, Additional results**

1094 **Fitting models with Monte-Carlo Cross validation**

1095 We fitted models to individuals using Monte-Carlo cross-validation to
1096 illustrates variability between individuals. Figure 33 shows mean LLs of each
1097 model per move (panel A) and mean LLs per move of each individual's best
1098 fitting planner and heuristic. The labeling of individuals as "Planning" or
1099 "Not Defined" is based on non-overlapping bootstrapped 95 CI of best-fitted
1100 heuristic and planner (not the bootstrapped 95 CI for difference between
1101 means). The distribution of best fitting models across individuals is shown
1102 in Figure 34.

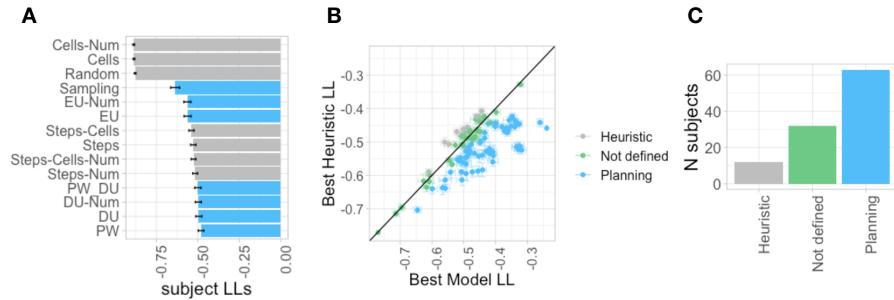


Figure 33: Experiment 2. Models fitted with Monte-Carlo cross-validation. **A.** Mean LL per decision with 95% CI, averaged over participants. **B.** Mean LL per decision for an individual's best-fitting planning model and best-fitting heuristic. Each dot represents an individual, error bars are 95%CI. 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

1103 **Computing correlations**

1104 Correlation analysis presented in the main text focuses on decisions visited
1105 by all participants, so that each data-point used in computing correlation is

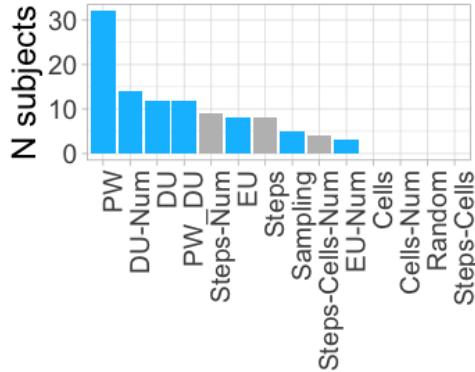


Figure 34: Experiment 2. The distribution of best-fitting planners across individuals

1106 based on the same population of people. In Figure 35 we show an alternative
 1107 analysis, aimed to maximize the number of data-points used to compute
 1108 correlation, by using decisions that were visited by at least 20% of participants.

1109 Here, the correlation of the best-performing PW-DU model with peo-
 1110 ple is $r = .96(95CI[.95, .97])$, and the correlation of the optimal EU is
 1111 $r = .82(95CI[.76, .86])$. These correlations are significantly different, with
 1112 bootstrapped difference between their means of [.1.2], indicating that PW-DU
 1113 predicts the aggregate population behavior better than the optimal EU model.
 1114 The 95% CI of bootstrapped difference between correlations of PW-DU (a
 1115 planner with the highest correlation) and Steps (a heuristic with the highest
 1116 correlation) was [0.020.05], suggesting that the DU-Num model predicts the
 1117 aggregate population better than the myopic Steps heuristic. These results
 1118 remained consistent with conclusions presented in the main text as we re-ran
 1119 this analysis for percentages $\in [10, 50]\%$.

1120 Below, we show additional analysis in which we drop the softmax parame-

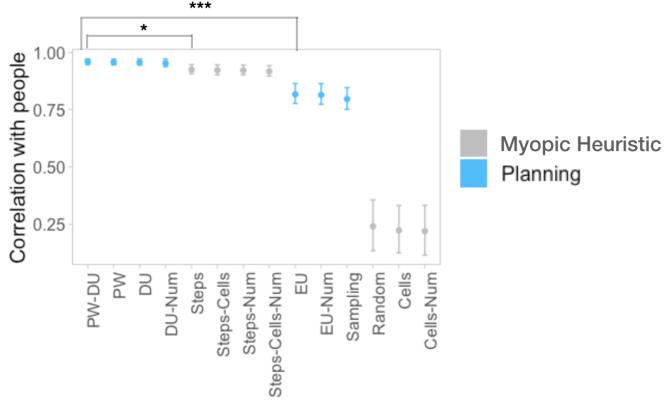


Figure 35: Experiment 2. Bootstrapped correlations of models' predictions with choice probabilities aggregated across the experimental population. The analysis includes all decisions visited by at least 20% of participants. Error bars indicate 95% confidence intervals.

1121 ter, and fit the remaining parameters of each model as whichever combination
 1122 of parameters agrees with the highest number of agent's choices. Then, for
 1123 each model, given its best fitting parameters, we count how many times the
 1124 agent's decision corresponds to the action this model suggests is an optimal
 1125 choice. The results of this analysis agree with the results of our likelihood-
 1126 based analysis, showing that planning models explain a significantly higher
 1127 fraction of people's decisions.

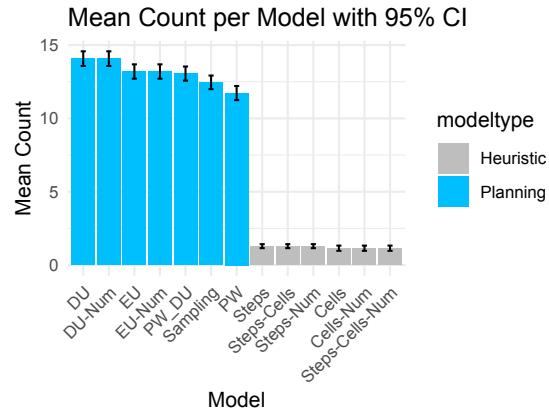


Figure 36: Experiment 2, first decisions. The counts indicate the number of an individual's decisions that were optimal under a model, given that model's best fitting parameters. Error bars indicate 95% confidence intervals across people.

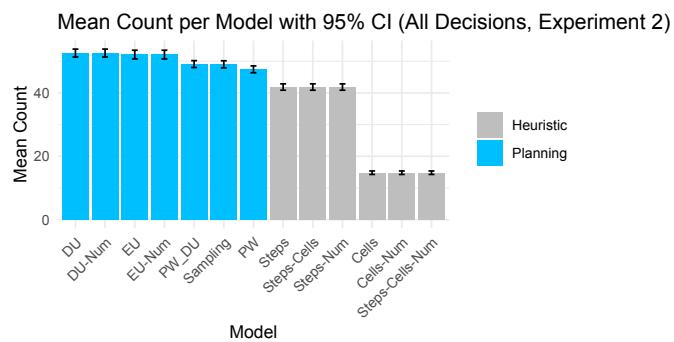
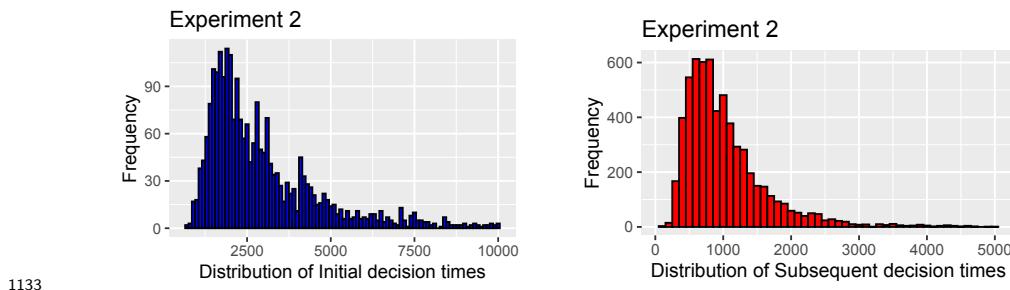


Figure 37: Experiment 2, all decisions. The counts indicate the number of an individual's decisions that were optimal under a model, given that model's best fitting parameters. Error bars indicate 95% confidence intervals across people.

1128 **Distribution of decision times in Experiment 2**

1129 The distributions of decision-times used to compute regression models are
 1130 shown below. We removed Initial decision-times longer than 10s, and Subse-
 1131 quent decision-times longer than 5s for homoscedasticity needed to compute
 1132 regression models.



1133

1134 **Does probability weighting explain decision times?** We compute
 1135 linear regression of the form $time_i = \alpha_0 + \alpha_1 \beta_i + \alpha_2 \tau_i + e$, where β_i and τ_i
 1136 are fitted to people in model PW. In this analysis we removed data from 4
 1137 participants fitted with $\beta > 1$, so that only people fitted with $\beta \in (0, 1]$ are
 1138 used. Here a larger β corresponds to more accurate probability perception.
 1139 In the fitted regression model we find a positive slope at β_i for Initial decision
 1140 times, and a negative slope for Subsequent decision times - indicating that
 1141 β in the PW model can reflect precision of computation. The results of the
 1142 fitted regression model are shown in Table 6.

Times	p	F-statistic	α_0	Slope τ	Slope β
Initial	< .0001	F(2,2240)=11.54	2664	-18 $p = .8$	901 $p < .0001$
Subs.	< .0001	F(2,5839)=23	1080	34 $p = .06$	-283 $p < .0001$

Table 6: Linear regressions predicting decision times by parameters fitted by model PW.

Table 7: Parameter Estimates by Model, Experiment 2

model	tau	gamma	beta	k	bits	budget	c
Cells	5.80			0.00			
Cells-Num	1.26			0.00	0.10		
DU	0.97	0.44	1.00				
DU-Num	0.62	0.37	1.00		3.77		
EU	2.20	1.00	1.00				
EU-Num	2.18	1.00	1.00		8.64		
PW	1.14	1.00	0.35				
PW_DU	0.91	0.53	0.65				
Random	1.00						
Sampling					214.92	11.90	
Steps	1.04			1.00			
Steps-Cells	0.86			0.86			
Steps-Cells-Num	0.24			0.88	0.58		
Steps-Num	0.28			1.00	0.54		

Distribution of fitted parameters, Experiment 2

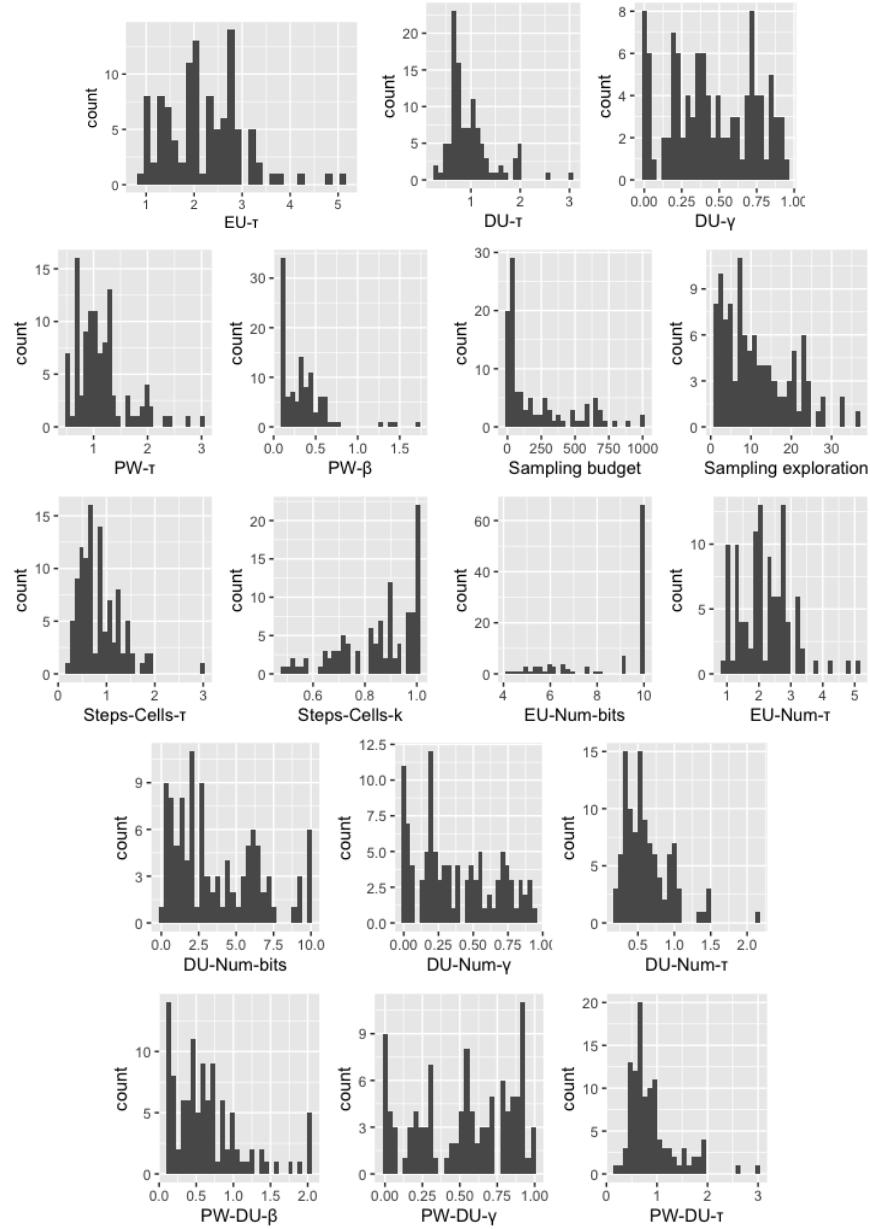


Figure 38: Experiment 2. Distribution of parameters fitted at individual level

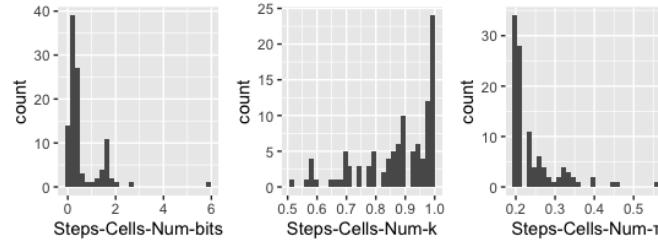


Figure 39: Experiment 2. Distribution of parameters fitted at individual level

1144 5.5 Maze Search Task

1145 The MST task was extensively piloted to ensure the clarity of instructions
1146 and a sufficient amount of practice to make the task intuitive to humans.
1147 A version of MST has been used to study how do people evaluate the
1148 goodness of plans made by others [6], however human planning in MST
1149 has not yet been studied by detailed computational modeling, which is
1150 the goal of current work. An online versoin of MST can be accessed at
1151 <https://marta-kryven.github.io/mst.html>.

1152 **Task Instructions:**

1153 **Screen 1 – Instructions**

1154 Welcome to our study!

1155 IMPORTANT

1156 This study runs best in Firefox, on a desktop/laptop.

1157 The study will NOT run on Safari, or a mobile device.

1158 In this study you will look for an exit in a maze.

1159 After this task, you will be asked to provide demographic information.

1160 The study is expected to take about 20 minutes.

1161 Thanks for participating!

1162 (For brevity, we omit the informed consent statement at the end of this page)

1163 button: [I AGREE]

Screen 2 – Instructions

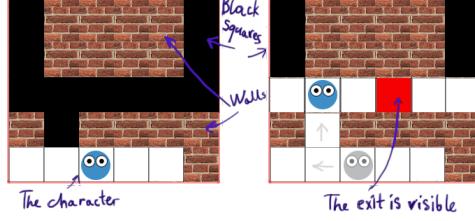
INSTRUCTIONS (PLEASE READ CAREFULLY)

1166 Your task is to exit the maze by reaching the red square in as few steps as
1167 possible.

1168 You can move one square at a time by clicking on the white squares next to
1169 your character.

1170 You cannot see through the walls. The squares you cannot see yet are black.
1171 The exit is equally likely to be behind any of the black squares.

A maze looks like this:



1173 YOU CAN GET A BONUS!

Planning your path wisely will pay off.

1175 The better you plan your path, the fewer steps you'll take and the more
1176 bonus you can earn.

1177 A bonus of \$3 will be awarded if you're in the top 20%, OR

1178 A bonus of \$2 will be awarded if you're better than average, OR

¹¹⁷⁹ A bonus of \$1 will be awarded if you're better than the bottom 30%.

button: [Let's practice!]

1181 Screen 3 – Practice mazes

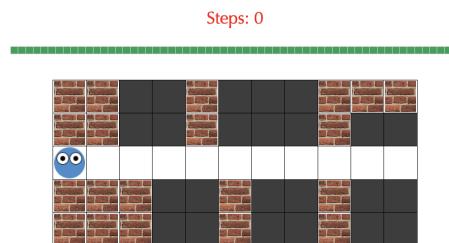
Practice Maze X of Y

1183 Let's look at this map. There are some black squares, a brick wall, and your
1184 character.

1185 There is ONE exit in this maze. This exit could be behind any one of the
1186 black cells.

1187 You can move your blue character by clicking one of adjacent white cells.

1188 Please find the exit in as few steps as possible.



1189 **Screen 4 – Instructions Quiz**

1190 **Great, you have finished Practice!**

1191 Please answer the quiz questions below to move on.

1192 Question 1: My task is to ..

1193 • visit every square in the maze

1194 • see how lucky I am

1195 • solve the mazes in as few steps as possible

1196 • click as fast as possible

1197 Question 2: Exits are always placed ...

1198 • in the bottom left corner

1199 • anywhere in one of the black cells

1200 • in the first place I search

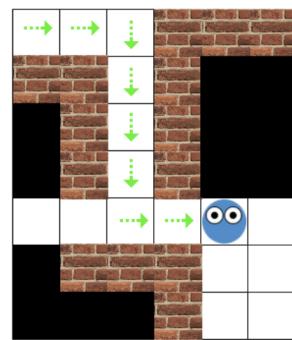
1201 • in the top right corner

1202 Question 3: Which image correctly shows parts of the maze the character

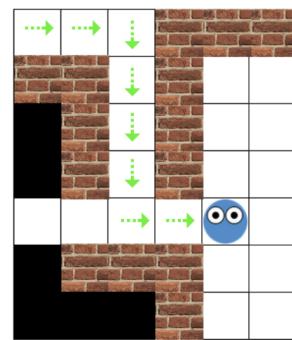
1203 has not seen yet (black squares)?

1204 button: [Submit]

○ Image A ○ Image B



A.



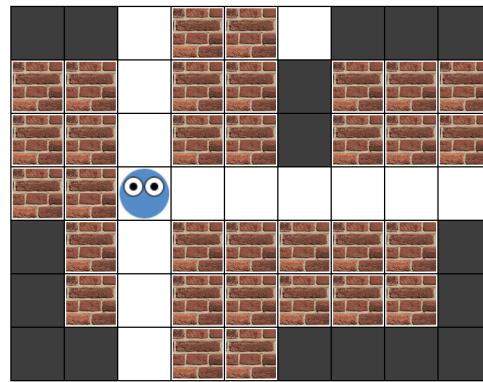
B.

1205 **Screen 5 – Experiment**

1206 Maze X of Y

1207 Please find the exit in as few steps as possible.

Steps: 0



1208 **Screen 6**

1209 Thank you!

1210 How did you make your decisions about which way to go?

1211 Text input: [...]

1212 button: [Submit]

1213 **Screen 7 - Demographics**

1214 Your age: [...]

1215 Your gender: [...]

1216 OPTIONAL: Please leave any comments about the study here, we welcome

1217 any feedback.

1218 [...]

1219 button: [Submit]

1220 5.6 Mazes Used Experiment 1

1221 The Figures below show mazes used in Experiment 1. The mazes were
1222 presented in a randomized order. The exit location was chosen randomly
1223 at the time of experiment design. Here, the eye symbol is used to illustrate
1224 locations of the room-revealing states, this marker was not part of experiment.

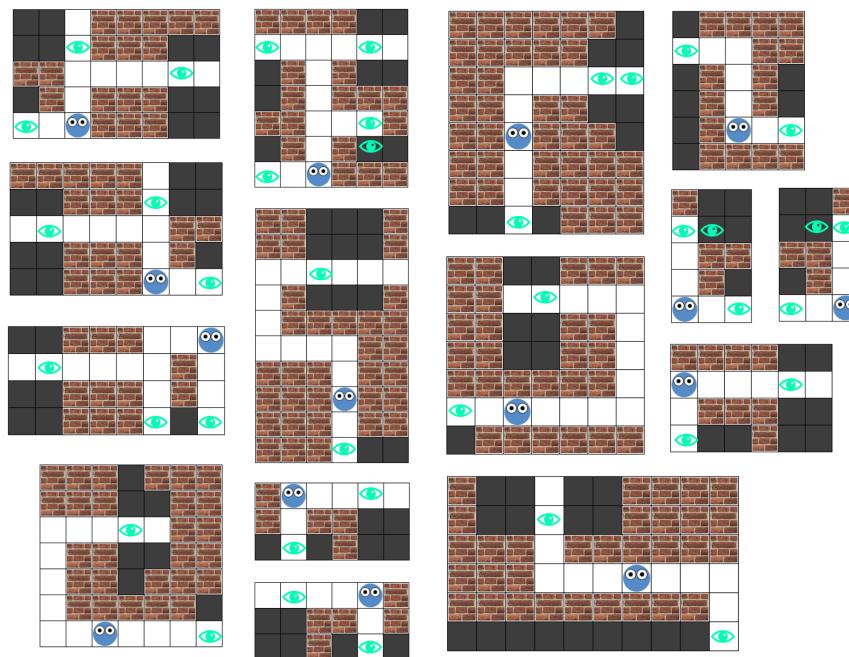


Figure 40: Mazes used in Experiment 1. The eye symbol marks the locations of room-revealing states.

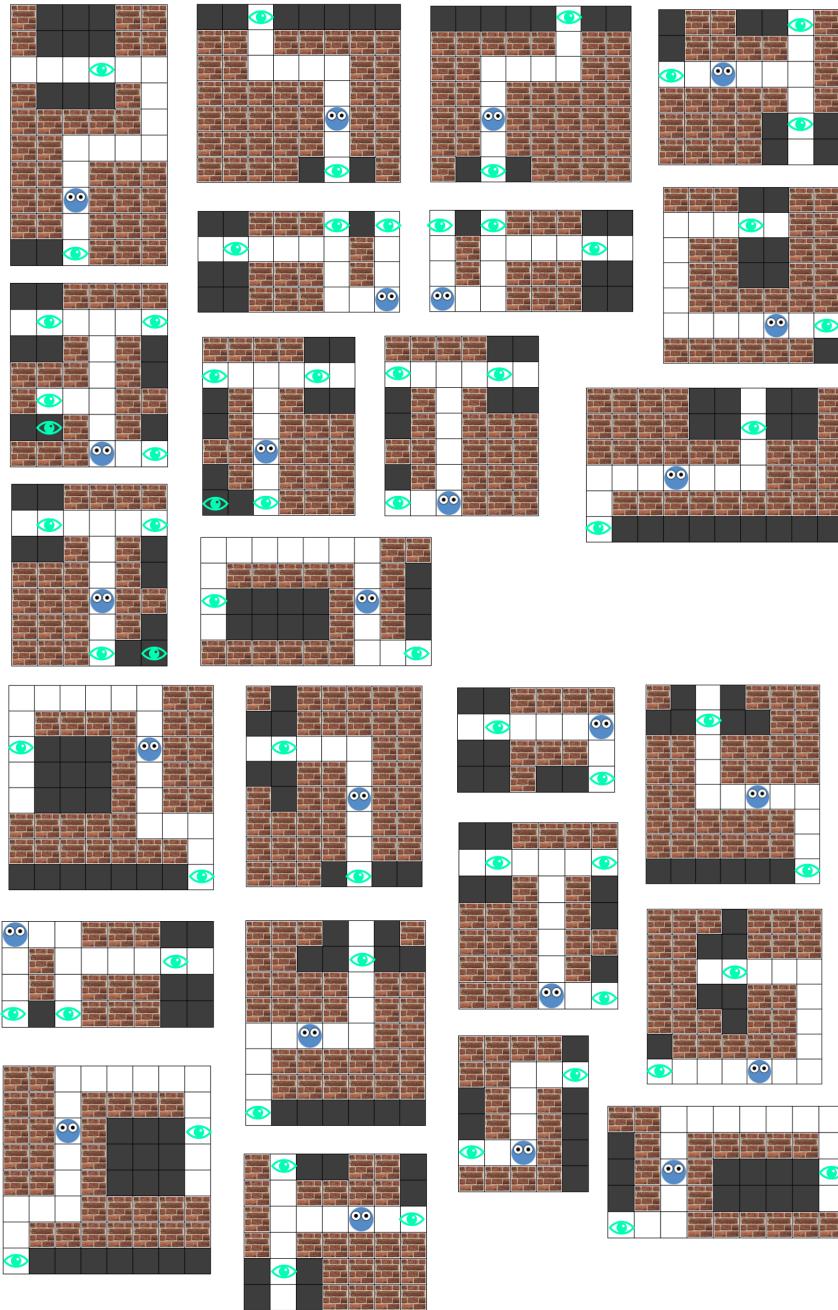


Figure 41: Mazes used in Experiment 1. The eye symbol marks the locations of room-revealing states.

1225 5.7 Mazes Used Experiment 2

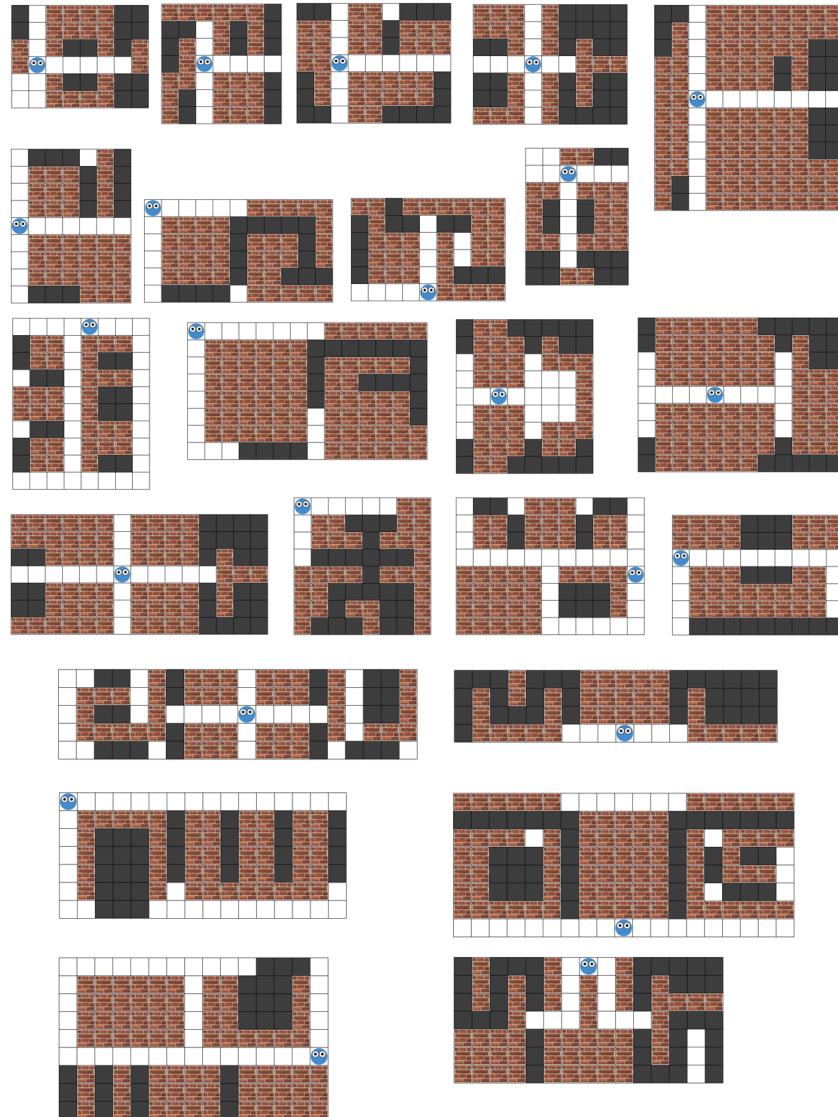


Figure 42: Mazes used in Experiment 2. The mazes were presented in a randomized order. The exit location was chosen randomly at the time of experiment design.