Adventures of human planners in Maze Search Task

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

***Abstract*— Humans make efficient plans during everyday navigation and natural spatial search, while these tasks still remain challenging for algorithms. Which mental computational models do we have that makes this possible? We investigate three computational principles that may be leveraged by people — approximate expected utility maximization, discounted utility, and probability weighed utility—in the context of a novel spatial Maze Search Task. These computational principles are well studied in classic bandit tasks and monetary gambles, but they have not been evaluated on naturalistic spatial tasks that involve sequential decision making. We found that accounting for a combined effect of these three principles explains aggregate human behavior better than models that include just one, or two of these principles, or any of the four behavioral heuristics. We also found substantial individual differences, revealing that humans are best explained by a diversity of planning strategies rather than a single best model. Our results take a step toward uncovering common computational qualities of human spatial planning that may generalize to natural human behaviors in daily life.**

I. INTRODUCTION

To build autonomous agents that seamlessly cooperate with people, we first need to understand how people plan. It is especially important to understand human planning in the context of daily activities—such as during navigation, and spatial search. While humans can plan remarkably well in such contexts [1], these tasks are notoriously hard for algorithms. In particular, uncertainty and partial observability can make spatial planning intractable [2].

Imagine looking for your keys before leaving the house. Certain locations are easily accessible, but unlikely to contain the searched object. Other locations are highly likely but further away. In which order should you search them?

Existing computational models have explored aspects of planning in classic laboratory tasks, such as bandits, gambles, and games [3]–[8]. However, their findings have not been evaluated on sequential spatial behaviors that share greater resemblance to the kinds of situations that people encounter in daily life. Notably, three relevant computational principles have been documented in the cognitive science and economics literature on decision-making tasks related to planning. **First**, people may *approximately maximize expected utility*, with action probability proportional to its utility, to account for occasional sub-optimal choices [9]–[11]. **Second**, humans may limit their planning horizon [3], [5], for example, when evaluating delayed monetary rewards [12], [13], or possible game moves [5]. This can be modeled by *discounting future utilities* [11], [14]. **Third**, humans have been shown to *weight probabilities* in utility computation, in a way that produces less accurate estimates of values at the extremes [12], [13], [15]–[17]. For example, people often over-estimate probabilities of extremely unlikely events [15], and perceive extreme quantities on a logarithmic scale [16]. Furthermore, Prospect Theory explains people's choices in gambles by non-linear weighting of the probabilities of outcomes [12], [13], as has been recently validated in a large multi-national replication [6]. However, these results are limited by the simple gambling tasks studied.

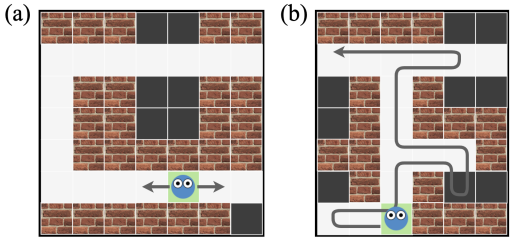

Fig. 1. **(a)** An example of a Maze Search Task with two unobserved rooms. A room is defined as a cluster of dark cells that can be revealed together. The player can move within the white cells which are visible and empty; bricked cells are walls that the player can not see or move through; black cells are unrevealed and may contain the hidden exit. The hidden exit is equally likely to be in any of the black cells, and the player's goal is to reach the exit in the fewest steps possible. **(b)** A hypothetical search trajectory in the Maze Search Task with four unobserved rooms. The player's starting location is indicated by the green square.

We hypothesized that each of these principles, or their combination, may be used during spatial search. At the same time, previous work found substantial individual difference in human problem-solving strategies of tasks that require planning, such as games [5], [7], as well as the use of trivial sub-optimal heuristics that may be employed to save effort and time [5], [8]. Thus, we hypothesized that individuals may vary in their use of planning strategies and heuristics, for example, as a result of individual differences in motivation, or cognitive resources available at the time.

We evaluate the combined use of these three computational principles in a behavioral experiment. To the best of our knowledge, we are the first to apply a combination of these principles to sequential decision-making in a spatial task. We found that the model combining all the three principles was best at predicting human behavior. We also found that individuals exhibited a variety of planning behaviors, consistent with all of the evaluated strategies. Our results suggest that approximate utility maximization, utility discounting and

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

[2]Department of Mechanical Engineering, Massachusetts Institute of Technology

*Equal contribution

{mkryven, syu2, maxkw, jbt}@mit.edu

probability weighted utility may be fundamental to human decision-making and attest to the importance of modeling a variety of strategies to predicting individual-level behaviors in more naturalistic contexts.

II. MAZE SEARCH TASK

We use a Maze Search Task (MST) to study human decision-making. In this task, subjects navigate through a series of partially observable, two-dimensional grid-world mazes [18]. Each maze is made up of walls, corridors, rooms, and a single hidden exit. Here, a room is defined as a cluster of dark cells that are revealed together from a common location. The subjects are instructed to reach this hidden exit in as few steps as possible. The exit becomes visible as a red tile when the subject reveals the hidden location.

An example maze is shown in Fig. 1 (A). The player's starting position is indicated by the smiley face avatar on a green cell. The player can move to any adjacent tile that is not a wall and illuminate the unobserved tiles (colored in black) by bringing them into the avatar's line of sight. The player moves through a maze until the exit is reached as shown in Fig. 1(B). The player is told that each of the black tiles is *equally likely* to hide the exit, and to reach it in the *fewest* steps possible. The player's objective is to find a route that minimizes the expected number of steps to the exit, given its possible locations. Prior to launching the experiment, this MST procedure was extensively piloted to ensure that it is intuitive to humans while presenting non-trivial challenges, such as trading off costs against rewards [18]. The current work formalizes human decision-making in MST with a family of quantitative models formulated as computational hypotheses.

III. COMPUTATIONAL MODELS

To plan a search trajectory in a MST, we represent observations as the states in the planning process and compute a policy for navigating between the states. Making an observation in a MST refers to revealing an entire room, or clusters of hidden cells that can be revealed together from a common position. The state-transition model for a given maze is structured as a tree, where each state node N_i is defined by the location of the agent making an observation, and the maze area observed so far. For example, in the state space representation shown in Fig. 2(A), the root node indicates the starting location, and the adjacent nodes indicate the subsequent possible states.

Based on this state space, we define four models that evaluate possible paths by estimating the expected number of steps to the exit. The models differ in how they assign value to a state while they share a common mapping of a state value to a probability of choosing it by the noisy maximization function, or commonly known as the softmax function, defined below.

$$\sigma(\mathbf{Q})_k = \frac{\exp(-Q_k/\tau)}{\sum_j \exp(-Q_j/\tau)}. \quad (1)$$

Here σ denotes the noisy maximization function that converts a list of real-numbered values into a probability distribution, Q_k is the value of the considered state, and Q_j are the alternatives to choosing Q_k. Parameter τ controls the noise of this mapping such that smaller values of τ lead to a stronger preference for higher values, and $\tau \to \infty$ yields a uniform probability distribution. Since the state value function Q_j estimates the expected number of steps to reach the exit, we use the negative sign to ensure that paths with smaller number of steps are more likely.

Expected Utility (EU) model measures the expected number of steps to the exit as follows:

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

Here, $C(N_i)$ is the set of all future states following N_i; p_i is the probability that the exit is found at N_i; s_i is the number of steps to reach N_i from the root node; and e_i is the expected number of steps to the exit from N_i if the exit is found at N_i. Thus, the *expected utility* value $Q_{eu}(N_i)$ of visiting a N_i is given by the sum of expected number of steps to the exit, and the expected steps to the exit from the remainder of the maze, assuming that subsequent choices are optimal.

Discounted Expected Utility (DU) model is almost identical to the EU model, except it discounts future values by a factor of $\gamma \in [0, 1]$ as shown in Eq (3). A small γ implies a more myopic agent, and setting $\gamma = 1$ results in a policy equivalent to the policy of the EU model.

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

Probability Weighed Utility (PWU) model is based on one of the key assumptions of Prospect Theory, which states that humans overestimate small probabilities and underestimate large probabilities. We express this assumption by the probability weighting function $\pi : [0, 1] \to [0, 1]$ as $\pi(p) = exp(-|\ln(p)|^\beta)$. The value function for this model then becomes:

$$Q_{PWU}(N_i) = \pi(p_i)(s_i+e_i)+\pi(1-p_i) \min_{C_j \in C(N_i)} Q_{PWU}(C_j) \quad (4)$$

Here β controls probability weighting. Larger β values yield more overestimation on lower probabilities and underestimation on larger probabilities. Smaller β has the opposite effect. Fig. 2 illustrates the predictions of the EU, DU, and PWU models for a simple two-room maze, as a function of the models' parameters.

Lastly, we define the **Combined Model (Comb)** that uses a combination of probability weighting and discounting, resulting in three free parameters: τ, β, and γ. We also define four myopic heuristics, which evaluate state values based on the current state alone without considering future states.

- **Steps Heuristic (SH)** only considers the number of steps to a node from its parent: $Q_{SH}(N_i) = s_i$.
- **Cells Heuristic (CH)** only considers the number of cells that are newly observed at a node: $Q_{CH}(N_i) = -c_i$.
- **Steps-Cells Heuristic (SCH)** combines steps and revealed cells: $Q_{SCH}(N_i) = k \cdot s_i - c_i$, where k is a

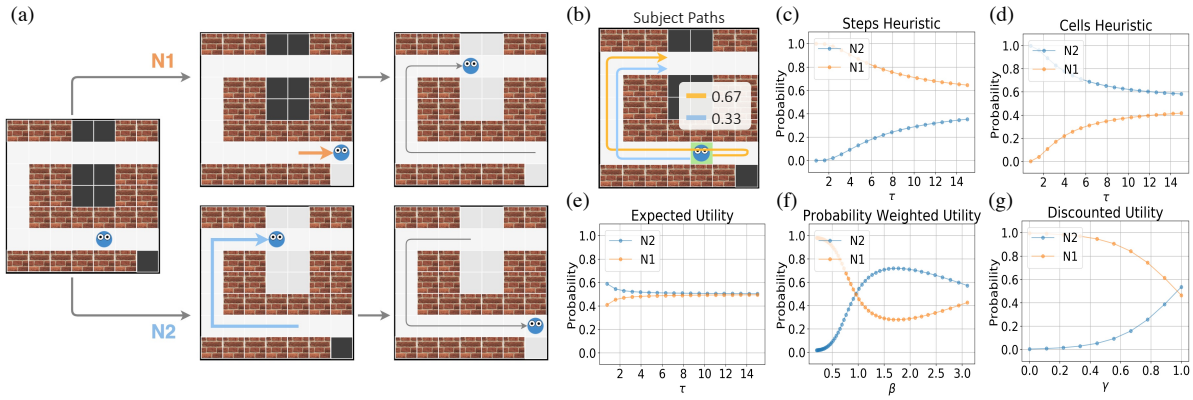

Fig. 2. **(a)** An example of the state space representation for a maze with two hidden rooms. The tree root is the starting location. The adjacent nodes indicate the subsequent possible states. **(b)** The empirical probabilities of the two possible search trajectories in human data – 67% and 33% of the subjects chose node N1 and N2, respectively. **(c)** The Steps Heuristic prefers visiting the closer room (N1) first as $\tau \to 0$ and becomes indifferent between the two nodes as $\tau \to \infty$. **(d)** The Cells Heuristic prefers visiting the bigger room (N2) first and becomes indifferent between N1 and N2 as $\tau \to \infty$. **(e)** The Expected Utility model prefers visiting the bigger room first (N2) as $\tau \to 0$, but becomes indifferent between N1 and N2 as $\tau \to \infty$. **(f)** To illustrate the predictions of the DU model and the PWU models, we fixed $\tau = 1$. The PWU model prefers N1 when $\beta < 1$, and N2 when $\beta \geq 1$. **(g)** The DU model prefers N1 when γ is closer to 0, and N2 when γ is closer to 1.

free parameter. This definition combines step and cell information about the immediate successor states.

- **Random Model (Rand)** assigns the same value to all nodes. Without loss of generality, we choose to set all node values to 1: $Q_{Rand}(N_i) = 1$.

IV. Behavioral Experiment

The eight alternative strategies described above make different predictions about how one should trade-off the likelihood of finding the exit with the costs of movement. For example, the DU model can explain a preference for smaller rooms nearby over larger rooms further away. The PWU model can predict an underestimation of highly likely outcomes, such as the exit being in a large room, and overestimation of unlikely outcomes, such as the exit being in a tiny room. The Steps heuristic can explain preferring the closer rooms regardless of their shape and size. Thus, in designing the experiment we chose a large diversity of mazes that elicit different predictions of behavior from our models. At the same time, our models differ in the amount of computation they require, and in the precision of their estimates. Thus, it is also possible that individuals may use all of these models, possibly due to individual differences in motivation, attention, and cognitive resources.

Method The experiment included 40 mazes, where each maze consists of at least two and at most five rooms. The starting locations were pre-determined, and the exit locations were randomly chosen at the time of design. All subjects saw the same set of mazes in a randomized order. The experiment was conducted in a web browser using a JavaScript interface. Subjects first read a consent page and a short description of the experiment. Following consent, subjects read a detailed description of the task, followed by 3 practice trials and a short quiz about the objective of the task. Subjects could not proceed to the experiment until they submitted the correct answer to the quiz. On each trial a subject was placed at the starting position, and navigated a maze by moving in the four cardinal directions using the mouse until the exit was reached. After completing experiment, subjects answered a demographic questionnaire, and provided a free-form description of any strategies they used in the experiment.

Subjects We recruited 120 subjects via Amazon Mechanical Turk (63 male, 56 female, average age of 39 with standard deviation of 12). One subject was excluded for incorrectly answering the instruction quiz more than twice, so the final analysis was done on 119 subjects.

Analysis of behavior We first analyze the subject's behavior in a model-free way to validate our state space representation. All subjects took direct paths between observation locations, which we represent as states in the decision process. Occasional deviations from the direct paths comprised fewer than 1% of the total moves. Thus, we segmented subjects' trajectories into decision states, aligned to maze tree representations. On average, subjects made 51 decisions (SD = 3.3) during the experiment. The exact number of decisions differed between subjects, due to individuals taking different search trajectories.

Fig.3A shows the expected performance of various models measured in total steps taken during the experiment, if following a greedy strategy under the given model. The variability in the model performance comes from tie-breaking in mazes were several paths are valued as equivalent. The optimal performance is achieved the EU model with an average total steps of $M_{EU} = 413.6$, $SD_{EU} = 14.3$. The best-performing heuristic model is the Cells Heuristic with $M_{CH} = 436.6$ and $SD_{CH} = 9.9$. Most humans took longer paths than the EU or Cells Heuristic model. Fig.3B shows human performance compared to the models' simulated performance, parametrized by the human parameter

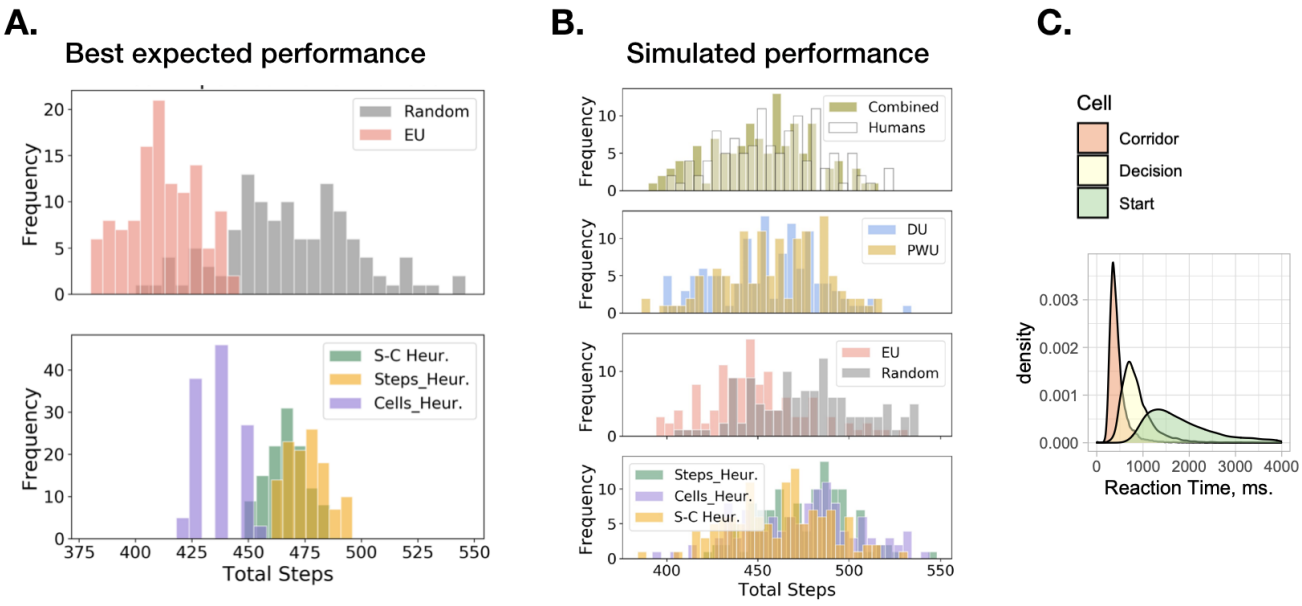

Fig. 3. Model-free results, Experiment 1. **A.** The expected performance of models measured as the total number of steps aggregated over all mazes, if following a greedy strategy under the given model. The variability in the model performance comes from tie-breaking in mazes were several paths are valued as equivalent. The EU model achieves the optimal overall performance. **B.** Comparing human performance to the expected performance of models, if the model parameters are sampled from the parameters fitted to the human population. **C.** Reaction times in milliseconds in different maze locations. People take a few seconds to study the map before moving, move quickly through corridors, and pause at decision locations – whenever new information is observed.

distribution. The average number of steps taken by humans was $M_h = 458.8,\ SD_h = 27.8$.

We coded each step in the human paths as *Start*, if the cell was the starting location, *Decision* if the cell corresponded to an observation location, and *Corridor* otherwise. Fig.3 shows the distribution densities of reaction times (RT) of subjects in milliseconds as they navigated through each type of cell. The Corridor cells were the fastest ($M = 457.5,\ SD = 229.2$)ms., Decision cells took ($M = 932,\ SD = 460$)ms., and the starting cells took ($M = 1804,\ SD = 710.5$) ms. Thus, subjects generally took a few seconds to process each maze, suggesting that people pre-plan their paths, and spent longer in Decisions compared to Corridors ($t(3725) = 60.9,\ p < .0001$, 95% confidence interval of the difference $(462, 483)ms.$), indicating different cognitive processing in the two types of locations.

The RT also decreased with subsequent decisions – for example, in mazes with four rooms the second decision generally took longer than the third decision. The linear regression model of RT as a dependent variable against where the $depth$ of the decision tree at the current decision as an independent variable, was significant ($p = .03$, $F(1, 1601) = 4.84$, coefficient significance $p = .03$) with the regression equation $RT = 841ms + 42msdepth$, suggesting that people may be re-evaluating their plans at each decision.

Model-based analysis We fitted the parameters of each model to each individual's decisions by Maximum Likelihood Estimation, and by 4-fold cross-validation. Both methods achieved consistent and very similar parameter estimates. The median parameters fitted to the subject population were as follows: EU($\tau = 2.01$), DU ($\tau = 0.95$, $\gamma = 0.88$), PWU ($\tau = 0.95$, $\beta = 0.7$), Combined ($\tau = 1.1$, $\beta = 1.14$, $\gamma = 0.86$) Steps-Cells Heuristic($k = 1.1,\ \tau = 2.97$), Steps Heuristic($\tau = 7.8$), Cells Heuristic($\tau = 10$). To analyze the fit of each model to the aggregate statistics of the subject population we measured bootstrapped correlations between the predicted choice probabilities, given median parameters fitted to the subject population, and empirical choice frequencies aggregated over subjects. The model fits are summarized in Fig. 4. The correlations of the Combined, DU and PWU, models were the highest. The Steps-Cells Heuristic was also highly correlated with the aggregate human behavior, as shown by the partially overlapping 95% confidence intervals of Steps-Cells Heuristic and the best fitting models.

Model Selection via k-fold Cross Validation To examine individuals' use of different decision-making strategies, and assess the fit of each model to individuals, we split each subject's decisions into an 80-20% train and test split, fitted the models to the training set using four cross-validation folds, and validated the fit on held-out data. Cross validation controls for over-fitting and different numbers of parameters in the models. The fits of each model to individuals shown in Figure 4A, suggest that people used a variety of strategies, including all of the evaluated models and heuristics, with the Combined model fitted as best explaining the highest fraction (25%) of individuals. The Steps-Cell Heuristic was the next most popular strategy, fitted as best explaining about a fifth of the subjects. Figure 4D shows pairwise comparison of cross-validation fits between the Combined model, and the 6 other best-fitting strategies in the order of fraction of subjects better

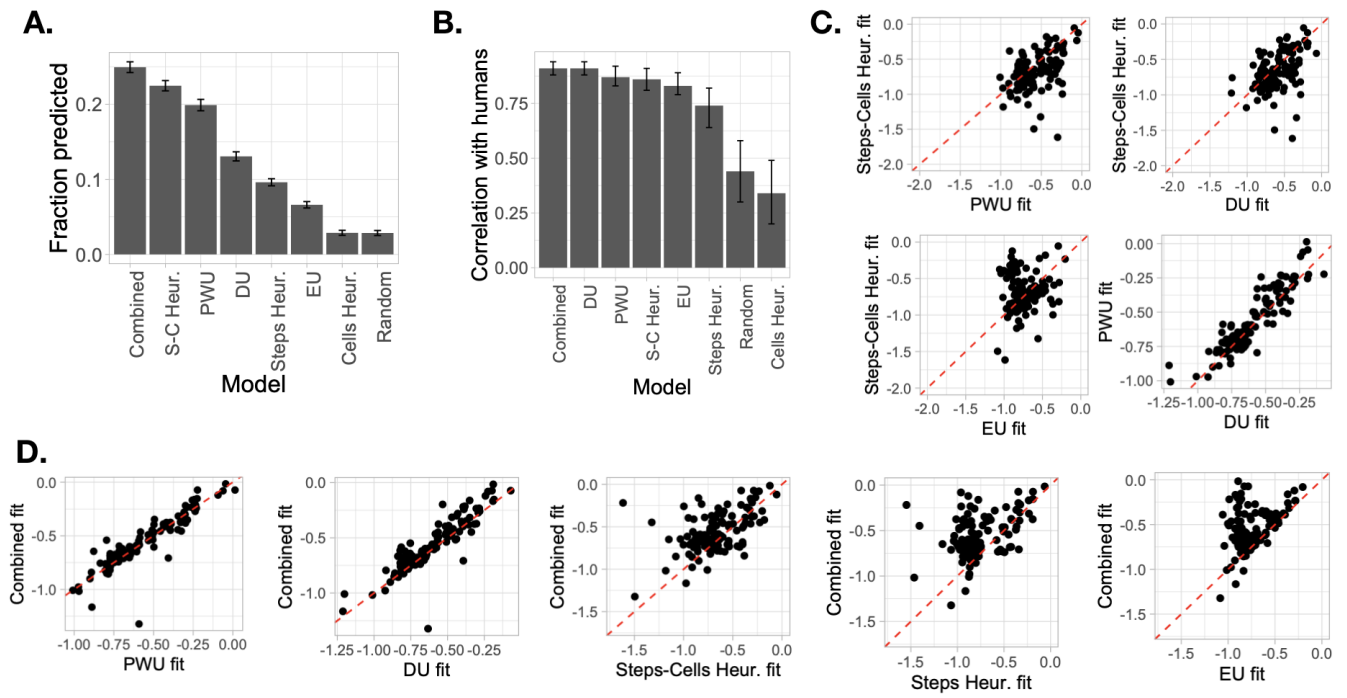

Fig. 4. Model-based results, Experiment 1. **A.** Fractions of individual subjects best predicted by each model, and by each heuristic, according to 4-fold cross-validation. **B.** Bootstrapped correlations of the predictions of all models and heuristics with the aggregate data. Error bars indicate 95% confidence intervals. **C.** Pair-wise comparison between model fits to individuals according to 4-fold cross-validation, comparing PWU, DU, Combined, and the Steps-Cells heuristic. Each dot represents a person. The red diagonal line indicates equally good fits - any dots that fall right on the middle of this line would be equally well explained by either of the compared models. **D.** Pair-wise comparison between the fits of the Combined model, and the fits of six other strategies, shown in the order of fraction of subjects better predicted by the Combined model.

predicted by the Combined model, which ranged between 0.59 (Combined vs. PWU) to 0.83 (Combined vs. the EU model).

V. Discussion

In current work we have generalized the examination of decision-making principles from simple gambling tasks to sequential spatial behaviors in a Maze Search Task (MST). We used behavioral experiment and modeling to investigate the use of three cognitive computational principles known to influence decision-making processes: approximate expected utility maximization (EU), discounted utility (DU), and probability weighed utility (PWU), applying them to naturalistic spatial behaviors.

We found that the Combined model, which accounts for approximate utility maximization, utility discounting and probability weighting, was consistently best at explaining human behavior. The Combined model outperformed other models that formalize only one or two of the three principles, as well as any of the four alternative behavioral heuristics. While we found that individuals used a variety of planning strategies, the Combined model consistently outperforms other models and heuristics in pair-wise comparison, predicting the highest fraction of individuals. Our results suggest that these three principles may be fundamental to human decision-making, and generalize to natural domains.

Qualitative inspection of subjects' responses to the free-form decision-making questions revealed three popular answers. First, about a half of the subjects reported prioritizing the closest room, consistent with the Steps heuristic. A smaller group reported preferring larger rooms, but made no reference to distance, as described by the Cells heuristic. The third most popular answer referred to minimizing steps while maximizing the sizes of visited rooms, suggesting a more sophisticated planning. Lastly, a few individuals reported guessing, as may have been formalized by the Random model. While our computational models formalize these strategies, it is unclear how to relate the model fits to the human self-reported strategies. For example, few of the evaluated individuals were consistently random, and a striking few were best explained by the Cells Heuristic. This could mean that human planning computations are not cognitively penetrable, although people may observe their actions and describe them in a simplified way. This could also mean that humans give inherently ambiguous descriptions of algorithmic and strategies to communicate them with less language. For example, depending on the configurations of the environment, reporting prioritizing of closer rooms could describe behavior typical of DU, and reporting prioritizing bigger rooms could reference PWU-like computations.

The MST takes a step toward measuring planning in natural behaviors, however it is limited by the grid-world topology, in which the layout of the entire environment can be seen at all times. Future work should extend the evaluation

of the three principles to street networks, environments with multiple goals, and multi-dimensional topologies. Our analysis also leaves open the possibility that the planning strategies evolve over time, for example, due to people optimizing over the set of alternative heuristics [7], fatigue, or learning. Subjects could also apply different strategies flexibly to different spatial scenarios. Subjects could be influenced by the observed exit locations as they proceed through the experiment, for example, if subjects observe that larger rooms have higher likelihood of hiding the exit, they could become biased toward favoring larger rooms towards the end of the experiment. Future studies will also investigate how human planning computations may emerge on the algorithmic level from using sampling-based computations with sparse samples, such as those based on MCTS [19]. This research fills the gap of modeling sequential spatial behaviors that share greater resemblance to real-life scenarios, and takes a step toward designing more human-like planning algorithms that can be leveraged in robotics to support better human-robot collaboration, and inverse planning inference of human actions.

REFERENCES

[1] C. Bongiorno, Y. Zhou, M. Kryven, D. Theurel, A. Rizzo, P. Santi, J. Tenenbaum, and C. Ratti, "Vector-based pedestrian navigation in cities," *arXiv preprint arXiv:2103.07104*, 2021.

[2] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.

[3] Q. J. Huys, N. Lally, P. Faulkner, N. Eshel, E. Seifritz, S. J. Gershman, P. Dayan, and J. P. Roiser, "Interplay of approximate planning strategies," *Proceedings of the National Academy of Sciences*, vol. 112, no. 10, pp. 3098–3103, 2015.

[4] C. M. Wu, E. Schulz, M. Speekenbrink, J. D. Nelson, and B. Meder, "Generalization guides human exploration in vast decision spaces," *Nature human behaviour*, vol. 2, no. 12, pp. 915–924, 2018.

[5] B. Van Opheusden, G. Galbiati, Z. Bnaya, Y. Li, and W. J. Ma, "A computational model for decision tree search.," in *CogSci*, 2017.

[6] K. Ruggeri, S. Alí, M. L. Berge, G. Bertoldo, L. D. Bjørndal, A. Cortijos-Bernabeu, C. Davison, E. Demić, C. Esteban-Serna, M. Friedemann, *et al.*, "Replicating patterns of prospect theory for decision under risk," *Nature human behaviour*, vol. 4, no. 6, pp. 622–633, 2020.

[7] Y. R. Jain, F. Callaway, and F. Lieder, "Measuring how people learn how to plan.," in *CogSci*, pp. 1956–1962, 2019.

[8] B. Meder, J. D. Nelson, M. Jones, and A. Ruggeri, "Stepwise versus globally optimal search in children and adults," *Cognition*, vol. 191, p. 103965, 2019.

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

[10] K. Louie, L. E. Grattan, and P. W. Glimcher, "Reward value-based gain control: divisive normalization in parietal cortex," *Journal of Neuroscience*, vol. 31, no. 29, pp. 10627–10639, 2011.

[11] R. S. Sutton, A. G. Barto, *et al.*, "Introduction to reinforcement learning. vol. 135," *MIT press Cambridge*, vol. 5, pp. 21–22, 1998.

[12] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[13] D. Kahneman and A. Tversky, "On the interpretation of intuitive probability: A reply to jonathan cohen.," 1979.

[14] J. R. Doyle, "Survey of time preference, delay discounting models," *Delay Discounting Models (April 20, 2012)*, 2012.

[15] F. Lieder, T. L. Griffiths, and M. Hsu, "Overrepresentation of extreme events in decision making reflects rational use of cognitive resources.," *Psychological review*, vol. 125, no. 1, p. 1, 2018.

[16] G. Fechner, "Elements of psychophysics, volume 1. howes, dh (ed.) and boring, eg," 1860.

[17] D. Prelec, "The probability weighting function," *Econometrica*, pp. 497–527, 1998.

[18] M. Kryven, T. Ullman, W. Cowan, and J. Tenenbaum, "Thinking and guessing: Bayesian and empirical models of how humans search.," in *CogSci*, 2017.

[19] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton, "A survey of monte carlo tree search methods," *IEEE Transactions on Computational Intelligence and AI in games*, vol. 4, no. 1, pp. 1–43, 2012.