Study of Human Behavioral Models for Engineering Applications

by

Suhyoun Yu

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

© Massachusetts Institute of Technology 2022. All rights reserved.

Author .....
Department of Mechanical Engineering
July 10, 2022

Certified by .....
Joshua B. Tenenbaum
Professor, Department of Brain and Cognitive Sciences
Thesis Supervisor

Accepted by .....
Nicolas Hadjiconstantinou
Chairman, Department Committee on Graduate Students
Department of Mechanical Engineering

Study of Human Behavioral Models for Engineering Applications

by

Suhyoun Yu

Submitted to the Department of Mechanical Engineering
on July 10, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Input from and interaction with humans is quickly becoming an inseparable part of designing critical systems: electrical power grids are evolving into smart grids that allow for two-way communication between utilities and their users, and the demand for seamless human-robot collaboration in warehouses and urban rescue missions is ever-increasing. As such, proper understanding and modeling of human decision making behavior should be an integral part of designing critical systems. Unfortunately, in the traditional engineering context, humans have been assumed to be rational agents that behave in the way they ought to, rather than the way that they actually do.

This thesis lays the groundwork for incorporating a more general model of human decision making in engineered systems to improve the quality of interaction between a system and its users, and as a result, its overall performance and reliability. We investigate three computational principles known to influence human behavior: noisy utility maximization, discounting, and the probability weighting principle from Prospect Theory. We evaluate their individual and combined effects in the context of a naturalistic spatial planning task that required sequential decision-making. This task presents challenges that will occur across general contexts. All model selection analyses used show that a significant majority of human subjects' trajectories are best explained by models with discounting and, in particular, probability weighting.

We conclude with a simulation of a simplified 2D urban rescue mission, which demonstrates that these more realistic assumptions about human behavior reduce the cost to complete the collaborative mission by ~ 13%. The results reinforce the hypothesis that a more nuanced model of human behavior is critical for systems that heavily interact with humans. Better models better complement real human behavior and therefore enhance the overall efficiency and performance of a given system.

Thesis Supervisor: Joshua B. Tenenbaum
Title: Professor, Department of Brain and Cognitive Sciences

Acknowledgments

First and foremost, I thank my advisor, Josh B. Tennebaum, for his support and constructive feedback during lab meetings. He generously accepted me into his lab in my second year into the program, and provided me the opportunity to connect with amazing labmates in a completely new field. I am also utterly grateful to my two committee members, Alberto Rodriguez and Themistoklis Sapsis, who have closely guided me through the program over the past several years. Their attentiveness was great emotional support that gave me the courage to complete the work. I cannot thank enough my two collegues, Marta Kryven and Max Kleimen-Weiner, whom I worked most closely with for this thesis. I learned so much from their detailed technical feedback during our frequent meetings, and without their help and guidance, I would not have been able come this far.

I will dearly miss working with Adam and the fellow TAs for 6.009 Fundamentals of Programming. I learned so much about Python, code design, and staff management while working as a TA for this course over the past several semesters. Though funding was the initial motivation for signing up to work for this course, it was my respect for Adam that made me continue to work for 6.009. I could not have asked for a better instructor to work for.

Surviving through the pandemic and the rough moments at MIT would have been impossible without my boyfriend and former lab mate, Dongchan Lee, and my four dearest Dynamically Qualified friends, Boyu, Cameron, Rohit and Saviz—we studied for the dynamics qualifying exam, hence the name. Their presence constantly reminded me that I was not alone in this process, and that there are close people who could genuinely empathize with me whenever I felt unsure of myself. I feel special to have friends whom we would celebrate birthdays together, go on annual trips, and congratulate on each other's success. I genuinely hope that we keep this friendship for the decades to come.

I would also like to take a moment to thank my former advisor, Kostya Turitsyn and undergraduate adviser and, I dare say, a dear friend, Oscar Mur-Miranda. I miss

them both very much, and I am excited to let them know that I was able to complete my graduate work. They were the very people who first introduced me to the world of Mechanical Engineering, and I can never forget their warm encouragement when I was still new to everything.

Last but not least, I present my utmost gratitude and respect to my parents and my sister. Without my parents, I would not be, and without my sister, my life wouldn't be as interesting. Their unwavering support and trust with my studies have helped me to achieve this far. Though they live far away, their daily encouragements and messages of love have emotionally sustained me to get through stressful times at MIT.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Related Work

Critical, mechanical, and electrical systems have traditionally been designed maximize to reliability and minimize cost with respect to the normative trend or average demand without taking in much feedback from its individual users. In recent decades, however, there has been a surge in automated systems wherein human input and interaction are an integral part [16, 30]. Large-scale electrical grids are now evolving into smart grids that allow for two-way communication between utilities and their users; demand for seamless human-robot collaboration is rapidly diversifying, from warehouse management to urban rescue missions; private ride-sharing companies are faced with challenging planning problems while accounting for both the supply of drivers and demand of passengers. As such, establishing quantitative models to explain and predict choices made by humans has increasingly become an integral part in systems that heavily interact with people.

Unfortunately, human tendencies and decision-making behavior have usually been overlooked in the context of engineering; human agents were assumed to behave rationally, making decisions that they *ought to* rather than what they *actually do*. Even within the field of psychology or behavioral economics, expected utility theory reigned as the normative descriptive model for decision-making under uncertainty [13, 20]. In the recent decades, however, researchers observed from empirical studies

that the assumption that people make decisions by maximizing expected utility has its limitations [6, 7, 33, 18]. This has led to the development of new models such as prospect theory [18, 38].

Still, it is not clear if it is appropriate to blindly apply those principles to any context or application. As Kahneman and Tversky pointed out, traditional experiments conducted by behavioral economics are typically "contrived gambles for small stakes", and that there is usually a "large number of repetitions of very similar problems". The gambles they refer to are usually of the form $(p, x; q, y)$, where outcomes x and y can be earned with probability p and q, respectively. See Tables B.1 ~ B.3 for examples of such gambles used by Kahneman and Tversky in their original paper on prospect theory. They acknowledged that this paradigm could raise obvious questions regarding the "validity of the method and the generalizability of the results."

Prospect theory is widely accepted to this day, and there has been much effort to replicate the results on larger scale data and modern standards in the recent decade. In 2020, a large group of researchers conducted a multinational study—involving 4,098 participants, 19 countries, 13 languages—to validate that this theory is still relevant across different cultures [32]. In 2021, another group substantially increased the size of the experiment and designed a family of machine learning models based on this theory [28]. While the successful replication of the original work in these recent studies is reassuring, they still use the same experimental paradigm. Therefore, it is still unclear if the principles of prospect theory, or any descriptive behavioral models developed by psychologists and behavioral economists, is applicable in other fields.

1.2 Contribution

While theorists agree that human decision-making is usually an approximation of what we think ought to be done [12, 17], studies of its algorithmic principles in the laboratory have been confined to simple games or gambles, unlike the dilemmas, humans encounter in daily life [14, 40]. Here we investigate human planning in a new experiment paradigm named the Maze Search Task (MST) that involves spatial task

resembling natural search and navigation, and ask which computational principles underlie human decision-making in this task.

Computational Principles

We hypothesized that human decision-making may exhibit three computational cognitive principles, previously described in the literature, although not yet evaluated together, nor tested in a naturalistic context. First, even with known utilities, people do not always make the best choice, especially when the differences between utilities are small. Instead, the choice may be approximately rational, with action probability proportional to its utility. This principle of *noisy utility maximization* is widely used, for example, as Luce's choice in economics [25], as Boltzmann response mapping in neural networks, and as divisive normalization in neurosceince [24].

Second, humans may limit the planning horizon of multi-step choices [14], usually modeled by *a discount factor* [4], so that the value of future utilities is inversely proportional to time. Utility discounting is often used in Reinforcement Learning [35], as well as in economics when modeling the values of delayed monetary rewards [17, 38].

Third, humans could perceive probabilities in a non-uniform way, by over- or under-weighting them [23]. This principle has been introduced as *probability weighing* in prospect theory [17, 38]. A brief proof in support of this principle presented by Kahneman and Tversky is in Appendix B. On the sensory level, probability weighting is related to the Weber-Fechner law of just noticeable difference [8], which states that humans tend to perceive measurable quantities on a logarithmic scale, distorting the perception of extremes. The principles of prospect theory have been recently applied to scale reward in Reinforcement Learning [31, 29, 11] for system controls, and validated in a large-scale modern replication of classic gamble experiments [32], although to our knowledge has not been formally evaluated in humans on non-gamble tasks.

In this work, we tested the combined effect of these three principles using behavioural experiments and computational modelling. To our best knowledge, this work is the first to test the cognitive effects of probability warping in the context of

a spatial task. We found that the aggregate human behaviour was predicted reasonably well by the optimal planner with noisy utility maximization. However, analysis of individual differences revealed that individual humans seamlessly combined noisy utility maximization, discounting, and probability warping, with a significant increase in the prediction accuracy of individual behaviours when the three principles were combined. Finally, probability warping explained more of the individual differences in human behavior than discounting.

Maze Search Task (MST)

To study human decision-making in the context of intuitive and abstract environments, we developed a series of Maze Search Tasks (MST) [22]. In an MST, a human subject is presented with a two-dimensional, partially observable maze with the task to locate the unique exit. This exit is initially hidden, so the subject has to explore and uncover the hidden spaces. The goal for the subject is to find the exit in the maze in as few steps as possible. The subject is informed that the exit may be hidden under any hidden tile with uniform probability. Effectively, this means that the player's objective in MST is to find a route that minimizes the expected steps to the exit, given its possible locations.

Each maze in the MST is comprised of four types of cells; wall, open, hidden, and exit cells. An open cell (illustrated as a white square) is a cell that the avatar has full view of and may walk on. A collection or a string of such open cells creates a room or a corridor. Each room is sectioned off by a wall cell (illustrated with brick patterns) that the avatar cannot see through or walk on. A hidden cell (illustrated as a black square) is a cell that the avatar has not explored yet and has no information on whether it could be an open or an exit cell. An exit cell (illustrated as a red square) is initially hidden under one of the hidden cells, and it is to be revealed when it enters the avatar's vision. In each maze, there is one unique exit cell, and it may be placed anywhere in one of the hidden cells with uniform probability. An example maze can be found in Figure (1-1).

While 3D virtual structures are the more popular choice for environments in stud-

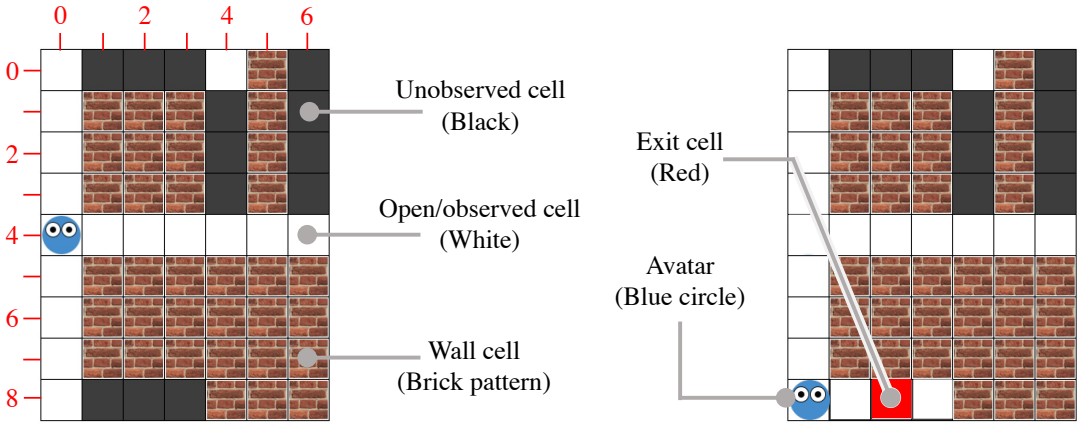


Figure 1-1: An example of a maze. (Left) Initial configuration of the maze that a subject sees at the start of the trial. The avatar to move is placed at position (row, column) = (4,0). This avatar could be moved to any open (white) adjacent cell. (Right) When the avatar moves all the way down to position (8, 0), it can open up unobserved (black) cells and convert them into open cells. The avatar must explore the maze by making new observations to locate the exit (red) cell that could be hidden under any black cell with uniform probability.

ies that involve human spatial navigation and orienteering [41], [2], [27], there are clear benefits to using simple 2D mazes. First of all, previous work has shown that this task is intuitive to human decision makers [22]. It also poses non-trivial challenges typical of sequential decision-making, such as trading off costs against rewards, and considering the opportunity costs that result from choosing one direction over the others.

Moreover, we can remove any factor that may inadvertently take up subjects' cognitive load during the experiment and affect their decision-making behavior by using 2D mazes. One such factor is visual attention, a cognitive process responsible for selecting partial information from the environment [1]. In other words, humans selectively take in information on the screen and may overlook details, so the more complicated the visual queue, the harder it is to analyze their behavior. Another cognitive load may come from the fact that humans need to build cognitive maps when navigating [36], [34], and constantly need to orient themselves when the structure of the environment is not fully visible at all times. The proposed 2D partially observable environment eliminates the concern of visual attention by simplifying the components

of the maze and unburdens subjects with the task of building and memorizing the structure by allowing for a full bird's-eye view of the structure at all times.

We designed MST stimuli likely to bring out their behavioural correlates to test the three computational principles. For example, we expected that discounting may cause a preference for nearby smaller rooms rather than larger rooms that are further away, even when the larger rooms should be preferred according to expected utility theory. Noisy utility maximization could lead to choosing longer expected paths, especially when the differences between the optimal values of alternatives are small. Nonlinear approximation of probability is expected to cause a combination of these effects and should be especially sensitive to the risk of incurring an unlikely but highly negative outcome, such as backtracking to a single remote unobserved cell, if it turns out to contain the exit despite the initial odds. In summary, we designed stimuli rich in trade-off decisions that often combined many of these features at once.

Structure of the Thesis

The remainder of the thesis is structured as follows. We present the formal definitions of the models in Chapter 2. Recursive models that take into account the aforementioned three computational principles are referred to as the *planning models*, as opposed to *heuristic models* that only consider one step ahead. Along with the definitions, we also provide examples to illustrate how each model behaves.

In Chapter 3, we describe the procedure and interface of the two experiments we conducted in this work. The first experiment investigates if humans tend to favor moving in a particular direction, namely left or right. This step was crucial for designing the mazes for the second experiment. The second experiment studies the individual and combined effects of the three computational principles. With the result, we assess how well a model fits aggregate subject behavior by plotting the distribution of bootstrapped correlation coefficients, and how well a model fits individual subjects via k-fold and Monte Carlo cross-validation.

Finally, we conclude the thesis by demonstrating the importance of correct modelling of human decision-making in engineering with a simulation of a human-robot

collaboration task. In Chapter 4, we identify the increasing demand for robot assistance in urban rescue missions and describe a simplified environment wherein a human and artificial agent cooperate to save a victim. We model the human agent's decision-making behavior based on the conclusion drawn from chapter 3 and measure the improvement in the collaborative performance when the robot correctly understands the human partner's behavioral model.

Chapter 2

Models

2.1 Decision Tree

Before we introduce the models, we need to first define a *decision tree*. We formally represent a maze as a decision tree to support mapping of a subject's trajectory to a sequence of decisions. A node in this decision tree is defined by the sequences of observations made by the avatar and the position at which it made the last observation. The rationale for using this decision tree is three folds. One, a human doesn't make a decision at each cell in the maze, but rather after gaining new information. Second, since the decision tree only considers the sequence of observations made rather than the actual path, we can reduce potential noise created as a result of a subject changing its mind during a path or accidentally moving back and forth. Third, we can reduce computational cost by not having to consider each cell in the maze. Formally, we can solve a MST problem by representing observations as the nodes in the planning process, and computing a policy for how to navigate between them.

We also define a *decision node* as a node in the decision tree with two or more children nodes. Consider the example illustrated in Figure 2-1. Node (A), or the root node, is the initial configuration of the maze where the avatar hasn't moved nor made any observations yet. There are three different observations that the avatar can make from the initial position; hence the root node has three children nodes, namely (B), (C) and (D). By definition, the root node in this example is a decision node, and any

leaf node is not a decision node. When the avatar takes the path [(4,0), (3,0), (2,0), (1,0), (0,0)] and observes the room at the top, then it reaches node (B). Note that even if the avatar takes the path [(4,0), (3,0), (2,0), (3,0), (2,0), (1,0), (0,0)], the only observation it makes is the room at the top, so we still consider that the avatar has reached node (B). The definition of our node also implies that the first child node of (B) and the first child node of (C) are also different nodes because the sequence of observations made by the two nodes are reversed.

2.2 Model Definitions

We will present two types of models: heuristic and planning. A heuristic value function is only capable of looking one step ahead, i.e. it only takes into account the information of immediate children nodes. A planning value function, on the other hand, is capable of *planning ahead* and can look ahead into all future steps. Each planning value function is formulated to reflect one of the four aforementioned principles in human planning: utility maximization; discounting of future cost and reward; and re-weighing of probabilities.

A value function can be interpreted as a model to estimate the expected number of steps to reach the exit from a given node, so the smaller the value the more preferable the node. All value functions depend on the same state space representation described in the previous section. To convert a node value into some *probability* of selecting it, we pass it through the softmax function defined below:

$$\sigma(\boldsymbol{v})_k = \frac{\exp(-v_k/\tau)}{\sum_j \exp(-v_j/\tau)} \tag{2.1}$$

Here σ denotes the softmax function that converts a list of real-numbered node values to a probability distribution, such that the probabilities are proportional to the values. Value v_k denotes the value of the considered node, and v_j denotes the alternatives to choosing v_k. Parameter τ controls the noise of this mapping, so that smaller values of τ lead to a less noisy preference for higher values, and large τ yield a more uniform probability distribution over which node to visit next. The negative

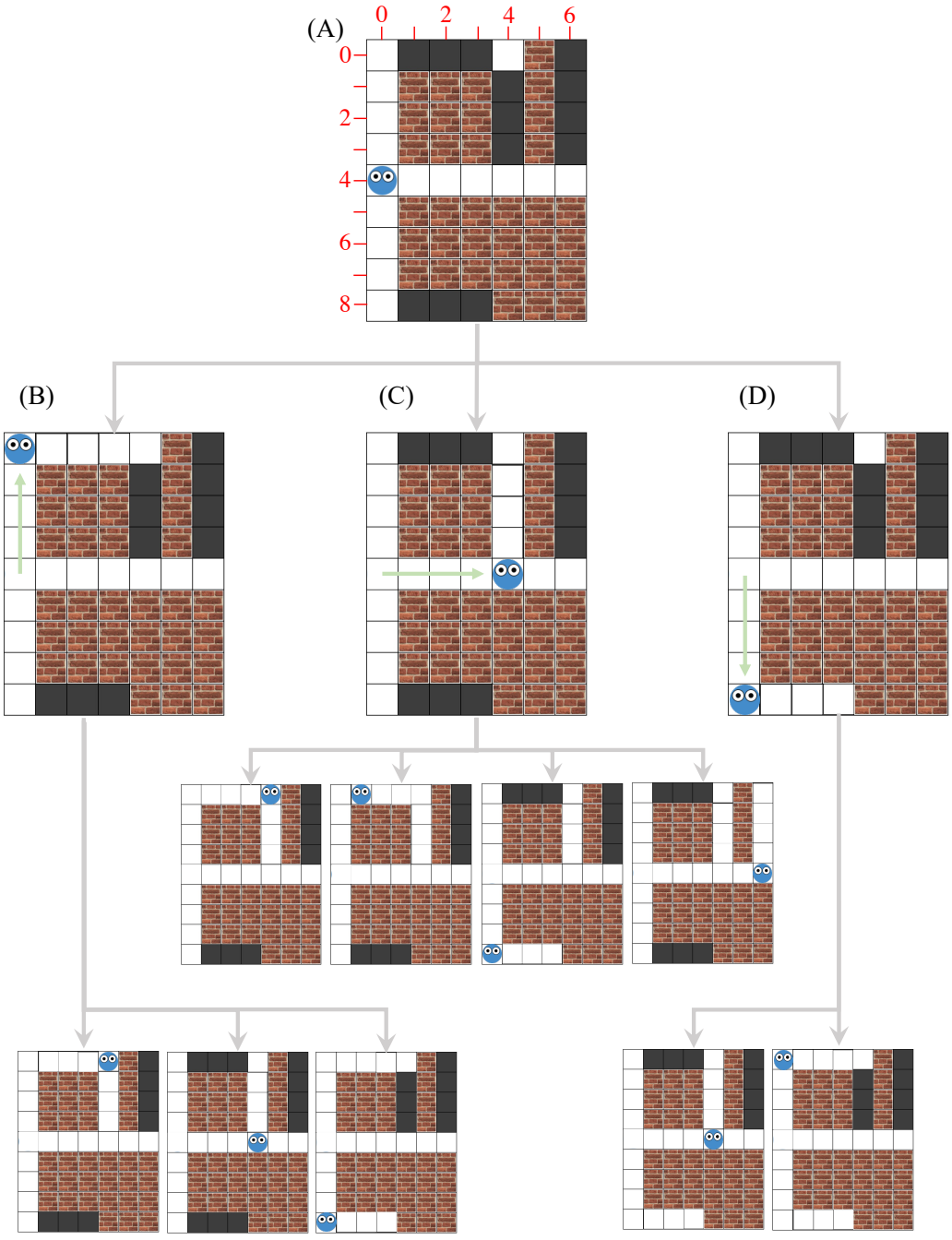


Figure 2-1: Visual example of a decision tree. This example only illustrates up to the second layer of the tree. **(A)** The subject sees this configuration at the start of the trial in this maze. Because the avatar hasn't made any observations yet, this configuration is referred to as the *root node*. Since this root node has three children nodes, namely (B), (C), and (D), it is also a decision node. Node (B) is reached by observing the top room; node (C) is reached by observing the middle room; and node (D) is reached by observing the bottom room.

sign in front of node values ensures that shorter paths result in higher probabilities.

For the remainder of this section, we define the following notations. $V(N_i)$ denotes the value of node N_i; $C(N_i)$ is the set of all children nodes of N_i; p_i is the probability of finding an exit at node N_i; s_i is the total steps taken to reach node N_i from its parent node; and e_i is the expected number of steps to the exit from node N_i in the event that the exit becomes visible in node N_i.

2.2.1 Heuristic Models

We present four types of heuristic models; Random, Heuristic Steps (HS), Heuristic Cells (HC), and Heuristic Steps Cells (HSC). As the name suggests, the Random model chooses a child node with uniform probability at any given node. HS only aims to minimize the number of steps to get to a node from its parent node. HC always goes to the node with the highest probability of finding the exit, i.e. the node with the most black cells. HSC takes the linear combinations of steps from parent node, and chance of finding an exit in a node. Formally, we can write HSC as follows:

$$V(N_i) = \lambda \cdot s_i + (1 - \lambda) \cdot c_i. \tag{2.2}$$

Here, $\lambda \in [0, 1]$ is some scalar that determine the model's dependence on number of steps from parent node versus immediate probability of finding an exit. Note that we get the HS model when $\lambda = 1$, and we get the HC model when $\lambda = 0$.

2.2.2 Planning Models

Expected Utility (EU) Model

The EU model defines the value function as the sum of utilities weighted by its probability of occurring. In theory, this value function should outperform any model in any environment.

$$V(N_i) = s_i + p_i \cdot e_i + (1 - p_i) \min_{c_j \in C(N_i)} V(c_j) \tag{2.3}$$

This definition of the value function is analogous to recursively computing *expected utility* of a path; the value of the next state is equal to the sum of immediate expected path to the exit from a state (if found) and the expected path to the exit from the remainder of the maze, assuming that subsequent states are chosen optimally.

Consider the value comparison of two paths in Figure (2-3). The corresponding paths are illustrated in Figure (2-2a). From the starting position at (0,0), going to the right (blue path) will allow an agent to illuminate black cells as it moves down the maze. Indeed, the EU model plans ahead and evaluates that the value of taking the blue path is better if one plans ahead. While going down (orange path) can reveal more squares with the same number of steps, the agent must go out of its way to observe the black cells it left out at the top if the exit was not found at the first guess. But since the HSC model is only capable of looking one step ahead, it can only prefer the orange path.

Discounted Utility (DU) Model

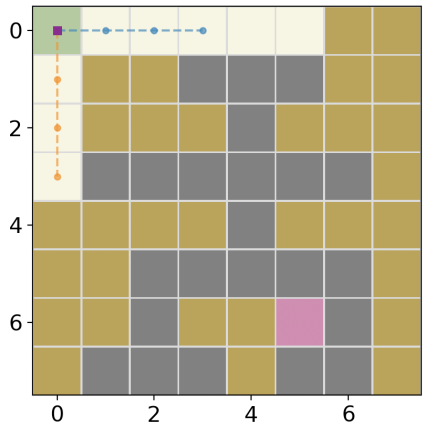
The DU model is constructed by adding a discount factor $\gamma \in [0, 1]$ to the recursive term in the EU model. Note that the DU model converges to the EU model when $\gamma = 1$, and converges to the HS model when $\gamma = 0$.

$$V(N_i) = s_i + p_i \cdot e_i + \gamma(1 - p_i) \min_{c_j \in C(N_i)} V(c_j) \tag{2.4}$$

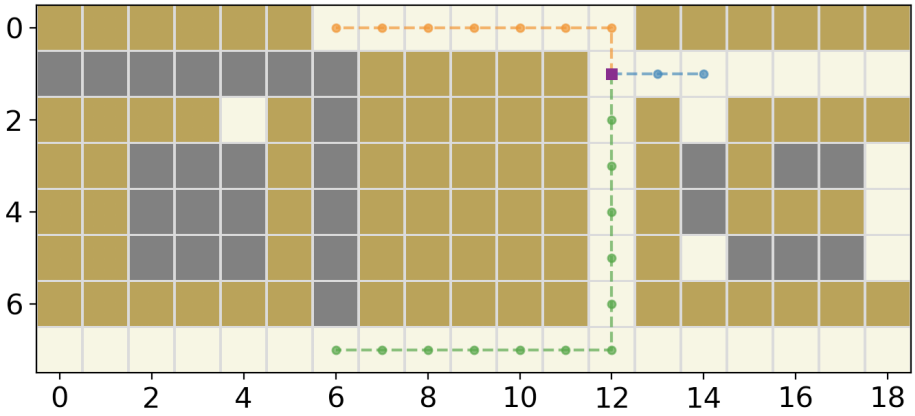
A small γ implies steeper discounting of future values, representing a more myopic agent. Figure 2-1 (d) shows an example of how the probabilities of choosing the next state under this model may change with γ, after applying Equation 2.1 to obtain probabilities. In this example smaller γ increase the probability that the closer, but less likely exit location will be prioritized over the further, but more likely location.

Probability Weighted Utility (PWU) Model

The PWU model is constructed by replacing each probability p_i term with *weighted* probability $\omega(p_i)$. According to prospect theory, humans tend to overestimate small



(a) Illustration of path for node values in Figure 2-3



(b) Illustration of path for node values in Figure 2-5.

Figure 2-2: Illustration of paths for node values in Figures 2-3 and 2-5. These examples are used to demonstrate how different models behave.

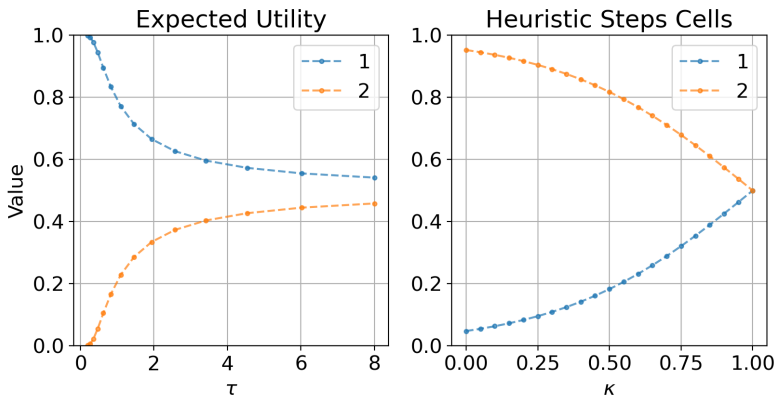


Figure 2-3: Comparison of the behavior of EU and HSC models. From the starting position at (0,0), going to the right (blue path) will allow an agent to illuminate black cells as it moves down the maze. Indeed, the EU model plans ahead and evaluates that the value of taking the blue path is better if one plans ahead. While going down (orange path) can reveal more squares with the same number of steps, the agent must go out of its way to observe the black cells it left out at the top if the exit was not found at the first guess. But since the HSC model is only capable of looking one step ahead, it can only prefer the orange path.

probabilities and underestimate large probabilities. To represent this re-weighting of probabilities, we define the following weighting function $\omega : [0,1] \to [0,1]$:

$$\omega(p) = exp(-\alpha|\ln(p)|^{1/\beta}) \tag{2.5}$$

$$V(N_i) = s_i + \omega(p_i)(e_i) + (1-\omega(p_i)) \min_{C_j \epsilon C(N_i)} V(C_j) \tag{2.6}$$

This definition of the weighting function satisfies the boundary conditions $\pi(0) = 0$ and $\pi(1) = 1$. The parameter α determines where the weighting function intercepts with the $\pi(p) = p$ line; larger α lower the intercept position. The parameter β controls probability warping, so larger β values yield sharper overestimation on lower probabilities and underestimation on larger probabilities, and smaller β results in closer approximations of the optimal choice. For simplicity, we fix $\alpha = 1$ and only consider the effect of β. See Figure (2-4) for how the behavior of this weighing function changes for different values of β.

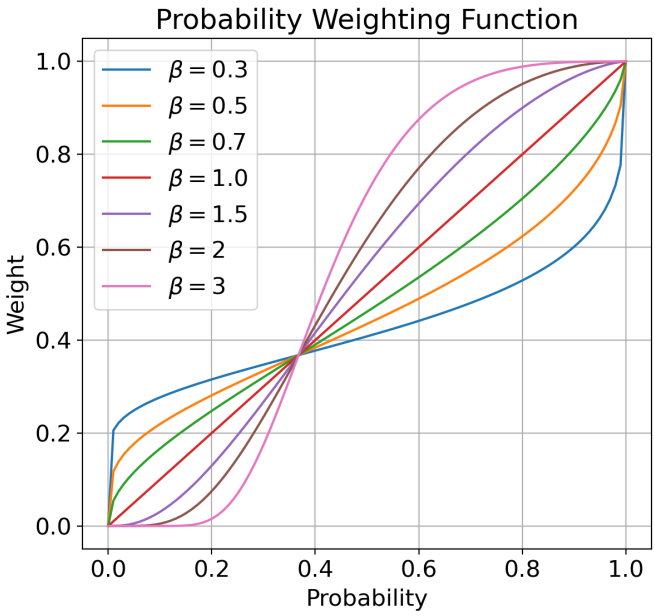


Figure 2-4: Probability weighting function for various β values. The behavior of underestimating large probabilities and overestimating small probabilities can be observed when $\beta < 1$. The behavior is reversed for $\beta > 1$.

Comparison of behavior of the EU model against the other two computation models on paths illustrated in Figure (2-2b). Note that the softmax function (Eq. 2.1) is incapable of reversing the order of values, so the EU model will always yield the same ranking of the children nodes given a node. But the DU and PWU models have the ability to reverse the order of preference by reweighing the value of future decisions. In this case, going left to take the orange path may be the optimal decision under expected utility theory, but it is likely that a human subject would want to check the small room that is much closer since the cost is not too high.

The Combined Model

The Combined model considers both discounting and probability warping in its node value definition, and it is defined by all three parameters, τ, γ and β.

$$V(N_i) = s_i + \pi(p_i)(e_i) + \gamma(1 - \pi(p_i)) \min_{C_j \in C(N_i)} V(C_j) \tag{2.7}$$

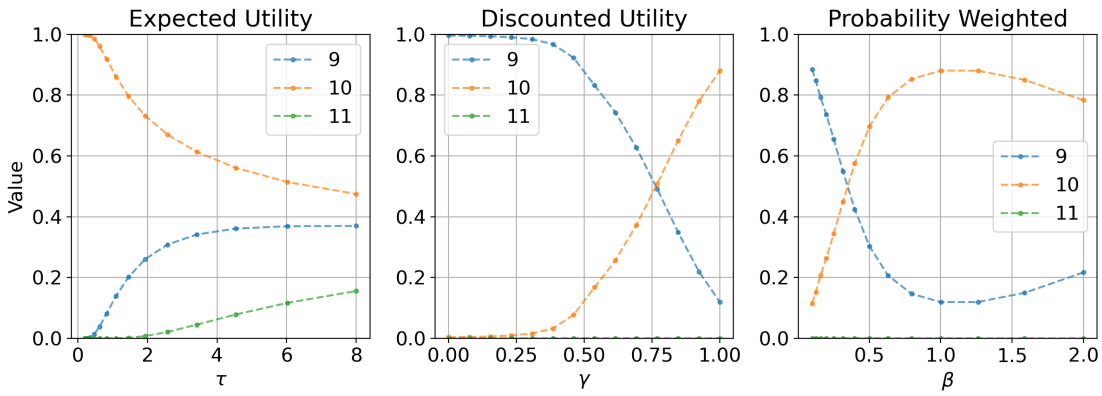


Figure 2-5: Node values for paths illustrated in Figure (2-2b) under three different planning models, namely EU, DU and PWU. Note that the softmax function (Eq. 2.1) is incapable of reversing the order of values, so the EU model will always yield the same ranking of the children nodes given a node. But the DU and PWU models have the ability to reverse the order of preference by reweighing the value of future decisions. In this case, going left to take the orange path may be the optimal decision under the expected utility theory, but it is likely that a human subject would want to check the small room that is much closer since the cost is not too high.

Chapter 3

Experiment with Maze Search Task

3.1 Experiment Procedure

The experiment was conducted online via a web interface hosted on MIT XVM. To enter the experiment, a subject first sees the survey page on Amazon Mechanical Turk with quick instructions and a link to the actual experiment. A screenshot of this survey page is in Figure 3-1. Upon completing the experiment, a subject is provided with a unique ID that starts with the letter "S", followed by six to nine digits. They are asked to go back to the survey page and enter this ID in the box provided.

The first page in the link in the consent page with a short introduction to MSTs (Figure 3-2). Following the consent, subjects were provided with a specific description of their task, which is to locate the exit in the maze in as few steps as possible (Figure 3-3). To check their understanding of the task and familiarize subjects with the MST environment, we provided seven practice mazes. Note that these practice trials were not used in the analysis. Following the practice trials, subjects were presented with three questions related to their tasks(Figure 3-4). They were required to answer correctly to all three questions in order to move on to the actual experiment. In case they answered incorrectly to at least one question, a popup notification appears on the screen instructing them to go through the practice trails again. This process repeats until they correctly answered all three questions.

At each experiment trial, the subjects see four items on the display screen: counter for the number of mazes they have seen so far, a one line reiteration of the task, a colored steps bar to visually indicate the number of steps taken so far in a given maze, and finally the maze itself. The steps bar is initially fully colored in dark green, but its length shrinks and the color slowly transitions from green to yellow, to orange, and finally to red as the subject takes more steps. An example of the experiment screen is illustrated in Figure 3-8. The bar is to visually remind the number of steps that the subject has taken so far, and unburden the subjects from having to mentally keep track of their steps. This way, we allow them to focus on deciding their next best moves.

At the beginning of each trial, a subject is placed at the starting position marked by the blue circular avatar. At any given position, the subject can move the avatar by clicking any adjacent cell that is not a wall. Note that the avatar can only be moved one cell at a time; clicking a wall or a non-adjacent cell won't move the avatar. Once the exit is located, the subject must exit the maze by placing the avatar in the exit (red) cell to move on to the next trail. The next trial will appear automatically as soon as the subject places its avatar on the exit cell.

After completing all trials, subjects answered a version of a Cognitive Reflection Test (CRT) [10]. They were then asked to provide their age, gender, and a free-form description of any strategies they used in the experiment. On the last page, subjects were shown how well they did with respect to other subjects on a scale of 1 to 10, with 10 being the 100^{th} percentile, and 1 being the 10^{th} percentile (Figure 3-6).

3.2 Experiment 1 — Inspect Left Right Bias in MST

We considered for a potential behavioral bias in navigating through the maze environment. In particular, we were concerned that perhaps subjects will be more biased towards favoring a particular direction, namely left or right. An eye-tracking study shows that people tend to spend disproportionately more time staring at the left side a web [9]. Another study that included a navigation task in a virtual environment

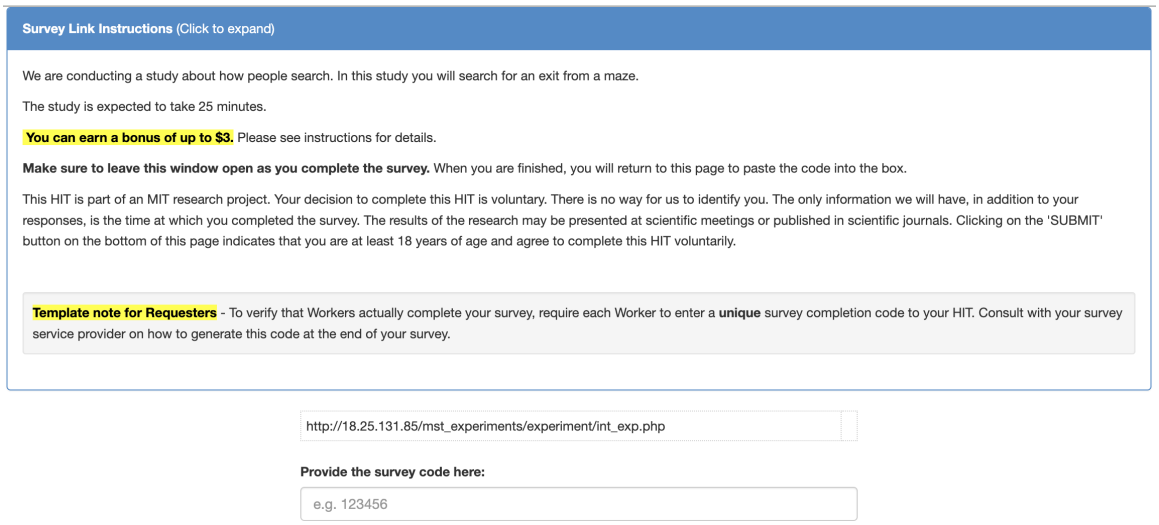


Figure 3-1: A subject first sees this screen on Amazon Mechanical Turk when they enter the experiment. They are instructed that the experiment is estimated to take at most 20 minutes, and that their responses will be used for research. When they enter the experiment by clicking on the link provided, the subject will be assigned a unique ID on the server side. After they go through all trials and answer the questions on the last page, they are asked to provide that unique subject ID in the box to exit the Mturk task.

concluded that people tend to turn left or right depending on their age or gender [42]. To investigate whether we should consider this factor for our experiment, we ran an experiment prior to the main one.

The same interface and format were used for this experiment; subjects went through a series of instruction pages and practice trials and moved on to the main trials. When subjects completed all trials, they were asked the same set of demographic and CRT questions. The difference between this and the main experiment is the set of mazes used. This experiment was aimed at investigating potential bias toward a particular direction when navigating in a 2D grid world.

3.2.1 Data Collection

To serve this purpose, this experiment is comprised of 40 mazes, where the first 20 mazes were designed uniquely, and the remaining 20 mazes were created by taking the mirror reflection of the original 20 mazes. Most mazes used in this experiment

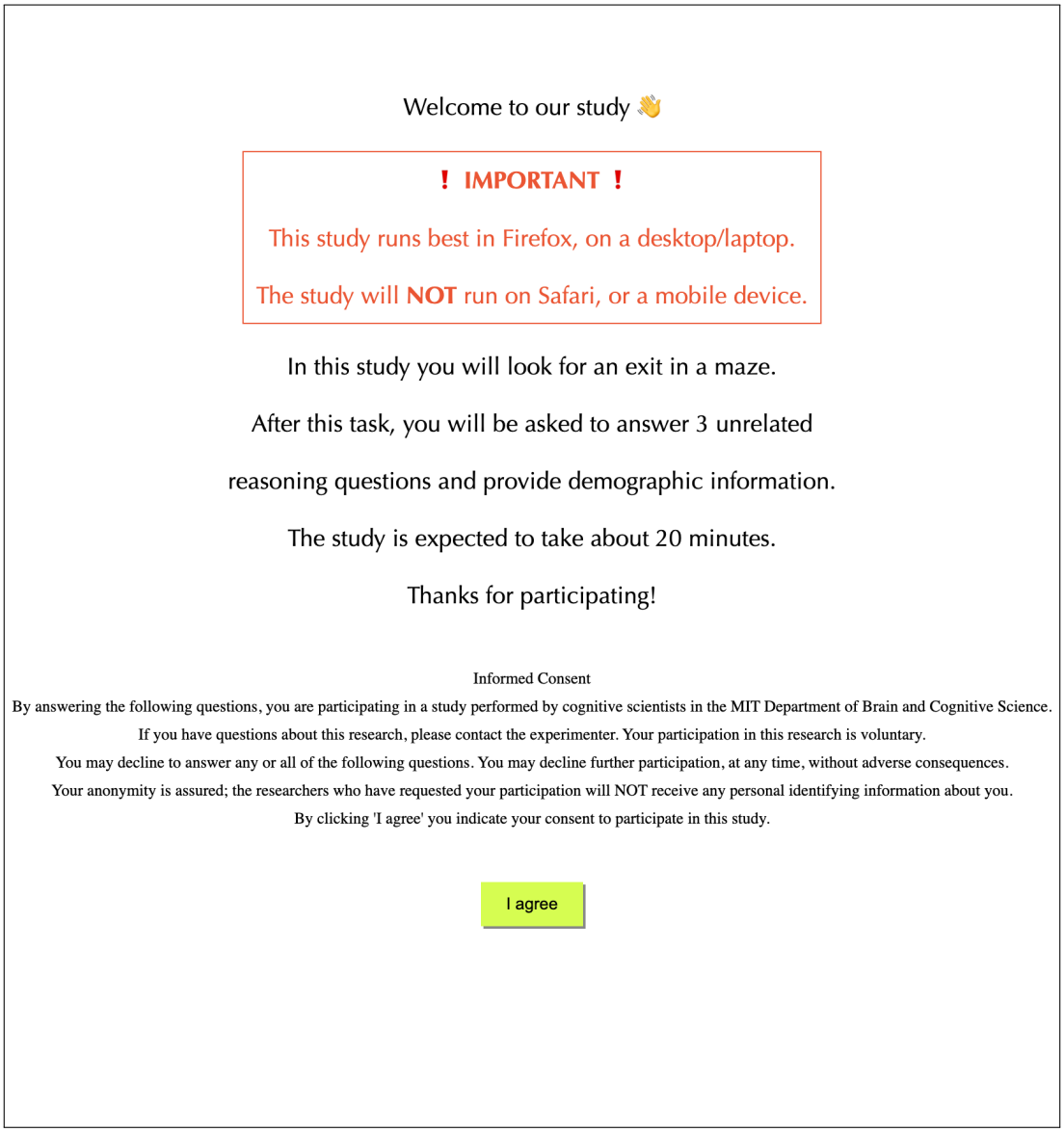


Figure 3-2: Screenshot of the consent page that a subject sees when they click the link to the experiment. They are once again informed that the experiment is expected to take at most 20 minutes, and that their responses will be used for research. They need to hit the green "I agree" button at the bottom to move on to the next page.

INSTRUCTIONS (PLEASE READ CAREFULLY)

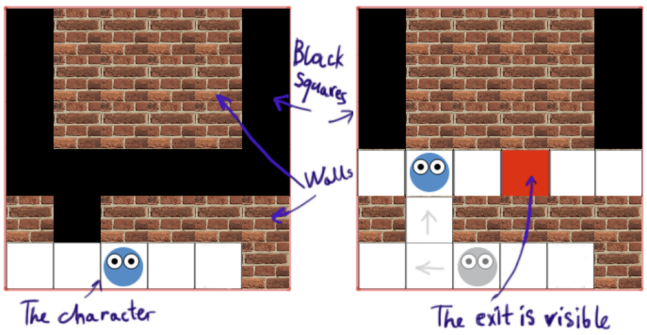
Your task is to exit the maze by reaching the red square in as few steps as possible.

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

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

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

A maze looks like this:



YOU COULD GET A BONUS!

Planning your path wisely will pay off.

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

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

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

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

Let's practice!

Figure 3-3: The instruction page provides a more detailed description of the structure and goal of MSTs. This explanation is accompanied with an example maze to illustrate the different cells that it is composed of, and how the avatar can move onto adjacent cells. The instruction also highlights that the goal is to locate and reach the exit in a maze in as few steps as possible, and that the exit is equally likely to be hidden under any of the black squares. The subject is also informed that they may earn up to $3 of bonus based on their performance. This bonus was stated explicitly to incentivize subjects to plan ahead when they make decisions.

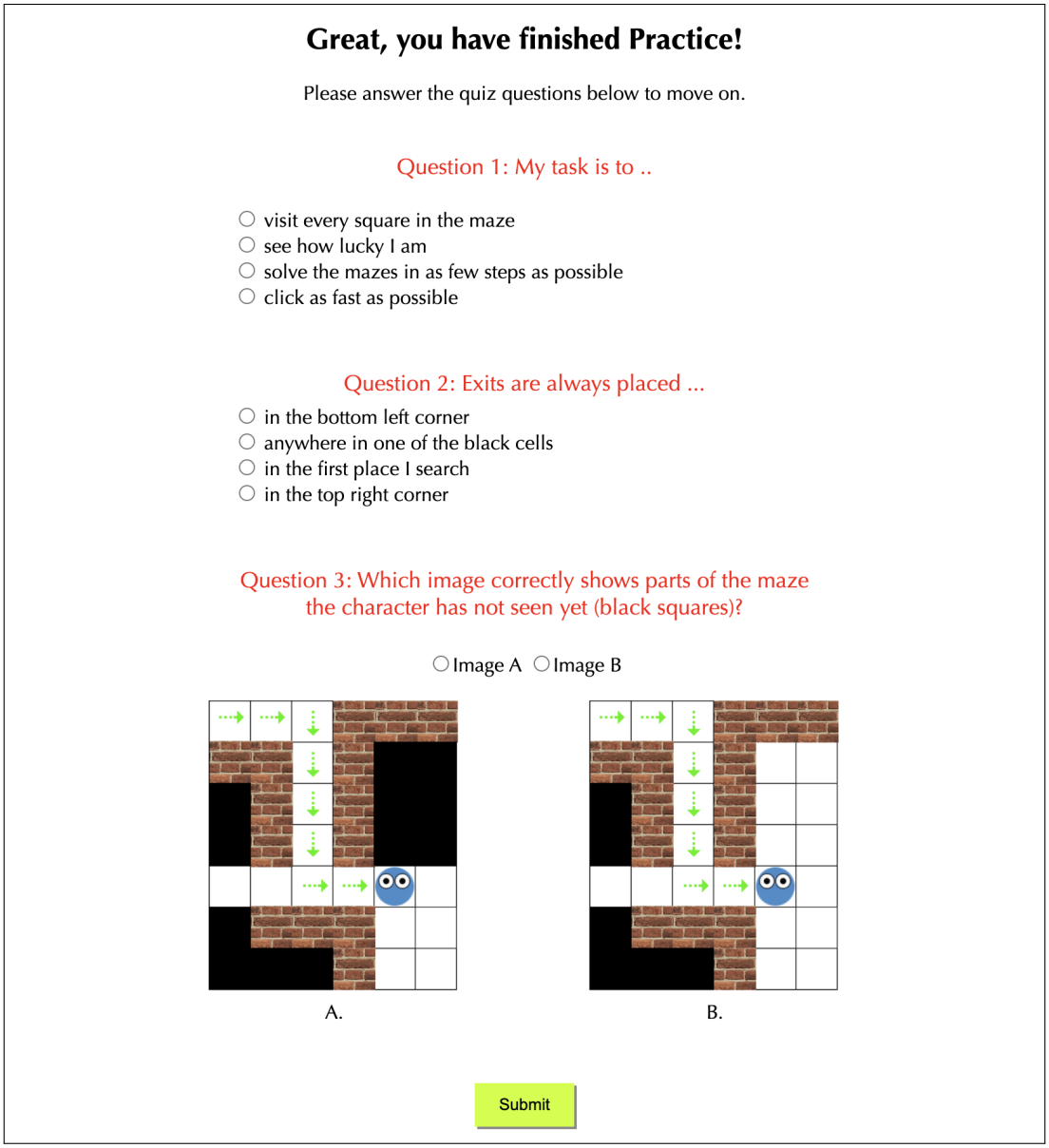


Figure 3-4: After 4 practice trials, a subject is presented with 3 questions about the experiment. They must answer correctly to all 3 questions in order to move on to the actual experiment. If they answer incorrectly to at least on of the questions, a pop-up screen appears to notify them that they must go through the practice trials again. This process is repeated until the subject submits the correct answer for every item. The correct answers to questions one, two, and three are the third, second, and second options, respectively.

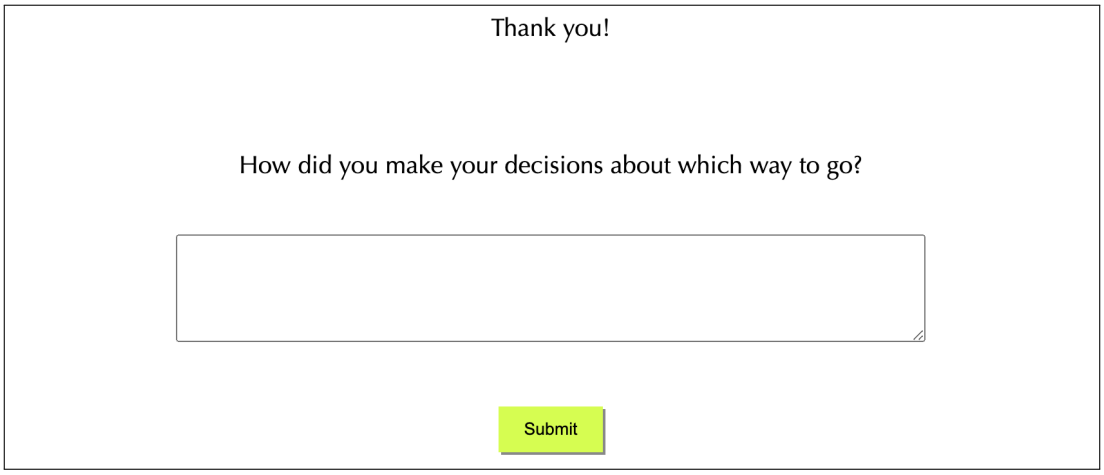


Figure 3-5: A subjects sees this screen as soon as they are done with the 23 MST trials. In the box provided, the subject is asked to explain the strategy they were using. There was no constraint on the number of words they may submit.

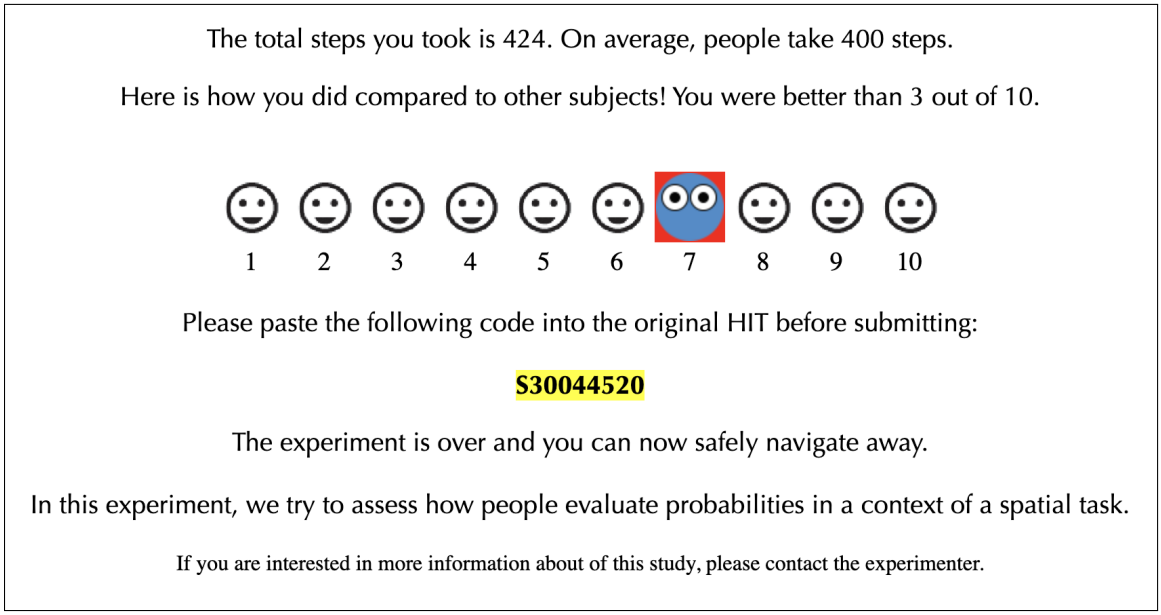


Figure 3-6: Final page of the experiment. A subject is given a brief summary of its performance in the experiment in terms of the number of steps. They are also informed of their relative performance compared to the other subjects from a scale of 1 to 10. Scale 1 indicates that the subject's performance was in the 90^{th} percentile, while scale 10 indicate that the subject's performance was in the 10^{th} percentile. In this page, they are also provided a unique subject ID that they are instructed to submit in the box on the Mturk screen.

are small, meaning that there are usually two options to choose from at the root node. We collected 104 subjects through Amazon Mechanical Turk. Among the 104 subjects, we rejected 4 subjects for going through the practice trials more than twice and used responses from 100 subjects to investigate any left-right bias in direction.

3.2.2 Statistical Test Result

We conclude that there is no statistically significant bias toward a direction, both on an individual subject level (Table 3.1) and a maze level (Table 3.2). In Table 3.1, we summarize the number of times a subject chose the node on the left or right with respect to the center of the maze. If the subject does not favor a particular direction, and instead chooses a node based on its features—such as the chance of finding an exit, or the number of steps required to take to get to the node—, then the two numbers should be comparable because for each uniquely designed maze there is a mirrored version of itself.

Indeed, the histogram summaries of the same result in Figure 3-7 visually show that no clear favor in direction exists. On the left is the distribution of the number of times a subject chose the node on the left subtracted by the number of times they chose the node on the right (M = −0.44, SD = 3.94). The distribution peaks around 0 and is tightly concentrated in the [-5, 5] range. On the right, we have two overlapping histograms, one for the distribution of subjects on the number of left decisions, and the other for the distribution of subjects on the number of right decisions. The distribution is almost identical, and a Kolmogorov test yielded that there is no statically significant difference in the distributions.

In Table (3.2), we summarize the distribution of subjects on the nodes in a maze and their counterparts in the mirrored version. For instance, in the first row of the first column of the table, we see that in Maze 1, 53 and 47 subjects chose 1 and 2, respectively. In the mirrored version of Maze 1, 56 subjects chose the counterpart of node 1 in the mirror-reflected version Maze 1, and 44 subjects chose the counterpart of node 2 in the mirror-reflected version Maze 1. From the summary, we can observe that the distribution of subjects is almost identical, and therefore conclude that there

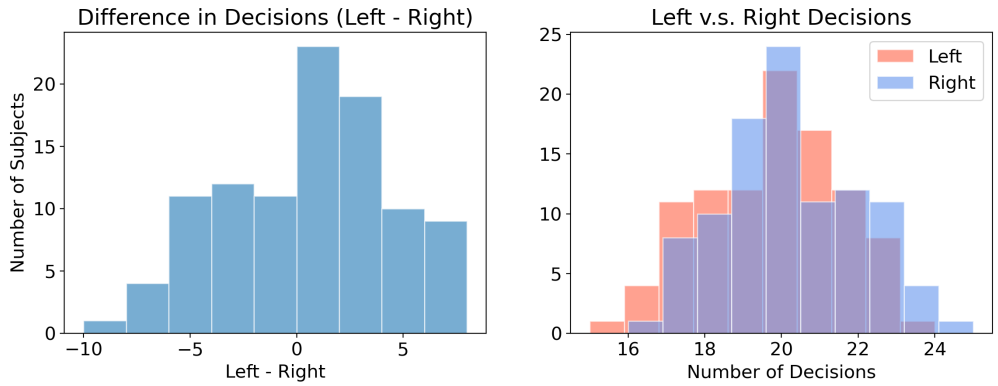


Figure 3-7: Visual summary of Table (3.1). (Left) Histogram of number of left decisions subtracted by right decisions ($M = -0.44$, $SD = 3.94$). (Right) Two overlapping histograms: one is the distribution of number of left decisions made by subjects, and the other is the distribution of number of right decisions made by subjects.

is no bias towards any direction.

3.3 Experiment 2 — Investigate Computational Principles

3.3.1 Stimuli

There are a total of 30 MST mazes used in the experiment: 7 were used as practice mazes, and 23 were used in the actual experiment. Each maze was first designed with a predetermined starting location. The exit position was assigned after the maze structure was set, for it had to be randomly selected among one of the hidden cells with uniform probability. All subjects saw the same set of mazes with the identical starting and exit locations, but the order in which they saw these mazes were randomly shuffled. The full list of 30 mazes is available in the Chapter 5.

3.3.2 Data Collection

We collected data from a total of 115 subjects (35 female, 80 male, $M(age) = 33$, $SD(age) = 9$) via Amazon Mechanical Turk. If a subject incorrectly answered the instruction questions more than twice, then we assumed that the subject did not

Subj No.	Left	Right	Subj No.	Left	Right	Subj No.	Left	Right
1	17	23	34	18	24	67	17	23
2	17	23	35	19	21	68	21	19
3	23	17	36	21	19	69	23	17
4	17	23	37	18	22	70	17	23
5	17	23	38	17	23	71	17	23
6	19	21	39	20	20	72	18	22
7	18	22	40	20	20	73	19	21
8	20	20	41	23	17	74	20	20
9	21	19	42	19	21	75	19	21
10	17	23	43	22	18	76	20	20
11	21	19	44	18	22	77	20	20
12	23	17	45	20	20	78	20	20
13	22	18	46	21	19	79	23	17
14	22	18	47	18	22	80	20	20
15	22	18	48	18	22	81	20	20
16	21	19	49	15	25	82	18	22
17	22	20	50	17	23	83	20	20
18	16	24	51	20	20	84	17	23
19	18	22	52	22	18	85	20	20
20	22	18	53	20	20	86	16	24
21	19	21	54	21	19	87	22	18
22	21	19	55	20	20	88	18	22
23	20	20	56	20	20	89	21	19
24	17	23	57	20	20	90	22	18
25	19	21	58	21	19	91	22	18
26	19	21	59	21	19	92	22	18
27	21	19	60	20	20	93	21	19
28	19	21	61	21	19	94	18	22
29	17	23	62	21	19	95	19	21
30	21	19	63	20	20	96	23	17
31	20	20	64	23	17	97	20	20
32	22	18	65	21	19	98	18	22
33	20	20	66	23	17	99	24	16
						100	21	19

Table 3.1: Column "Left" lists the number of times a given subject first chose a node that is on the left of the maze. Column "Right" lists the number of times a given subject first chose a node that is on the right of the maze. We hypothesize that a subject chooses a node based on its features, such as the chance of finding an exit. The result summarized in this table shows that the values in the two columns named Left and Right are mostly comparable, therefore supports that people are in fact not biased towards taking a particular direction.

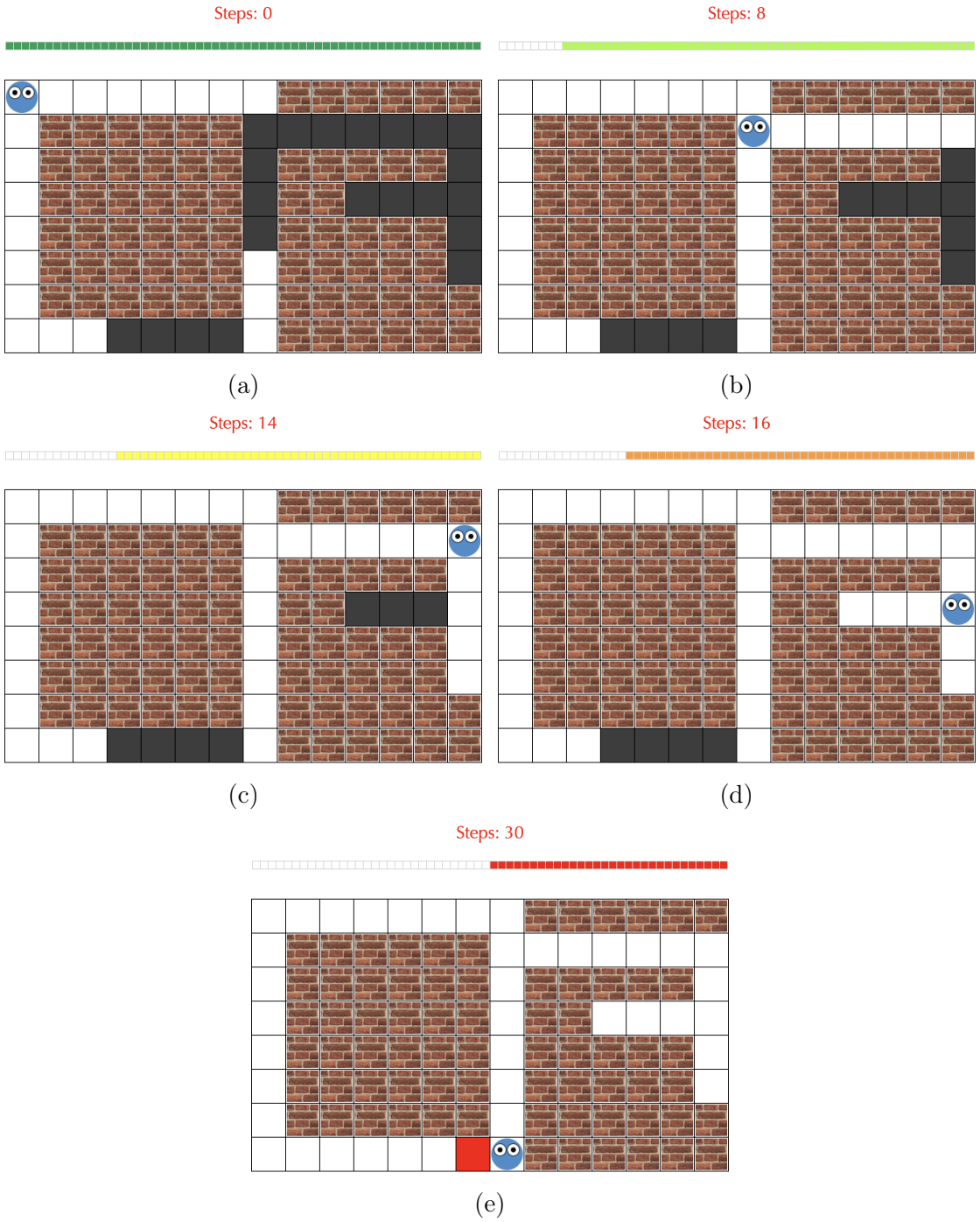


Figure 3-8: Example of what a subject sees on its screen during experiment trials. **(a)** At the very start of the trial, a subject starts from the top left corner and has accumulated 0 steps. **(b)** The subject decides to first visit the column of 4 black squares on the right. It travels to the right and accumulates 7 steps. Notice that the bar at the top shrinks and changes to a lighter color. As illustrated in **(c)**-**(e)**, the color of the bar also gradually changes to red. **(c)** The subject has located the exit (red) cell.

Maze No.	Original	Mirrored	Maze No.	Original	Mirrored
1	(53, 47)	(56, 44)	10	(39, 61)	(48, 52)
2	(68, 32)	(70, 30)	11	(30, 22, 48)	(29, 24, 46)
3	(29, 71)	(32, 68)	12	(67, 25, 8)	(67, 25, 08)
4	(71, 29)	(69, 31)	13	(65, 29, 6)	(64, 30, 6)
5	(60, 40)	(68, 32)	14	(37, 63)	(34, 66)
6	(39, 61)	(42, 58)	15	(51, 40, 8)	(51, 40, 9)
7	(52, 48)	(53, 47)	16	(67, 33)	(67, 33)
8	(47, 53)	(48, 52)	17	(30, 36, 34)	(42, 26, 32)
9	(60, 40)	(58, 42)	18	(42, 58)	(38, 62)
10	(51, 21, 24, 4)	(51, 21, 27, 2)	20	(46, 4, 13)	(45, 42, 13)

Table 3.2: Table summary of subject distribution on node choices in a maze and its mirrored counterpart. For example, starting from the root node of maze 1, 53 subjects chose node 1 and 47 subjects chose node 2. In the mirrored version of maze 1, 56 subjects chose the counterpart of node 1, and 44 subjects chose the counterpart of node 2. Overall, the distribution is comparable, implying that the aggregate human behavior was based on the structure of a maze rather than its orientation.

understand the task or was not invested enough in this experiment and excluded that subject's data for the analysis. We excluded 8 of such subjects and used data provided by 107 subjects in the analysis. Key statistics on subjects' experiment performances are summarized in Figure 3-9.

Parameter Fit

We would first like to comment on model parameter fits. Histogram summaries of parameter fits for the EU, PWU, DU and HSC models are presented in Figure 3-10. The median parameters fitted to the subject population were as follows: EU(τ = 1.936), DU (τ = 0.826, γ = 0.462), PWU (τ = 1.098, β = 0.317), and HSC (τ = 0.826, κ = 0.95).

We omit the fit for the Combined model to first focus on the effects of discounting and probability weighting individually. We also omit the fits for the other heuristic models because they are special cases of HSC. First, we can see that the distribution of fitted τ for EU is concentrated around a higher median when compared to the other two computational models and HSC. Recall that the value of τ is, to some extent,

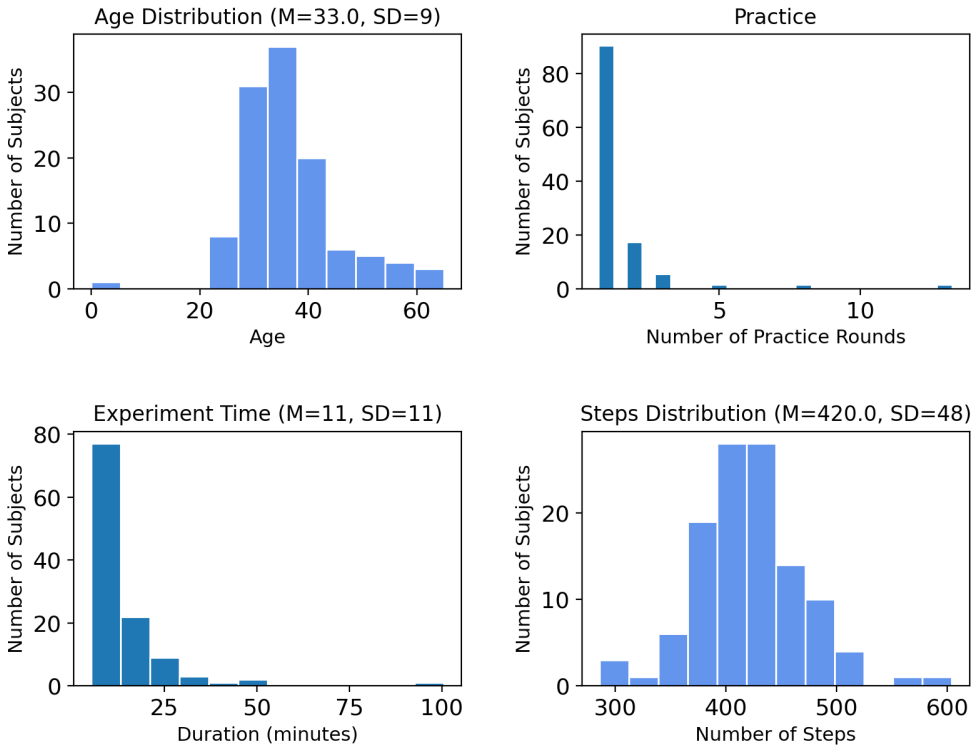
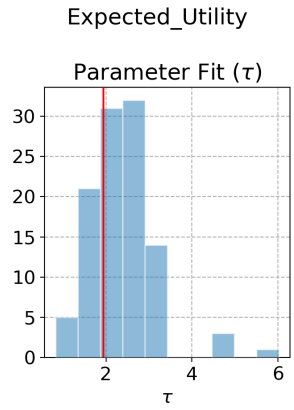


Figure 3-9: Summary of key statistics on subjects. Bar plot on the number of practice trials (upper right) that subjects went through shows that almost all subjects only needed on round of practice trials. Those who needed more than two trials of practice rounds were not considered for the final analysis. Histogram on total time spent in minutes (lower left) indicates that most subjects finished well under the expected time we stated in the instructions, and the distribution of the number of total steps (lower right) has a nice bell-shaped curve concentrated around 380 ~ 460 steps.

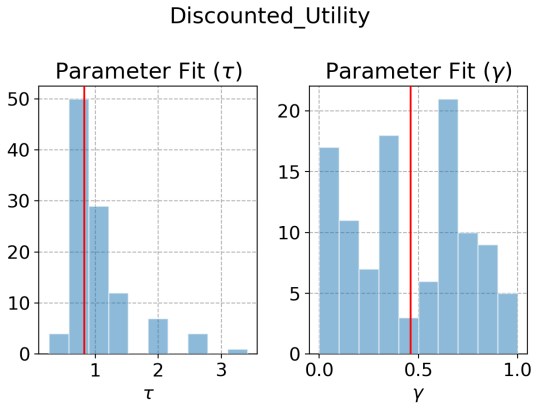
an indication of noise in the model—when $\tau \to \infty$, the model converges to a random model. This observation implies that the EU model is not flexible enough to explain human decision-making, and that at least one other parameter is needed to capture a missing component. That said, it is difficult to draw any conclusions just by the fit of τ values because the values of nodes are at slightly different scales for different models.

Total Steps Simulation

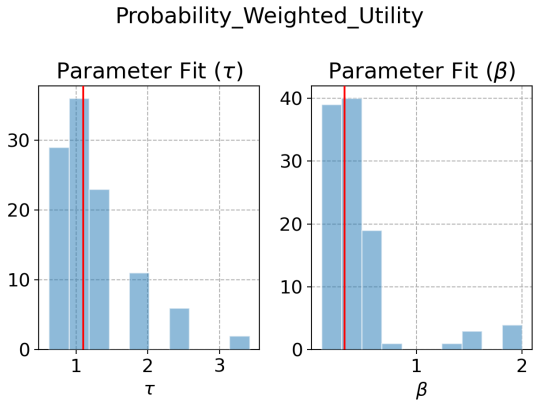
Before comparing model fits to subject data, we would also like to demonstrate that EU is indeed theoretically the best performing model for minimizing the expected



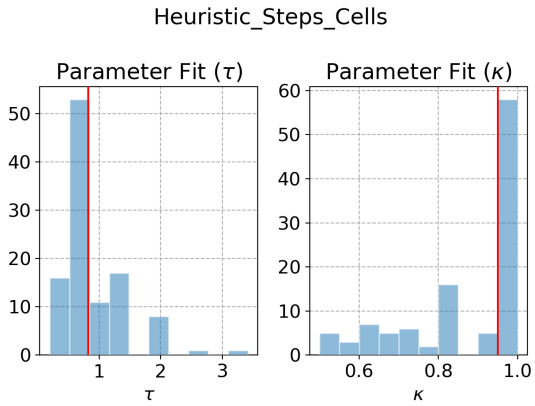
(a) $\tilde{\tau} = 1.936$



(b) $\tilde{\tau} = 0.826$, $\tilde{\gamma} = 0.462$



(c) $\tilde{\tau} = 1.098$, $\tilde{\beta} = 0.317$



(d) $\tilde{\tau} = 0.826$, $\tilde{\kappa} = 0.95$

Figure 3-10: Parameter fits for four models: EU, DU, PWU and HSC. Median of parameter values are specified in the caption of each subfigure. **(a)** The distribution of τ is generally lower for DU, PWU and HSC models, implying that noise tends to be lower for these four models. It is difficult to conclude that EU is not an adequately descriptive model for human decision-making from parameter fit result alone. But considering that the node values from four models are comparable, we can observe that the second parameter probably contributes to reducing the noise in model behavior. **(c)** One other interesting observation to make is that the median β value is very close to the proposed value in [38]. **(d)** Recall that when $\kappa = 1$, then the HSC model is equivalent to the HS model, and when $\kappa = 0$, then the HSC model is equivalent to the HC model. The distribution of fitted κ values imply that human subjects tend to put more weight on the cost to get to a node than the chance of finding an exit.

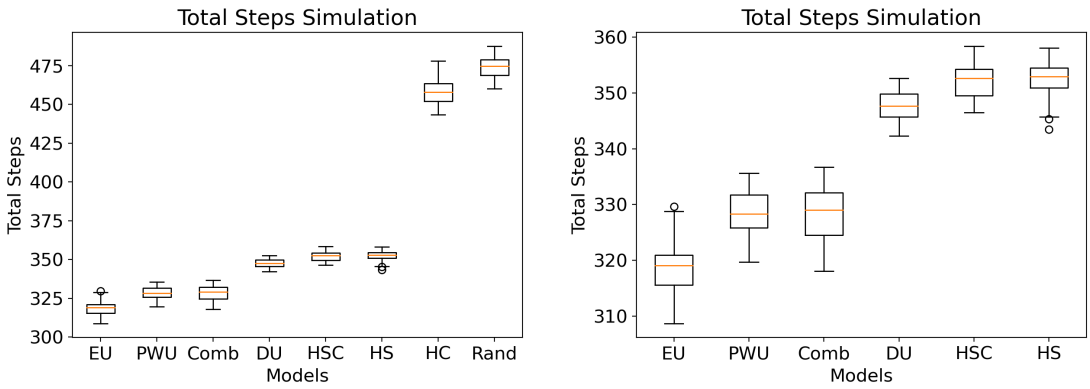


Figure 3-11: (Left) Box plot summary of total steps simulation of all models. The EU model in theory should be the best performing model because it minimizes the expected cost. This simulation verifies that the EU model is indeed the best performing model. (Right) Closeup of the result excluding the HC and random models. Heuristic models don't plan ahead, so it is reassuring to see that all planning models outperform the heuristic models in this simulation.

total steps to locate exits in MSTs. To show this, we simulate the total steps of each model with the median of parameters fitted to subjects. For a given maze, we randomly draw 100 exit positions, and we simulate the model to locate the exit for each randomly drawn position. The average of number of steps taken to locate the exits is the total steps for that maze. To get the overall total steps for a model, we sum up the total steps from all 23 mazes by repeating this process for each maze. Since all models presented in this work yield stochastic policies, we repeat the entire process for another 100 times to get a distribution of total steps. A box plot summary of total steps simulations in Figure 3-11 clearly shows that the distribution for EU is the lowest among all models. Other computational models also perform better than the heuristic models, as they should, but they do not outperform EU.

3.3.3 Model Comparison

Correlation on Aggregate Data

We assessed how well a model fits aggregate human behavior by studying the correlation between the model and human node values. Model node value was computed with the median of the fitted parameters. Meanwhile, we define human node value

of a node as the fraction of humans that chose that node from its parent node. For example, suppose that a decision node N has two children nodes, C_1 and C_2. If 100 subjects reached node N, and 20 and 80 subjects chose C_1 and C_2, respectively, then the human node values for C_1 is 0.2, and the human node value for C_2 is 0.8.

The bootstrap approach was used to find a distribution of correlation coefficients for each model. We first collected all 107 decision nodes that were encountered by at least 20 subjects during the experiment. At each iteration, we sampled a subsample of size 107 with replacement. The correlation coefficient between a model and human node values were found on that subsample. We repeated this process for 100 iterations. The distribution of correlation coefficient for each model is summarized as a box plot as shown in Figure 3-12.

This fit on aggregate subject behavior shows that the three planning models, Combined, DU and PWU are best at capturing the general human decision-making behavior. It is worth noting that the PWU and DU models are close to the Combined model even though they have one less parameter, implying that discounting or probability weighting alone is sufficient to capture human decision-making. A two-tailed t-test shows that the distributions of DU and HSC is significantly different, $t(108) = -0.396$, $p < 0.001$, and that the distributions of HS and EU are significantly different, $t(108) = -21.459$, $p < 0.001$. The fact that the distribution from the EU model is significantly lower than the other three computation models strongly implies that humans do not plan or behave based on the theoretically optimal policy. In fact, the two heuristic models, HSC and HS display much more accurate fits to the aggregate data than the EU model.

k-fold cross-validation for Individual Preference

Next, we evaluated the relative performances of the models using k-fold cross-validation. Note that by design, k-fold cross-validation implicitly takes over-fitting into account—a model that over-fits to training data will perform poorly on testing data. The models presented have different number of parameters, and when we assess the performance on one static data set, the model with the most number of parameters will always be

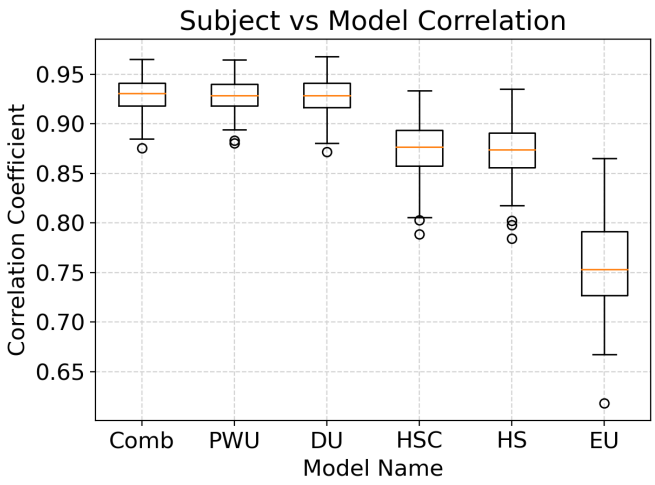


Figure 3-12: Box plot summary of the distribution of correlation coefficients between model and human node values for each model. Bootstrap approach was used to sample a random subsample of decision nodes at each iteration, and correlation coefficient was computed for that subsample. This subsampling of decision nodes with replacement was repeated for 100 iterations.

the best fitting model. This model selection method was chosen to make sure that the models are compared on a more equal footing.

To begin, we list out all decision nodes chosen by the subject, and randomly shuffle the order. The collection of decision nodes is then split into k roughly equal folds. At the k^{th} iteration, the k^{th} fold is set aside as the test set, and the rest of the folds are concatenated into a single training set. For each subject, we do parameter fitting to all models via maximum likelihood estimation. With the parameters fitted to the training set, we compute the average log-likelihood on the test set for each subject, and sum them over to get a single value for each model. The smaller this sum of average log-likelihoods, the better the model describes the decision-making behavior of human subjects.

To demonstrate the consistency of the result, we performed 3-, 4-, 5-fold and Monte Carlo cross-validations (Figures 3-13, 3-14). For Monte Carlo cross-validation, an 80%-20% split was used for sampling training and testing sets. The performance of a model was determined by the log-likelihood averaged over 50 iterations. In every case, the Random model consistently shows very low performance as expected,

implying that human decision-making is far from random. Planning models as a whole account for 73% ~ 79% of subjects. The fact that three planning models, PWU, DU, and Combined models outperform all heuristic models in all three cases implies that humans do take future consequences into account, i.e. humans do plan ahead. Meanwhile, the performance of EU is not as good as the other planning models; its value is either very close to or worse than HS or HSC, and it accounts for only 8% ~ 14% of all subject behavior. From this observation, we can infer that EU is not flexible enough to model human decision-making and that humans do not behave in the theoretically optimal way. On the other hand, simply increasing the number of parameters does not improve a model's performance. The Combined model is also consistently underperforming when compared to the DU and PWU models, conveying that discounting or probability weighting alone is sufficient to capture key characteristics of human decision-making.

3.3.4 Model Extension: Utility Transformation

Another key principle of prospect theory states that humans subjectively transform utility via some value function that is concave for gains, convex for losses, and steeper for losses than for gains [18]. In this section, we present two additional models that take utility transformation into account to further investigate the generalizability of principles of prospect theory.

The full form of a utility transformation function for relative gains used in prospect theory is $U(v) = \lambda \cdot v^{\alpha}$, where parameter λ controls the steepness, α controls the concavity of the expression, and input v denotes the relative gain with respect to the worst value within a maze. Recall that all node values are passed through a softmax function parameterized by τ. Here, τ also dictates the steepness or shallowness of the gradient between values. Since the roles of λ and τ overlap, we only consider one parameter, α, and use the expression $U(v) = v^{\alpha}$ as the utility transformation function.

We define two additional models to investigate utility transformation in the context of an MST. One is Compressed Utility (CU) which only transforms utility, and second is Full Prospect Theory (FPT) which takes into account both utility transfor-

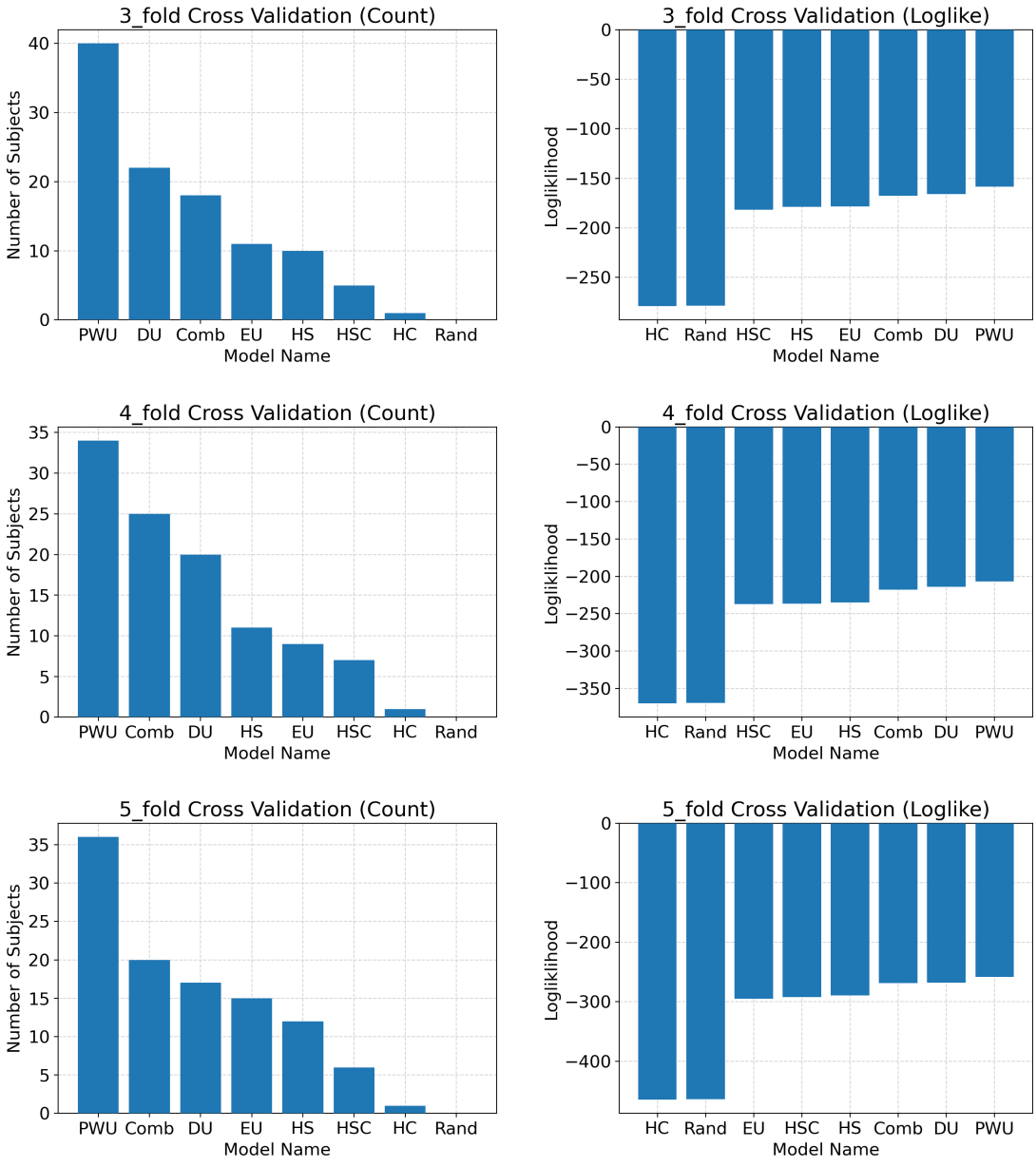


Figure 3-13: Plots of k-fold cross-validation. The three plots on the left column summarizes the relative performances of all models by displaying the number of best fitted subjects. The three plots on the right column displays summarizes the relative performances of all models by plotting the sum of average log-likelihood of all subjects.

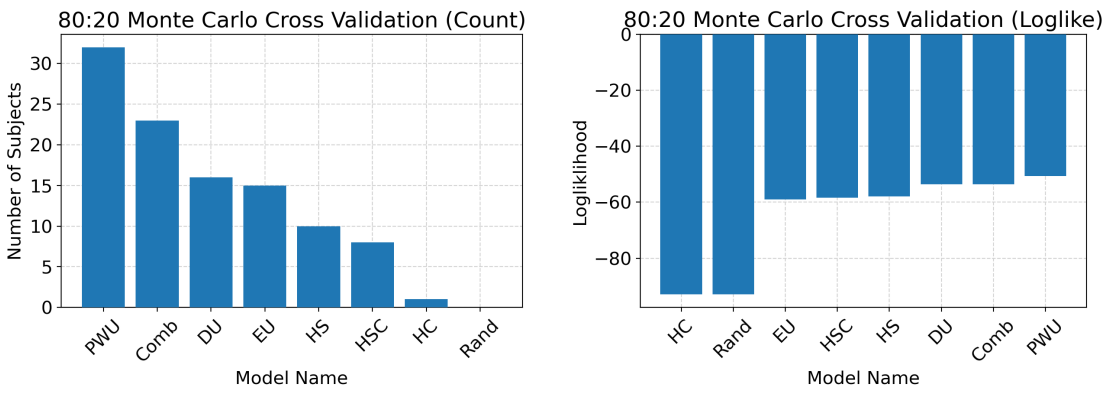


Figure 3-14: Results of Monte Carlo cross-validation with 80%-20% split and 50 iterations. The plot on the left summarizes the number of subjects best fitted to a given model, and the plot on the right summarizes the log-likelihood averaged over each iteration.

mation and probability weighting. The formal definitions are defined below:

Compressed Utility $$V(N_i) = U(s_i) + p_i U(e_i) + (1 - p_i) \min_{C_j \in C(N_i)} V(C_j)$$

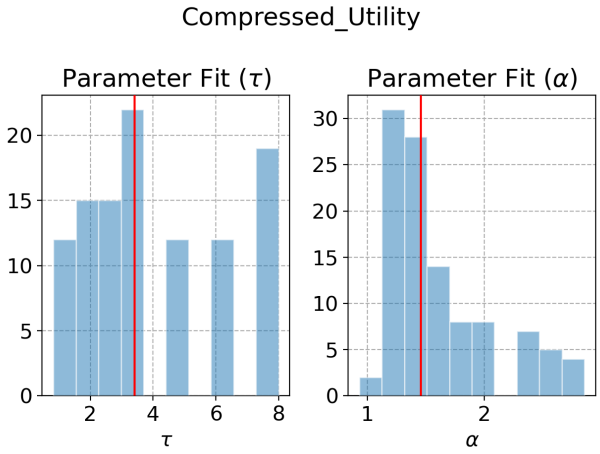
Full Prospect Theory $$V(N_i) = U(s_i) + \pi(p_i) U(e_i) + (1 - \pi(p_i)) \min_{C_j \in C(N_i)} V(C_j).$$

Model Analysis

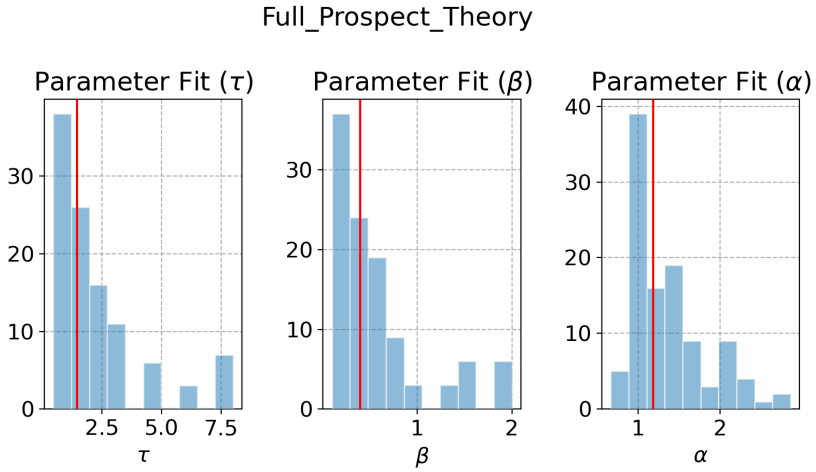
In Figure 3-16 is the histogram summary of fitted parameters for CU and FPT. The medians of fitted parameters for CU are $\tilde{\tau} = 3.415$, $\tilde{\alpha} = 1.5$, and the medians of fitted parameters for FPT are $\tilde{\tau} = 1.936$, $\tilde{\beta} = 0.399$, $\tilde{\alpha} = 1.1$. In both cases, we observe that fitted α values are mostly above 1, indicating that the utility transformation functions are concave, consistent with the principle for relative gains in prospect theory. However, most fitted α values are concentrated around 1, and the medians of α values are close to 1 for both models, implying that the utility transformation functions are approximately linear.

Bar plot in Figure 3-16(a) summarizes the number of subjects best fitted by a given model. PWU is the best model for almost 30% of all subjects, and it is still the most popular model. CU is the next most popular model, and combined with FPT, the two models best describe 24% of all subjects. In Figure 3-16(b), however, the performance

of CU and FPT significantly dwindles. The difference in the sum of log-likelihoods between EU and FPT is ΔLL(EU - FPT) = 26.985, and the difference between EU and CU is ΔLL(EU - FPT) = 77.349. This cross-validation result suggests that parameter α is fitting noise and causes the two models to overfit.

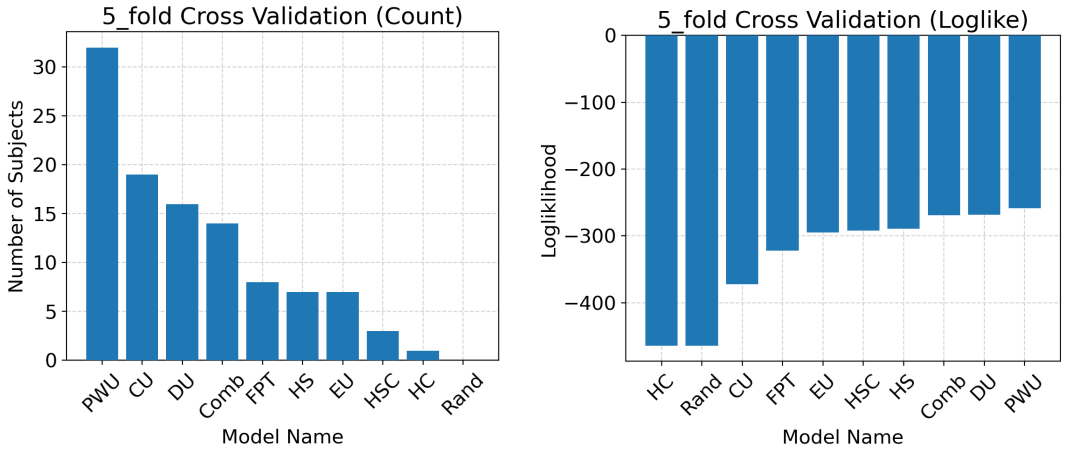


(a) $\tilde{\tau}$ = 3.415, $\tilde{\alpha}$ = 1.5



(b) $\tilde{\tau}$ = 1.936, $\tilde{\beta}$ = 0.399, $\tilde{\alpha}$ = 1.1

Figure 3-15



(a) (b)

Figure 3-16: Plots of 5-fold cross-validation. **(a)** Number of subjects that was best fitted by a given model. PWU still the most popular model that best fits more than 30 subjects. CU is the next popular model. Performance of FPT is not as good as the other planning models, but it fits more subjects than heuristic models or EU. **(b)** Sum of log-likelihoods, averaged over each fold. CU and FPT combined best describes 24% of the subjects,

3.4 Conclusion

Contribution

We first recognized the demand for descriptive human behavioral models in engineering applications. As such, we formalized three cognitive principles—noisy utility maximization, discounting, and probability weighting—, and examined their generalizability in MSTs, a sequential spatial decision-making task. We ruled out a potential behavioral bias in subjects when interacting with experiments conducted on screens, namely the left-right bias. We verified that there is not statistically significant bias in choosing the left or right action, but rather chose their moves based on their assessment of the associated utility.

We observed that a combination of noisy utility maximization, discounting, and probability weighting achieved best prediction of aggregate human behavior. EU was not the best at describing aggregate human behavior, even when compared to heuristic models. Moreover, k-fold cross-validation analysis shows that PWU significantly

outperformed all suggested models, implying that probability weighting is a key factor in human decision-making model. Our work is the first to evaluate probability weighting, a central principle of prospect theory, in the context of spatial search. We demonstrate that probability weighting is a fundamental to human decision-making, and generalize beyond gambles scenarios for which it was originally proposed [17].

After observing the importance of probability weighting in modeling human decision-making, we further extended the models to include utility transformation. We proceeded with the understanding that an MST may not have the correct structure for properly studying utility transformation because there is neither an explicit reward nor a clear reference point to which the values should be anchored to define relative losses and gains. Indeed, the results of parameter fit indicated that utility transformation functions are approximately linear, and the performances of the two newly introduced models were significantly worse than EU. As part of our future work, we could modify the current structure of MSTs to have explicit losses and gains to study utility transformation in human decision-making.

Individual Differences

Informally inspecting 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 our Heuristic Model. Second, a smaller group reported preferring larger rooms, but made no reference to distance. The third most popular answer referred to minimizing steps while maximizing the sizes of visited rooms, suggesting approximately optimal planning. Lastly, a few individuals reported guessing, or using intuition. Our computational models predict some of these observations. For example, discounting could make future backtracking seem negligible, and probability weighting could exacerbate this effect by overestimating the value of smaller rooms. However, it is unclear whether, and how much, humans are aware of their cognitive strategies.

Limitations

The MST takes a step toward measuring planning in natural behaviors, however it is impossible to do justice to the complexity of human planning with just one task. The use of two-dimensional grid-worlds with only one goal is convenient for developing a rigorous modeling methodology, and evaluating the core computational principles. However, these principles should be further evaluated in street networks, environments with multiple goals, and multi-dimensional topologies. Our analysis also leave open the possibility that the planning strategies evolve over time, for example, due to people optimizing over the set of alternative heuristics [15], due to fatigue, or as a result of learning efficient state-space representations [37]. Further, we derive the state-space representation by observing the distribution of human decision times, without specifying how the state-space representations should be defined in a general case, which is an active area of research in robotics. These limitations need to be addressed in future work.

3.5 Future Work

Future studies will also investigate how human planning computations may evolve over the time course of the experiment. For instance, individuals may shift towards more elaborate planning with experience, or shift toward a simpler heuristic with fatigue. It is significant that most human deviations from the optimal planner in the experiment were observed in smaller but closer observations. Very infrequently such "mistakes" involved moving toward a larger room, despite this decision being not optimal. We intend to investigate this relationship further in future work.

Our study attempts to discover the probability weighting effect of prospect theory from a sequential decision-making task and aims to explain irregularities in human decision-making. Indeed, the analysis of the experiment result reveals that the PWU model outperforms EU and also appears to do better than the Discount model. However, it is clear that the discount effect and probability weighting have overlapping effects, and we cannot determine the precise impact of each phenomenon on decision-

making. Future work will focus on designing and analyzing a richer set of experiments that involves higher decision space. There, we aim to investigate how probability weighting loads differently from discounting effect.

Another direction of future research is to investigate whether the computational principles proposed in this work generalize to other tasks. While many studies from behavioral economics focus on one-shot decisions in the context of monetary gambles, our experiments introduced sequential decision-making tasks that involve time and space. We are exploring to expand our experiment set to more complex structures that closely resemble real-life scenarios. We also hope to apply our findings in more specific and realistic applications to demonstrate the generalizability of these descriptive models.

Chapter 4

Simulation of Human-Robot Collaboration in Urban Rescue Mission

4.1 Background and Motivation

With the recent, several groundbreaking developments in robotics and artificial agents, combined with the concern with disasters in urban areas, the demand for autonomous robots in rescue missions is increasing. Indeed, there are platforms such as RoboCup-Rescue competitions that promote and gather researchers in this field [21], and simplified simulation tools are being developed to help people focus on high-level solutions [5]. With the right help from these agents, the human rescuers should be able to work more safely and efficiently, and consequently, increase the survival rate of victims. For seamless cooperation from robots in such dire and hazardous situations, establishing the right model for human decision-making should be an integral part in designing our robot rescuers.

Here, we consider a hypothetical environment wherein a human and an artificial agent cooperate to save a victim in an MST. The human and robot agents are placed in a partially observable gridworld that is identical to MSTs in structure. The human's

task is to locate and save the victim placed somewhere in this environment. The robot's goal is to make the search more efficient as a team by predicting the human's next move and taking an action that is complementary to that of the human. There is no doubt that the robot agent's ability to make accurate predictions on its fellow human agent's move is critical in increasing victim survival rate and saving cost.

4.2 Simulation Setup and Example

The robot-assisted rescue mission works as follows. To begin, the human agent makes a decision based on its assessment of the environment and situation. Note that the human agent does not take the robot agent's intention or behavior into account when making this decision. The robot agent simultaneously makes an autonomous decision on where it should go to save the victim. To make this decision, the robot agent tries to predict what its human partner would do in the given situation, and make the next best decision based on that prediction. For this human-robot collaboration to work effectively, it would be best if the robot agent's decision *complements* the human agent's decision, i.e. the paths taken by the two do not overlap. That is why the robot agent's understanding of its human partner is crucial. With the correct descriptive model for human decision-making, the robot partner will be able to make more accurate predictions of the human agent's decision, and therefore become a more reliable partner to collaborate with.

Here, we will walk through an example of how the human-robot cooperation changes based on the robot's assumption of the model for human decision-making. Consider the maze illustrated in Figure (4-1). The victim is placed at position (row, column) = (1, 0), and both the human and robot agents start at position (5, 4). The interaction between the two agents at each iteration is illustrated in Figure (4-4) and Figure (4-5). In both cases, the human agent behaves based on the PWU model with the median of the fitted parameters from the experiment. The robot agent is the *rational cooperator* to the human, thus making complementary decisions based on the EU model with $\tau = 1$. In Figure (4-4), the robot incorrectly assumes that the

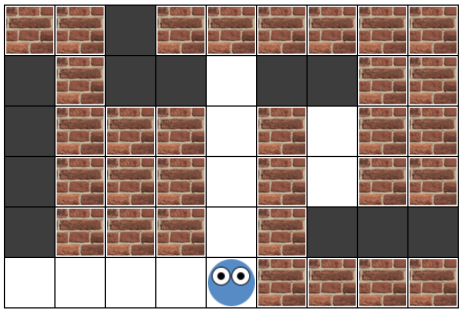


Figure 4-1: Maze used for the demonstration of human-robot collaboration in a rescue mission in MST style.

human agent behaves based on the EU model. The parameter used is once again the median of the fitted parameters from the experiment. in Figure (4-5), the robot correctly assumes that the human agent behaves based on the PWU model.

At the initial position, each agent has a choice between two nodes; one is to go up to position (1,4) and reach node 1, and the other is to go left to position (5,0) and reach node 2. The values of these two decisions for each model are summarized in Figure (4-2). According to the EU model, choosing node 1 is preferable to choosing node 2; however, the experiment shows that only 22 subjects chose node 1 while 85 subjects chose node 2. Indeed, the PWU model puts more weight on node 2 for $\beta \sim 0.35$, thus correctly predicting the aggregate human behavior. In Figure (4-5), we can see in the left plot that the human chose to first go up, and that the robot was able to correctly predict this decision. As a result, it was able to first go left and explore part of the maze that the human will not have reached yet. Since the robot is able to find the subject at the first iteration, no further search by the human-robot team is needed.

When the robot incorrectly predicts the human's move due to an incorrect assumption, it ends up taking the same action as the human agent as shown in Figure (4-4 (a)). The benefit of having two agents searching a maze is so to make the search more efficient by splitting the work. But if the two agents end up moving in the same direction, the benefit that comes from having two agents is lost. Since the victim was not located in the first iteration, this human-robot team must go through a second one. Once again, the preferred decision by most subjects was to take a left; while

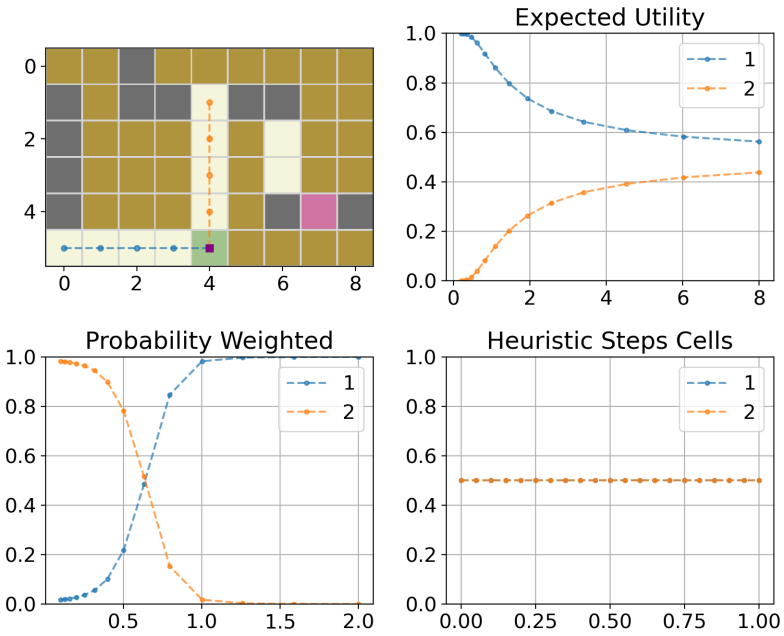


Figure 4-2: The upper left plot shows that node 0 (root node) has two children nodes: nodes 1 and 2. Node 1 is reached by going left (blue path) to reach position (row, column) = (5,0) and observing cells at positions {(1,0), (2,0), (3,0), (4,0)}. Node 2 is reached by going up (orange path) to reach position (1,4) and observing cells as positions {(1,2), (1,3), (1,5), (1,6)}. From the plot on the upper right, we can see that the EU model will always prefer node 1 over node 2. However, 22 subjects chose node 1 while 85 subjects chose node 2 in the experiment. Indeed, the PWU model (lower left plot) puts more value on node 2 for $\beta \sim 0.35$. Note that the HSC model (lower right plot) is incapable of distinguishing the two nodes because they both take equal number of steps to reach and observe the same number of hidden cells.

only 26 subjects went right to position (1, 6) to reach node 4, 56 subjects went left to position (1, 2) to reach node 5. The PWU model captures this behavior correctly for β 0.35, while the EU model can only prefer node 4. As a result, the robot ends up going in the same direction as the human in this iteration as well. To sum, the human was able to locate the victim in the maze in only 4 steps when its robot collaborator correctly predicted the human's next move, compared to 10 steps when it did not.

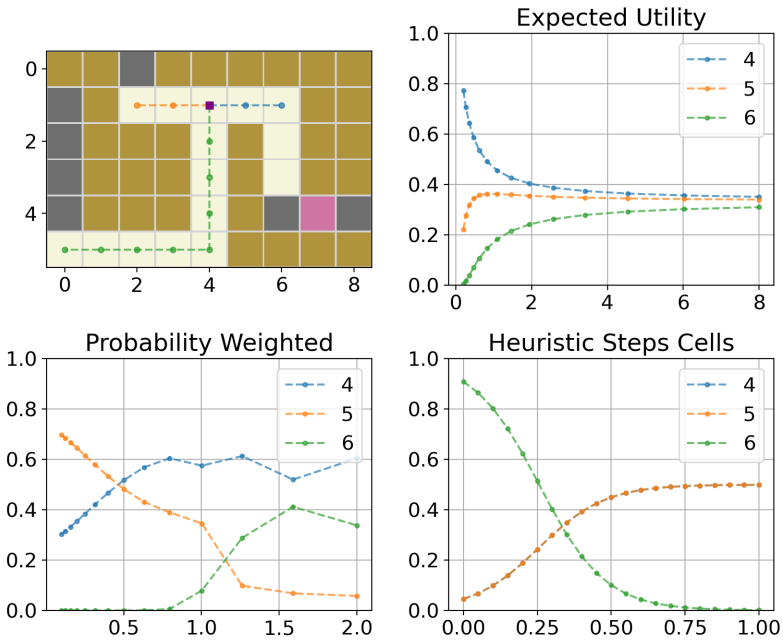
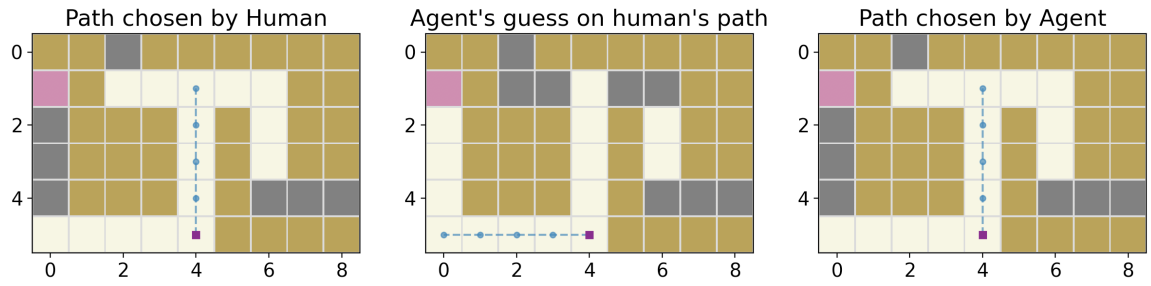
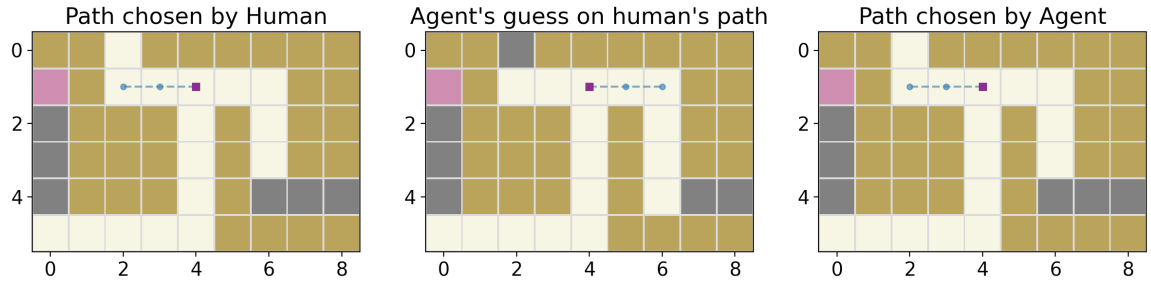


Figure 4-3: The upper left plot shows that node 2 has three children nodes: nodes 4, 5 and 6. Node 4 (blue path) is reached by going right to reach position (1, 6) and observing cell at position (4, 6). Node 5 (orange path) is reached by going left to reach position (1,2) and observing cells at position (0, 1). Node 6 is reached by moving down, then left to reach position (5, 0) and observing cells at positions {(1,0), (2,0), (3,0), (4,0)}. From the plot on the upper right, we can see that the EU model will always prefer node 4 over all other nodes. But only 29 subjects chose node 4, while 56 subjects node 5. No subject chose node 6. Once again, the PWU model (lower left plot) slightly prefers node 5 over node 4 for $\beta \sim 0.35$.



(a) The human agent first decides to go up. But the robot agent incorrectly predicts that the human will first go left. Based on this incorrect prediction, the robot agent tries to take the best complementary action, which is to go up. The robot is not able to help the human at this iteration because both ends up taking the same action.



(b) The human agent then decides to visit the single black square on the left. The robot makes the same mistake as it did in the previous iteration, and again ends up taking the same action as the human.

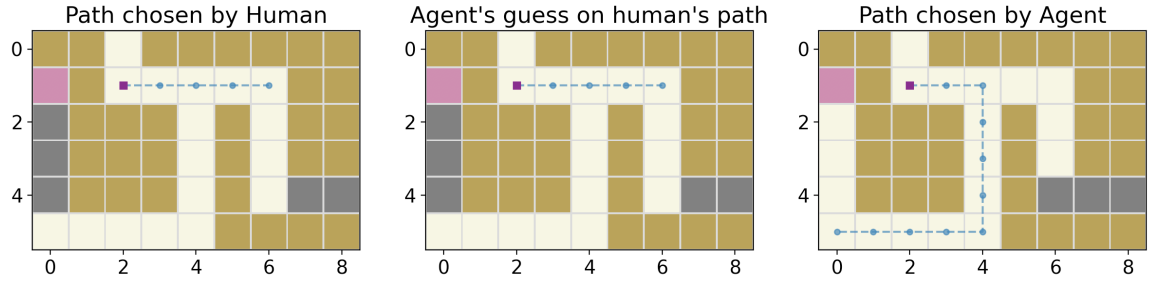


Figure 4-4: Each row of plots illustrate the human's decision (left), robot's prediction on the human's decision (middle), and the robot's decision based on its prediction (right) in a single iteration. The purple square indicates the position at which an agent makes the decision, and the blue dotted line is the path taken by the agent. The human agent makes decisions based on the PWU model, but the robot agent incorrectly assumes the human agent makes decisions based on the EU model. As a result, the human-robot team takes three iterations to locate the victim, and it costs the human agent 10 steps.

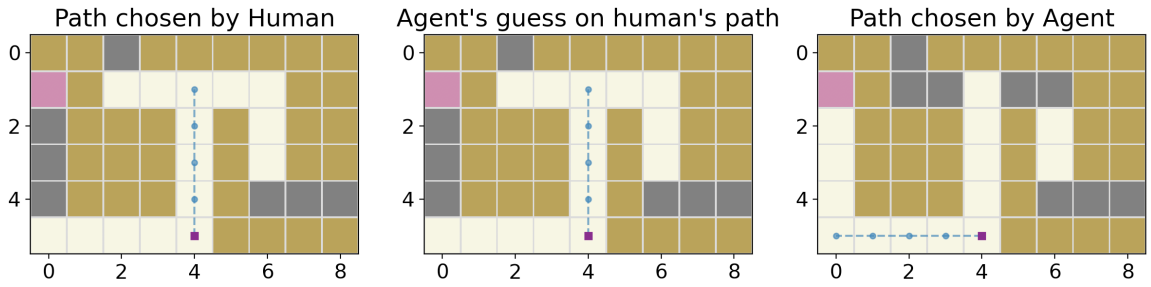


Figure 4-5: The robot agent correctly assumes that the human agent behaves based on the PWU model. Unlike the case in Figure (4-4), the robot agent is able to predict that the human will first go up and decides that going left would be the best complementary action to take. As a result, the human-robot team is able to find the victim in one iteration.

4.3 Simulation Result

To demonstrate the importance and effectiveness of the correct understanding of human decision-making in this rescue mission scenario, we compare the total steps taken by the human agent in the mission. The mission consists of the same 23 mazes used in the experiment. In each maze, we count the number of steps taken by the human until the victim is located, and we sum those steps across all 23 mazes to get a single total steps to count for a single mission. If the robot agent often makes correct predictions of the human's decision, then the human-robot team should not need many iterations to find the victim, thus reducing the number of steps that the human must take. On the other hand, a consistent prediction mistake by the robot would increase the number of times when the decisions overlap, thus increasing the number of iterations and total steps taken by the human.

We repeat this simulation for 100 times to get a distribution of total steps. Note that the stochasticity comes from the fact that decisions are not deterministic. When an agent is presented with a list of nodes to choose from, the agent makes a weighted random selection, where the weights are the values of nodes. The expectation is for the mean of the distribution of total steps to be much lower when the robot has the correct assumption about the human decision-making behavior than that of the case when it does not.

Indeed, the histogram in Figure (4-6) precisely illustrates this expectation. The

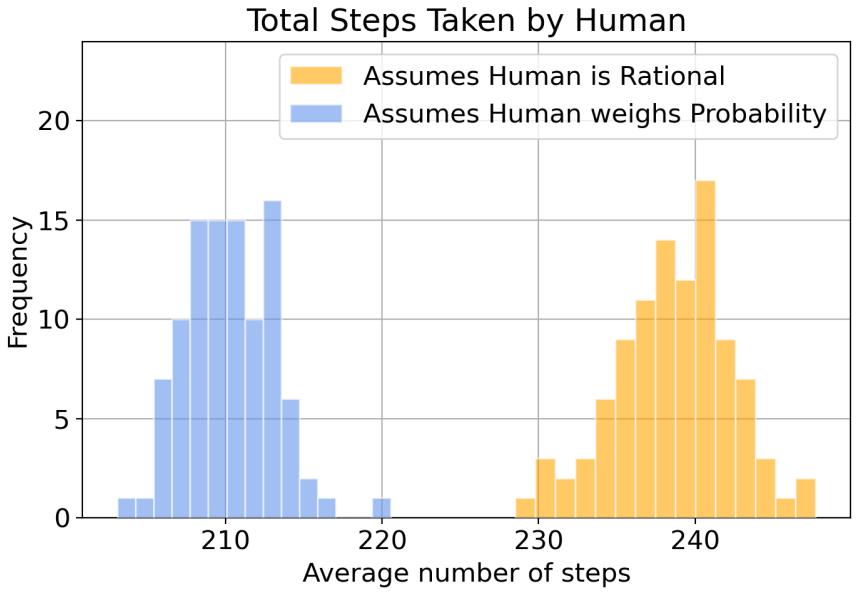


Figure 4-6: Histogram summary of total steps simulation for the two cases: one wherein the robot correctly assumes that the human agent bases its decision on the PWU model, and the other wherein the robot incorrectly assumes that the human agent behaves based on the EU model. We can visually see that the distribution of total steps for the correct assumption is much lower than for the case of the incorrect assumption. A two-sided t test rejects the null hypothesis that the means of the two populations are equal ($t(99) = 60.80, p < 0.01$).

peak of the distribution when the robot has the correct assumption ($M = 210$ steps, $SD = 2.9$ steps) is about 30 steps lower than the peak when the robot does not have the correct assumption ($M = 238$ steps, $SD = 4.7$ steps), and the two distributions do not overlap. A two-sided t test rejects the null hypothesis that the means of the two populations are equal ($t(99) = 60.80, p < 0.01$). Therefore, there is enough evidence to support that the average total steps taken by the human are much less when the robot collaborator makes the correct assumption about the human's decision-making behavior than when it makes an incorrect assumption.

Next, we verify that the total steps trend upwards as the assumed model for human decision-making gets further away from the actual model used. We do the same 100 simulations to get a distribution of total steps for ten different β values: {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}. The result of this simulation is summarized as a box plot in Figure (4-7). As expected, the distribution of total steps is similar to the

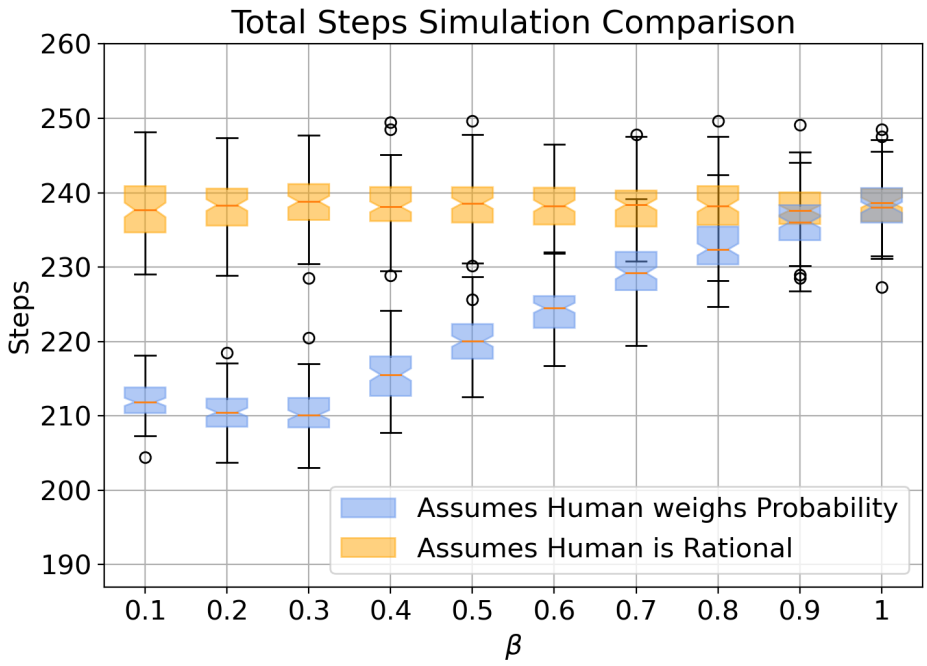


Figure 4-7: Box plot summary of total steps comparison for various β values. In all cases, the human made decisions based on the PWU model with median of the fitted parameters, and the robot agent made rational decisions based on the EU model with $\tau = 1$. For cases when the β value is close to the median, we can observe that the total steps tend to be at its lowest, mostly ranging from 208 ~ 215 steps. Note that $\beta = 1$ is equivalent to the EU model, and we indeed see that the distribution of the total steps for the two cases becomes closer as $\beta \to 1$.

one shown in Figure (4-6) when β is close to the actual value used by the human. However, we see that the distribution trends upwards as the β value diverges from the actual value, and finally converges to the one when the robot assumes the EU model for human decision-making. Recall that the PWU model is precisely the EU model when $\beta = 1$.

4.4 Conclusion

A simplified urban rescue mission simulation was set up to demonstrate the importance of understanding descriptive human behavioral model in the context of human-robot collaboration. We demonstrated that the robot's correct understanding

of human decision-making significantly improved the performance of the collaborative task.

4.4.1 Comments on Computational Cost

The definition of the models presented in this work are based on the environments' decision trees, so their computational complexities mostly depend on the size of the tree. Note that the branching factor for a node is determined by the *structure* of the given environment rather than its physical size. More formally, let k be the number of children nodes at the root decision node. Then the size of this environment's decision tree is bounded below by $\sum_{i=0}^{k} k!/(k-i)!$. But this lower bound is defined based on the assumption that the branching factor strictly decreases at each time a new observation is made, and characterizing the upper bound without knowing the structure is impossible. Consider the mazes illustrated in Figure (4-8) as examples. Both mazes have branching factors of 2 at their root nodes, but the maze on the left has a total of 5 nodes in its decision tree, while the maze on the right has a total of 4928 nodes in its decision tree. The cause of this large difference lies in the structures of the two mazes; on the left, the branching factor decreases as the avatar makes new observations, but on the right, the branching factor is non-decreasing for some iterations as the avatar makes new observations. With all that said, using a decision tree allows us to convert continuous space into discrete space, and it is still significantly more economical than using belief states, which are commonly used for partially observable environments. With belief states, the number of nodes required increases roughly by a factor of $m \times n$, where m and n are the numbers of rows and columns of the maze grid. Therefore, models that take in decision nodes as states can substantially decrease the computational complexity when computing for policies, and can be used as a more feasible solution for practical use in real-life scenarios.

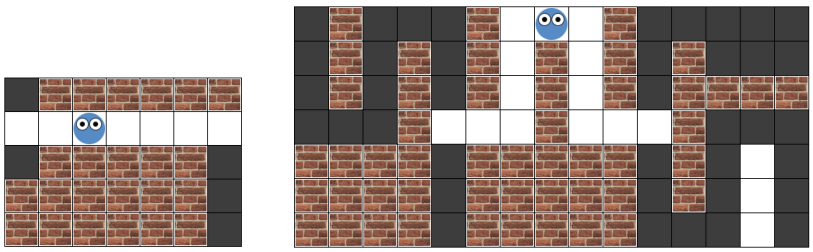


Figure 4-8: (Left) Total number of nodes in the corresponding decision tree is 5. (Right) Total number of nodes in the corresponding decision tree is 4928.

4.4.2 Future Work

The current simulation assumes that the robot and human agents decide on their actions simultaneously. This simplification was made under the assumption that there should be no delay in making decisions in a hazardous environment, but it also implies that the robot cannot receive any information about the human agent's behavior before making its prediction. In a real-world setting, it may be that the robot may delay its decision by a short amount of time to take in some information about the human's next action in order to make a more accurate prediction and improve the long-term collaborative performance.

Throughout the simulation, we also assume that the robot agent has a fixed behavioral models for humans. A human may not learn or significantly update its strategy when it is under time pressure in risky, unpredictable environments, but developing a more flexible robot agent that can dynamically adapt to its human partner opens up new potentials for the two to work on complex tasks that take longer to complete [39], [19].

Appendix A

List of Mazes in Experiment 2

In this section, we list all 7 practice mazes and 23 main mazes used in the experiment. The design object of practice mazes is to get subjects accustomed to MSTs, so these practice mazes are small in size, and there are at most two options to choose from at any decision node. These mazes are also designed to help subjects build intuition on how black squares are uncovered. The 23 main mazes are much larger in size, and the branching factor at decision nodes is usually three to four. Model selection analysis is performed on subject data on these 23 main mazes.

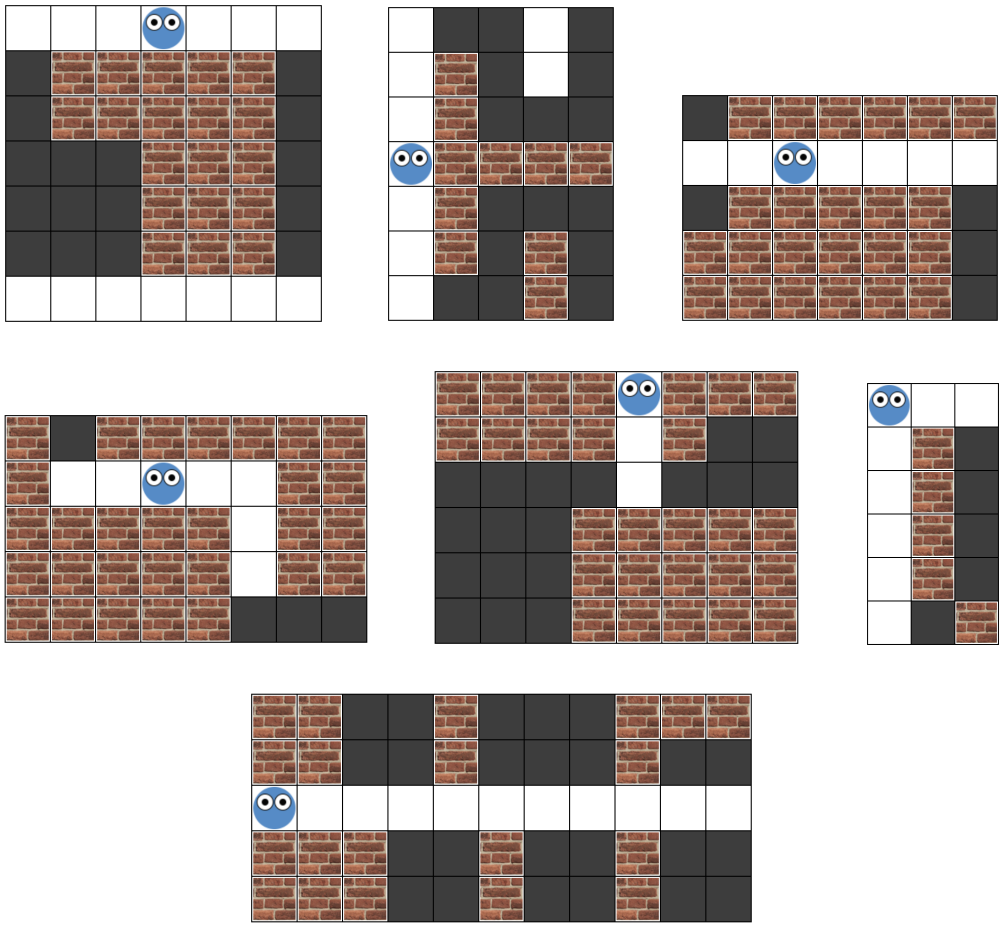
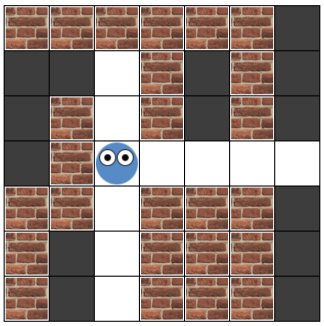
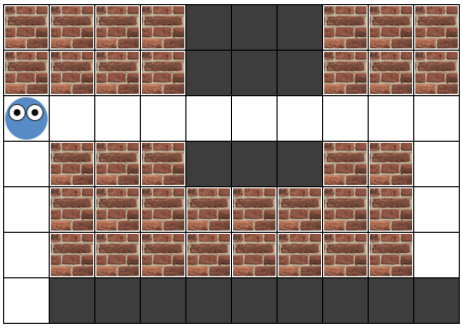
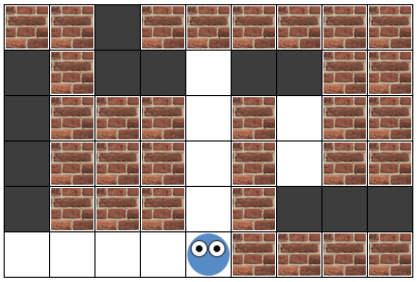
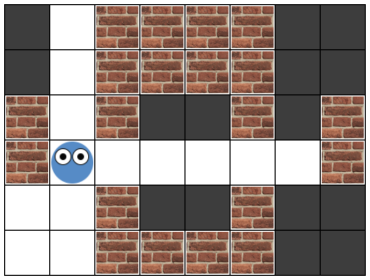


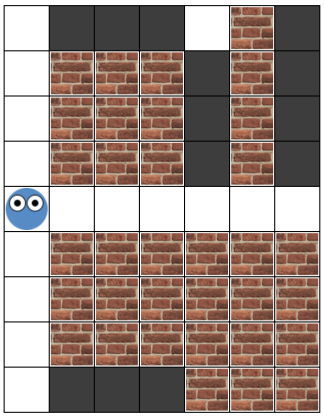
Figure A-1: 7 practice mazes. Each practice maze is small in size, and there are at most 2 options to choose from at any node. The purpose pf practice mazes is the get subjects accustomed with MSTs and build intuition on how black squares are uncovered. All subjects saw the same set of practice mazes, and unlike the main trials, these mazes were *not* randomly shuffled.

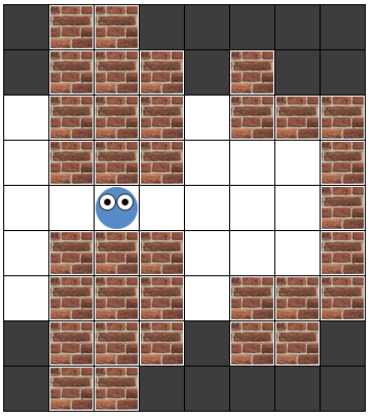


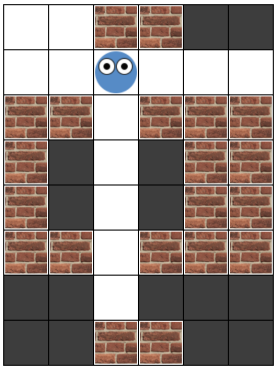


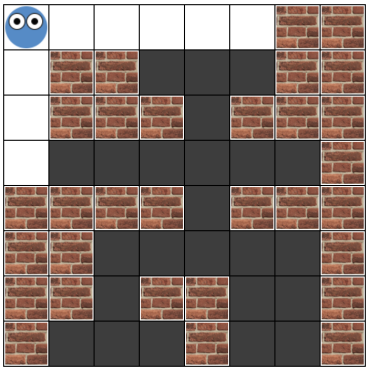


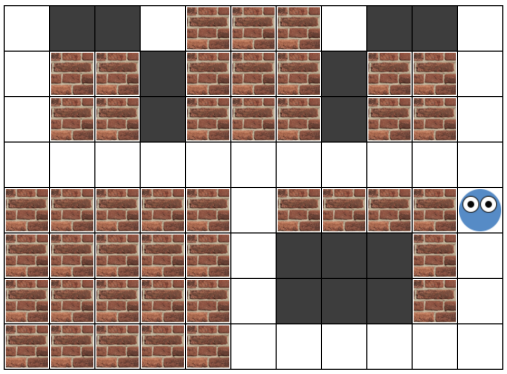


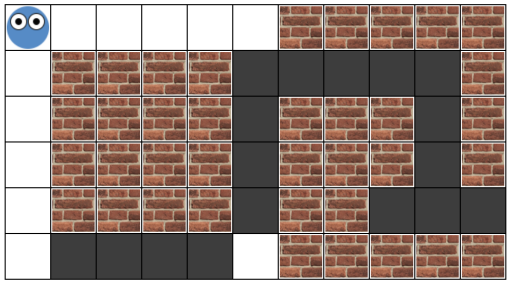


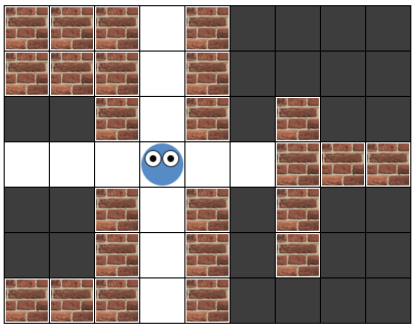


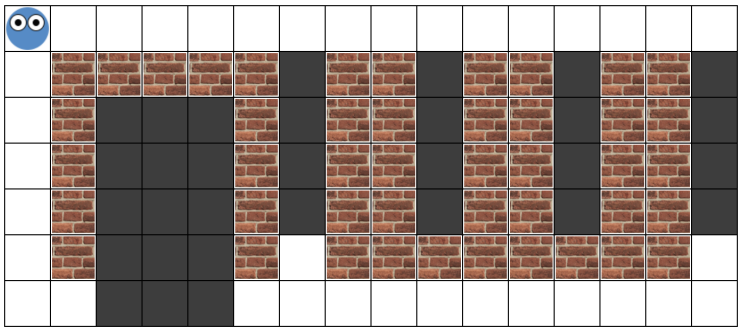


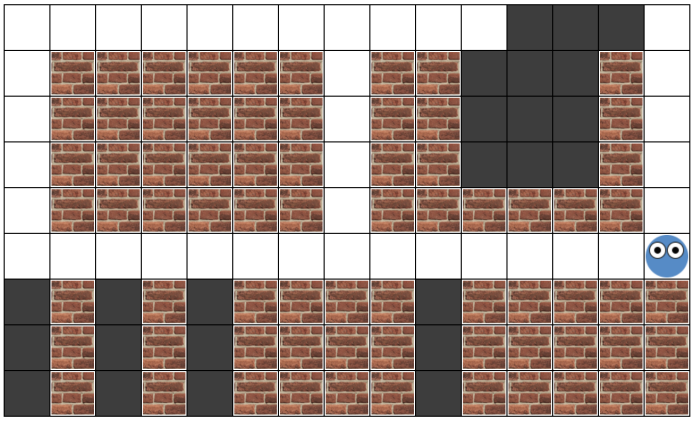


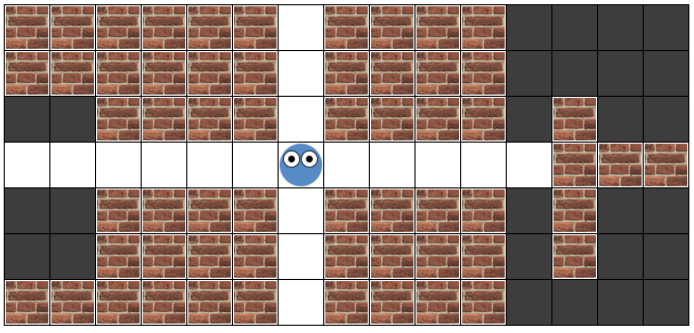


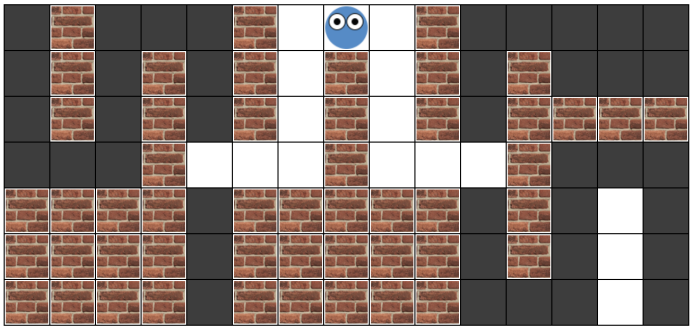


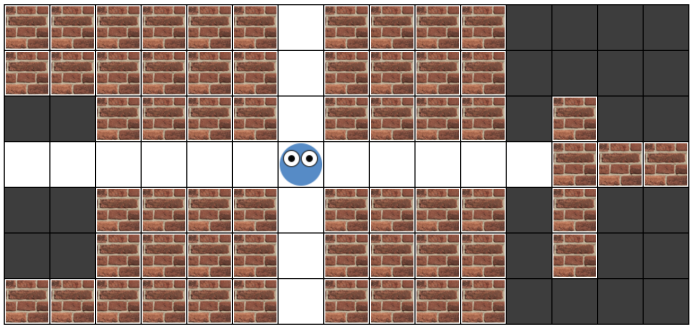


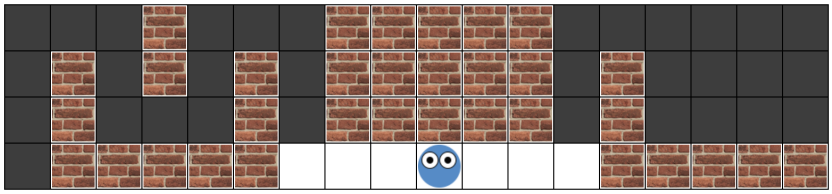


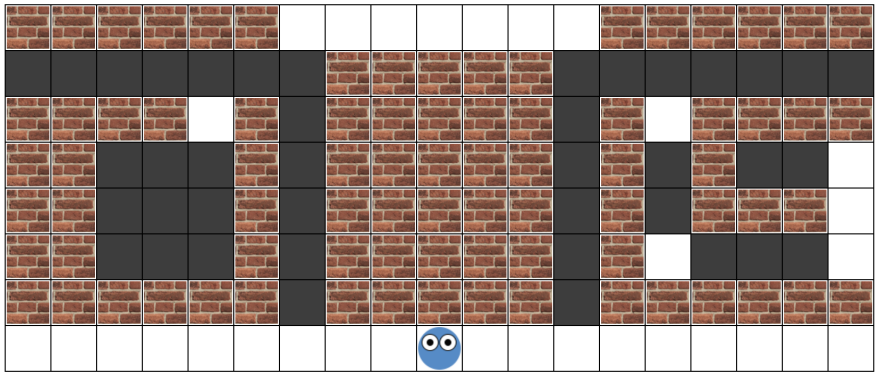


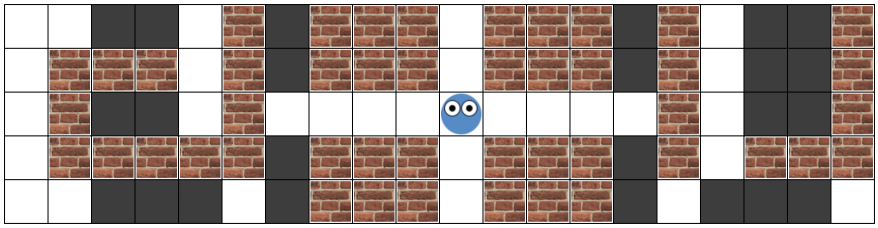












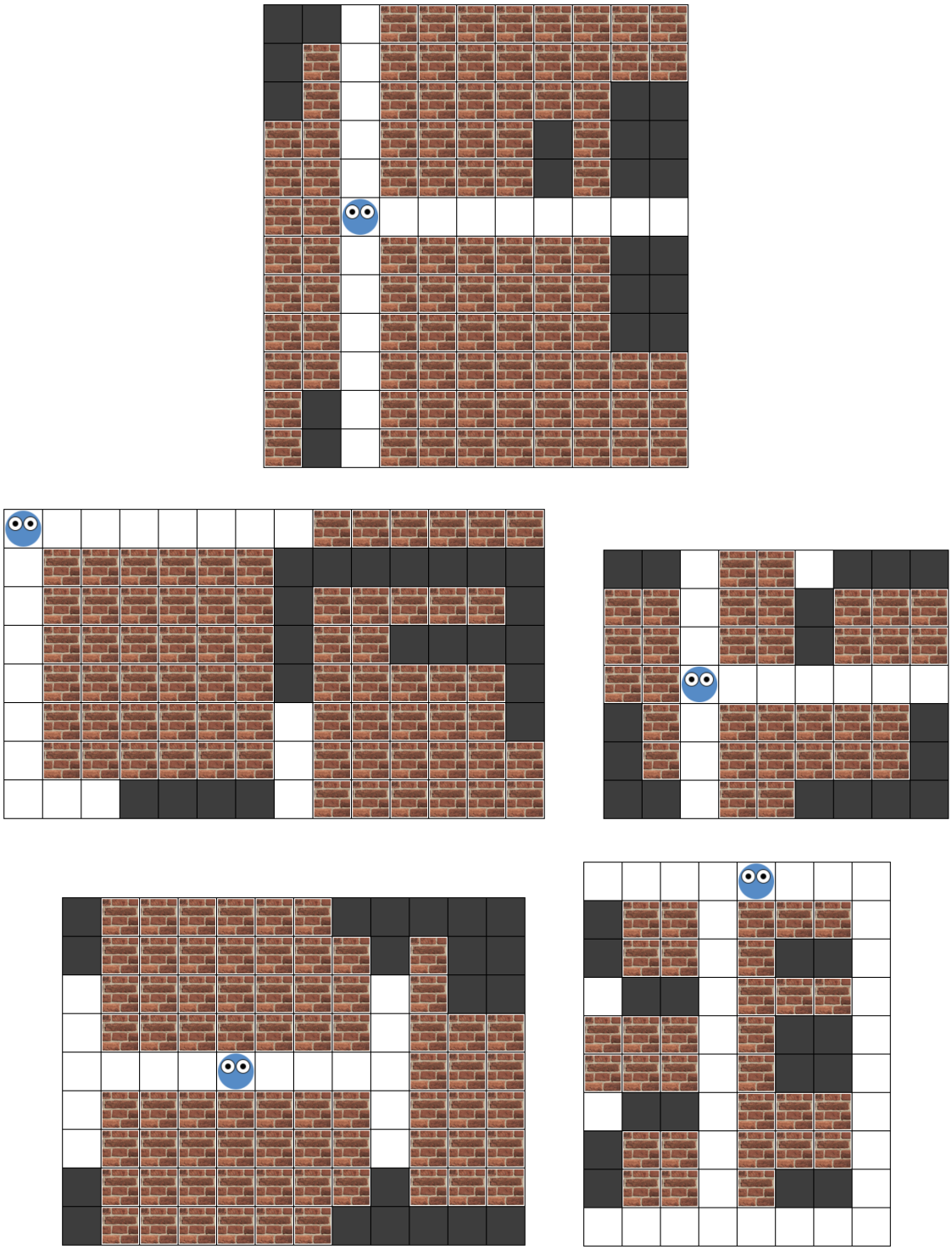


Figure A-5: 23 mazes that were used in the actual experiment. Model selection analysis was done on subject data from these mazes. These are generally larger in size, and there could be up to 4 options to choose from at a node.

Appendix B

Brief Proof of Probability Weighting in Prospect Theory

Expected utility is a weighted average of utilities of all possible outcomes, where the weight of an outcome is its probability of occurring. This model has been the dominant descriptive model used for understanding decision-making under uncertainty; however, empirical studies have shown that the expected utility theory is inadequate to model individual preferences and decision-making behavior ([18], [3], [26]). Prospect theory, developed by Kahneman and Tversky, is an alternative to expected utility theory that gained much traction in the late 20^{th} century. One of the conclusions that the authors reached from their empirical study is that people weigh an outcome by its *decision weight* rather than its probability of occurrence.

Decision weight is a nonlinear transformation of a probability, and it does not obey many of the usual probability axioms or properties. For instance, they violate the second probability axiom that states that the probability of the set of all outcomes is 1. Prospect theory asserts that this nonlinear transformation of probabilities made by people overweight small probabilities and underweight moderate to large probabilities. To support their claims, the authors designed a series of gambles roughly of the form $(p, x; q, y)$, which denotes a prospect of earning x with probability p and a prospect of earning y with probability q. See Tables (B.1 - B.3) for examples that were used in the original work by Kahneman and Tversky. In the experiment, subjects were

presented with a series of these two-choice gambles, and they were asked to choose the one they would prefer to play. The expectation is that a subject will choose a gamble that seems to return a higher utility and that this utility is not expected utility.

Let $\omega : [0,1] \to [0,1]$ be the weighting function that maps probabilities to decision weights. To first show that $\omega(p) > p$ for small p, consider the two choices presented in Table (B.1). According to expected utility, a subject should not have any preference for one gamble over the other. However, 72 subjects indicated that they would choose gamble A over gamble B. Somehow, earning \$5,000 with a probability of 0.001 is more favorable than receiving \$5 with certainty. This subject preference behavior implies that:

$$\omega(0.001) \cdot v(5,000) > v(5), \quad \text{or}$$
$$\omega(0.001) > v(5)/v(5,000) > 0.001 \quad (\because \text{concavitiy of } v, \text{ and } v(0) = 0).$$

A:	5,000 with probability	0.001	B:	5 with probability	1

Table B.1: Corresponds to Problem 14 in [18]. Gamble A was preferred by 72 subjects, and gamble B was preferred by 28 subjects.

More generally, we can show that the weighting function ω is concave for small values of p. Consider another two-choice problem from the original paper presented in Table (B.2). As with the previous case, expected utility theory would conclude that both choices are equally good. However, gamble A was the favorite of subjects; out of 100 subjects, 73 subjects responded that they would choose gamble A over gamble B for the higher prospect despite the chance of winning this higher reward

being halved. This subject behavior implies that:

$$\omega(0.001) \cdot v(6,000) > \omega(0.002) \cdot v(3,000), \quad \text{or}$$
$$\omega(0.001)/\omega(0.002) > v(3,000)/v(6,000).$$

Here, v is the value function that maps the absolute reward value to some subject value that an individual feels. We will omit the details about v since it is not relevant to this thesis, but it is important to note that v is a concave function for non-negative inputs. For simplicity, we will only consider positive prospects in this section. Since v is concave, and $v(0) = 0$, we can make the following observation:

$$\omega(0.001)/\omega(0.002) > v(3,000)/v(6,000) > 0.5, \quad \text{or}$$
$$\omega(0.002 \cdot 0.5 + 0 \cdot 0.5) > \omega(0.002) \cdot 0.5 + \omega(0) \cdot 0.5.$$

By definition of concavity and using the boundary condition $\omega(0) = 0$, we can conclude that ω is a concave function for small values of p. Similar questions and arguments can be used to demonstrate that $\omega(p) < p$ and that ω is convex for moderate to large values of p.

A:	6,000 with probability	0.001	B:	3,000 with probability	0.002

Table B.2: Corresponds to Problem 8 in [18]. Gamble A was preferred by 73 subjects, and gamble B was preferred by 27 subjects.

Finally, the authors also demonstrate that decision weights don't sum to unity, or more specifically that $\omega(p) + \omega(q) < 1$ for $p + q = 1$. This conclusion is drawn from two pairs of gambles shows in Table (B.3). According to experiment, 82 out of 100 subjects preferred gamble A. Subject responses imply that:

$$\omega(0.66) \cdot v(2,400) + \omega(0.33) \cdot v(2,500) + \omega(0.01) \cdot v(0) < v(2,400), \quad \text{or}$$
$$(1 - \omega(0.66)) \cdot v(2,400) > \omega(0.33) \cdot v(2,500)$$

At the same time, gamble C was preferred by gamble D, implying that $\omega(0.33) \cdot v(2,500) > \omega(0.34) \cdot v(2,400)$. Putting the two inequalities, we get that $(1-\omega(0.66)) \cdot v(2,400) > \omega(0.34) \cdot v(2,400)$, or $1 > \omega(0.66) + \omega(0.34)$.

A:	2,500 with probability	0.33	B:	2,400 with probability	1
	2,400 with probability	0.66			
	0 with probability	0.01			

C:	2,500 with probability	0.33	D:	2,400 with probability	0.34
	0 with probability	0.67		0 with probability	0.66

Table B.3: The first and second pairs corresponds to problems 1 and 2 in [18]. Gamble A was chosen by 18 subjects, and gamble B was chosen by 82 subjects. Gamble C was chosen by 83 subjects, and gamble D was chosen by 17 subjects.

Bibliography

[1] Charles H Anderson, David C Van Essen, and Bruno A Olshausen. Directed visual attention and the dynamic control of information flow. In *Neurobiology of attention*, pages 11–17. Elsevier, 2005.

[2] Srikanth Bandi and Daniel Thalmann. Space discretization for efficient human navigation. In *Computer Graphics Forum*, volume 17, pages 195–206. Wiley Online Library, 1998.

[3] Colin F Camerer. An experimental test of several generalized utility theories. *Journal of Risk and uncertainty*, 2(1):61–104, 1989.

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

[5] Eric Eaton. Gridworld search and rescue: A project framework for a course in artificial intelligence. In *In the Proceedings of the AAAI-08 AI Education Colloquium, Chicago, IL*, 2008.

[6] Hillel J Einhorn and Robin M Hogarth. Decision making under ambiguity. *Journal of business*, pages S225–S250, 1986.

[7] Daniel Ellsberg. Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669, 1961.

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

[9] Therese Fessenden. Horizontal attention leans left. *Nielsen Normal Group*, 22, 2017.

[10] Shane Frederick. Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19(4):25–42, 2005.

[11] Peter Geibel and Fritz Wysotzki. Risk-sensitive reinforcement learning applied to control under constraints. *Journal of Artificial Intelligence Research*, 24:81–108, 2005.

[12] Samuel J Gershman, Eric J Horvitz, and Joshua B Tenenbaum. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–278, 2015.

[13] Nils H Hakansson. Essays in the theory of risk-bearing, 1972.

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

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

[16] Anthony Jameson. Choices and decisions of computer users. *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, 3, 2012.

[17] Daniel Kahneman and Amos Tversky. On the interpretation of intuitive probability: A reply to jonathan cohen. 1979.

[18] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. 47(2):263–292, 1979.

[19] Ramakrishna Kappagantu and S Arul Daniel. Challenges and issues of smart grid implementation: A case of indian scenario. *Journal of Electrical Systems and Information Technology*, 5(3):453–467, 2018.

[20] Ralph L Keeney, Howard Raiffa, and Richard F Meyer. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press, 1993.

[21] Hiroaki Kitano, Satoshi Tadokoro, Itsuki Noda, Hitoshi Matsubara, Tomoichi Takahashi, Atsuhi Shinjou, and Susumu Shimada. Robocup rescue: Search and rescue in large-scale disasters as a domain for autonomous agents research. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, volume 6, pages 739–743. IEEE, 1999.

[22] Marta Kryven, Tomer Ullman, Bill Cowan, and Joshua Tenenbaum. Plans or outcomes: How do we attribute intelligence to others? 2021.

[23] Falk Lieder, Thomas L Griffiths, and Ming Hsu. Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological review*, 125(1), 2018.

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

[25] R Duncan Luce. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.

[26] Mark J Machina. Choice under uncertainty: Problems solved and unsolved. *Journal of Economic Perspectives*, 1(1):121–154, 1987.

[27] Eleanor A Maguire, Neil Burgess, James G Donnett, Richard SJ Frackowiak, Christopher D Frith, and John O'Keefe. Knowing where and getting there: a human navigation network. *Science*, 280(5365):921–924, 1998.

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

[29] LA Prashanth, Cheng Jie, Michael Fu, Steve Marcus, and Csaba Szepesvári. Cumulative prospect theory meets reinforcement learning: Prediction and control. In *International Conference on Machine Learning*, pages 1406–1415. PMLR, 2016.

[30] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, et al. Machine behaviour. *Nature*, 568(7753):477–486, 2019.

[31] Lillian J Ratliff and Eric Mazumdar. Inverse risk-sensitive reinforcement learning. *IEEE Transactions on Automatic Control*, 65(3):1256–1263, 2019.

[32] Kai Ruggeri, Sonia Alí, Mari Louise Berge, Giulia Bertoldo, Ludvig D Bjørndal, Anna Cortijos-Bernabeu, Clair Davison, Emir Demić, Celia Esteban-Serna, and Maja Friedemann. Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, pages 1–12, 2020.

[33] Leonard J Savage. *The foundations of statistics*. Courier Corporation, 1972.

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

[35] Richard S Sutton and Andrew G Barto. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge, 1998.

[36] Edward C Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.

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

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

[39] Lihui Wang, Sichao Liu, Hongyi Liu, and Xi Vincent Wang. Overview of human-robot collaboration in manufacturing. In *Proceedings of 5th international conference on the industry 4.0 model for advanced manufacturing*, pages 15–58. Springer, 2020.

[40] Charley M Wu, Eric Schulz, Maarten Speekenbrink, Jonathan D Nelson, and Björn Meder. Generalization guides human exploration in vast decision spaces. *Nature human behaviour*, 2(12):915–924, 2018.

[41] Zhutian Yang. *Modeling Human Planning in Maze Orienteering Problems*. PhD thesis, Massachusetts Institute of Technology, 2021.

[42] Peng Yuan, Ana M Daugherty, and Naftali Raz. Turning bias in virtual spatial navigation: age-related differences and neuroanatomical correlates. *Biological psychology*, 96:8–19, 2014.