HEURISTIC SATISFICING INFERENTIAL DECISION MAKING IN HUMAN AND ROBOT ACTIVE PERCEPTION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yucheng Chen

August 2023

HEURISTIC SATISFICING INFERENTIAL DECISION MAKING IN HUMAN AND ROBOT ACTIVE PERCEPTION

Yucheng Chen, Ph.D.

Cornell University 2023

Inferential decision-making algorithms developed to date have assumed that an underlying probabilistic model of decision alternatives and outcomes may be learned *a priori* or online. As a result, when these assumptions are violated, they fail to provide solutions and are limited in their ability to modulate between optimizing and satisficing in the presence of hard time or cost constraints, adverse environmental conditions, or other unanticipated external pressures. Cognitive studies presented in this dissertation demonstrate that humans modulate between near-optimal and satisficing solutions, including heuristics, by leveraging information value of available environmental cues. Using the benchmark inferential decision problem known as a "treasure hunt", this dissertation develops a general approach for investigating and modeling active perception solutions under pressure, learning from humans how to modulate between optimal and heuristic solutions on the basis of external pressures and probabilistic models, if and when available. The result is an active perception approach that allows autonomous robots to modulate between near-optimal and heuristic strategies, tested via high-fidelity numerical simulations and physical experiments. The effectiveness of the new active perception strategies is demonstrated under a broad range of conditions, including decision tasks in which state-of-the-art sensor planning methods, such as cell decomposition, information roadmap, and information potential algorithms fail due to adverse weather (fog) or significantly underperform because of time or cost limitations.

BIOGRAPHICAL SKETCH

Yucheng Chen joined Sibley School of Mechanical and Aerospace Engineering at Cornell University in the fall of 2017. Prior to that, he obtained his bachelor's degree in Aeronautics and Astronautics and a minor in computer science at Purdue University. Since he started his PhD, the research work spanned from statistical analysis of human behavior data to hardware validation of the decision making strategies extracted from humans. Besides his own research, he also actively collaborated with other lab members, particularly on hardware side: he helped conduct a quadcopter experiment to support a dynamics learning algorithm, and designed another outdoor quadcopter experiment to validate a novel occlusion avoidance path planning algorithm.

To my family and my second half Yuqin Zhang.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Silvia Ferrari. Throughout my PhD study, Dr. Ferrari teaches me a systematic way of doing research, from how to propose a high quality research question, distinguish questions from methods to how to organize results to draw compelling conclusions. In the meantime, she also allows me to explore different aspects in my research work, including algorithm development, data analysis, and hardware validation. I would also like to thank all lab members in Laboratory of for Intelligent Systems and Controls (LISC), who give me the sense of belonging and many inspirations in research. Particularly, I would like to thank Dr. Pingping Zhu and Dr. Chang Liu, who already became assistant professors, for giving me a lot of help at the beginning stage of my PhD. I would also like to thank Shi Chang, who was my roommate for four years, for discussing research ideas and enjoying delicious food with me.

Finally, I would like to acknowledge and thank my parents for their unconditional love throughout the 10 years that I am in the US. I also want to express the deep appreciation to my love Yuqin Zhang, as she always shows significant patience and support especially in my difficult time. I would not have gone this far without the love I receive.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Rational inferential decision-making theories in both human and autonomous robot studies assume knowledge of a world model, such that near-optimal or satisficing strategies may be achieved by maximizing an appropriate utility function and/or satisficing mathematical constraints [80, 81, 13, 62, 82]. When a probabilistic world model is available, either because it is learned online or *a priori*, a variety of approaches, such as optimal control, robot/sensor planning, and maximum utility theories may be applied to inferential decision-making problems for robot active perception, planning, and feedback control [22, 51, 77, 89, 21, 47, 49]. Many "model-free" reinforcement learning (RL) and approximate dynamic programming (ADP) approaches have also been developed on the basis of the assumption that a partial or imperfect model is available in order to predict the next system state and/or "cost-to-go", and optimize the immediate and potential future rewards, such as information value [3, 79, 68, 20, 87, 92].

Humans have also been shown to use internal world models for inferential decision-making whenever possible, a characteristic first referred to as "substantial rationality" in [81, 80]. As also shown by the human studies on passive and active satisficing perception presented in this dissertation, given sufficient data, time, and informational resources, a globally rational human decision-maker uses an internal model of available alternatives, probabilities, and decision consequences to optimize both decision and information value in what is known as a "small-world" paradigm [76]. In contrast, in "large-world" scenarios, decision-makers face environmental pressures that prevent them from building an internal model or quantifying rewards, because of pressures such as missing data, time and computational

power constraints, or sensory deprivation, yet still manage to complete tasks by using "bounded rationality" [84]. Under these circumstances, optimization-based methods may not only be infeasible, returning no solution, but also cause disasters resulting from failing to take action [31]. Furthermore, Simon and other psychologists have shown that humans can overcome these limitations in real life via "satisficing decisions" that modulate between near-optimal strategies and the use of heuristics to gather new information and arrive at fast and "good-enough" solutions to complete relevant tasks.

To develop satisficing solutions for active robot perception, herein, we consider here the class of sensing problems known as treasure hunt [11, 93, 20, 12, 10, 9, 8]. The mathematical model of the problem, comprised of geometric and Bayesian network descriptions demonstrated in [11, 21], is used to develop a new experimental design approach that ensures humans and robots experience the same distribution of treasure hunts in any given class, including time, cost, and environmental pressures inducing satisficing strategies. This novel approach enables not only the readily comparison of the human-robot performance but also the generalization of the learned strategies to any treasure hunt problem and robotic platform. Hence, satisficing strategies are modeled using human decision data obtained from passive and active satisficing experiments, ranging from desktop to virtual reality human studies sampled from the treasure hunt model. Subsequently, the new strategies are demonstrated through both simulated and physical experiments involving robots under time and cost pressures, or subject to sensory deprivation (fog).

The treasure hunt problem under pressure, formulated in Chapter 2 and referred to as *satisficing treasure hunt* herein, is an extension of the robot treasure hunt presented in [11, 93], which introduces motion planning and inference in the

search for Spanish treasures originally used in [83] to investigate satisficing decisions in humans. Whereas the search for Spanish treasures amounts to searching a (static) decision tree with hidden variables, the robot treasure hunt involves a sensor-equipped robot searching for targets in an obstacle-populated workspace. As shown in [21] and references therein, the robot treasure hunt paradigm is useful in many mobile sensing applications involving multi-target detection and classification. In particular, the problem highlights the coupling of action decisions that change the physical state of the robot (or decision-maker) with test decisions that allow the robot to gather information from the targets via onboard sensors. In this dissertation, the satisficing treasure hunt is introduced to investigate and model human satisficing perception strategies under external pressures in passive and active tasks, first via desktop simulations and then in the Duke immersive Virtual Environment (DiVE) [18], as shown in Fig. 1.1.

To date, substantial research has been devoted to solving treasure hunt problems for many robots/sensor types, in applications as diverse as demining infrared sensors and underwater acoustics, under the aforementioned "small-world" assumptions [21]. Optimal control and computational geometry solution approaches, such as cell decomposition [11], disjunctive programming [88], and information roadmap methods (IRM) [93], have been developed for optimizing robot performance by minimizing the cost of traveling through the workspace and processing sensor measurements, while maximizing the sensor rewards such as information gain. All these existing methods assume prior knowledge of sensor performance and of the workspace, and are applicable when the time and energy allotted to the robot are adequate for completing the sensing task. Information-driven path planning algorithm integrated with online mapping, developed in [94, 55, 25], have extended former treasure hunt solutions to problems in which a prior model of the

(a) (b)

Figure 1.1: Human participant solving treasure hunt problem first under no pressure (a) and then under sensory deprivation (fog pressure) (b) in the DiVE [18].

workspace is not available and must be obtained online. Optimization-based algorithms have also been developed for fixed end-time problems with partial knowledge of the workspace, on the basis of the assumption that a probabilistic model of the information states and unlimited sensor measurements are available [74]. This dissertation builds on this previous work to develop heuristic strategies applicable when uncertainties cannot be learned or mathematically modeled in closed form, and the presence of external pressures might prevent task completion, e.g., adverse weather or insufficient time/energy.

Inspired by previous findings on human satisficing heuristic strategies [31, 28, 32, 29, 63], this dissertation develops, implements, and compares the performance

between existing treasure hunt algorithms and human participants engaged in the same sensing tasks and experimental conditions by using a new design approach. Subsequently, human strategies and heuristics outperforming existing state-of-the-art algorithms are identified and modeled from data in a manner that can be extended to any sensor-equipped autonomous robot. The effectiveness of these strategies is then demonstrated with camera-equipped robots via high-fidelity simulations as well as physical laboratory experiments. In particular, human heuristics are modeled by using the "three building blocks" structure for formalizing general inferential heuristic strategies presented in [33]. The mathematical properties of heuristics characterized by this approach are then compared with logic and statistics, according to the rationale in [31].

Three main classes of human heuristics for inferential decisions exist: recognition-based decision-making [71, 35], one-reason decision-making [60, 29], and trade-off heuristics[53]. Although categorized by respective decision mechanisms, these classes of human heuristics have been investigated in disparate satisficing settings, thus complicating the determination of which strategies are best equipped to handle different environmental pressures. Furthermore, existing human studies are typically confined to desktop simulations and do not account for action decisions pertaining to physical motion and path planning in complex workspaces. Therefore, this dissertation presents a new experimental design approach (Chapter 3) and tests in human participants to analyze and model satisficing active perception strategies (Chapter 6) that are generalizable and applicable to robot applications, as shown in Chapter 7.

The dissertation presents separate analysis and modeling studies for human satisficing strategies in passive and active perception and decision-making tasks.

For passive tasks, time pressure on inference is introduced to examine subsequent effects on human decision-making in terms of decision model complexity and information gain. The resulting heuristic strategies (Chapter 4) extracted from human data demonstrate adaptability to varying time pressure, thus enabling inferential decision-making to meet decision deadlines. These heuristics significantly reduce the complexity of target feature search from an exhaustive search $O(2^n)$ to $O(nlog(n)+n)$, where n is the number of target features. Additionally, the heuristics exhibit superior classification performance to that of an "optimal" strategy using all target features for inference (Chapter 5), thus reflecting the less-can-be-more effect [31]. For active tasks, human motion strategies are modeled as heuristics (Chapter 6) when information gathering capabilities are limited, such as in adverse weather conditions. These proposed heuristic strategies are then applied and tested on robots equipped with onboard sensors, and compared with existing planning methods (Chapter 7) through simulations and physical experiments in which the optimal strategies have very poor performance. Regarding information cost pressure, a decision-making strategy using a mixed integer nonlinear program (MINLP) is developed on the basis of existing methods [11, 93]. The MINLP-based strategy, compared with human strategies, demonstrates superior performance (Chapter 7). This finding is consistent with expectations, given that information cost pressure does not fundamentally undermine the accuracy of the presumed world model, and optimization-based solutions can still provide high-quality decisions. Overall, the strategies presented herein provide a toolbox for decision-making under different pressures to solve the satisficing treasure hunt problem.

CHAPTER 2

TREASURE HUNT PROBLEM FORMULATION

This dissertation considers the treasure hunt problem of inferential decision making under pressure for the purpose of mobile information gathering from multiple r fixed targets. The goal of an information-gathering agent is to navigate in an obstacle-populated environment to observe the target features and infer their classification variables, and ultimately, discover all important targets or "treasures" in a workspace referred to as $\mathcal{W} \subset \mathbb{R}^2$. The body of the information-gathering agent is assumed to be a rigid geometry $\mathcal{I} \subset \mathcal{W}$.

The workspace is populated with q fixed rigid obstacles $\mathcal{B}_1, ..., \mathcal{B}_q \subset \mathcal{W}$ and r fixed point targets $\mathbf{x}_1, ..., \mathbf{x}_r \in \mathcal{W}$. Associated with the ith target, a set of n random variables represents target features $F_i = \{F_{i,1} \ ... \ F_{i,l} \ ... \ F_{i,n}\}$. The lth feature is discrete and random with finite range. A random and discrete hypothesis variable Y_i represents the ith target's property of interest. The range is $\mathcal{Y} = \{y_j | j \in \mathcal{J}\}$, where y_j represents the jth outcome of Y_i, and both F_i and Y_i are assumed unknown in prior.

The information-gathering agent can perceive information about targets and obstacles in its surroundings. This ability is facilitated by the concept of field of view (FOV), which is defined as follows:

Definition 2.0.1 (FOV) *The field of view of an information-gathering agent in workspace $\mathcal{W}$ is a compact subset $\mathcal{S} \subset \mathcal{W}$, from which the information-gathering agent can gather information on the environment.*

FOVs can be associated with different perception mechanisms, such as passive perception, wherein the agent collects information without actively interacting

with the environment, or interactive perception, whose outcome may depend on interactions with the environment.

Thus, the considered information-gathering agent is assumed to have two distinct FOVs: $\mathcal{S}_P$ for passive perception and $\mathcal{S}_I$ for interactive perception. Both FOVs are not omnidirectional. The agent relies on $\mathcal{S}_P$ to gather navigational information such as obstacle detection and target awareness. In contrast, $\mathcal{S}_I$ is utilized for observing inference-related information, such as target features, and obtaining information value.

The state of an information-gathering agent at t_k with the aforementioned perception capabilities is described as a vector $\mathbf{q}_k = [\mathbf{s}_k^T \quad \theta_k \quad \xi_k \quad \phi_k]^T$, where $\mathbf{s}_k = [x_k \quad y_k] \in \mathcal{W}$ is the position of the information-gathering agent with respect to the workspace $\mathcal{W}$, $\theta_k \in \mathbb{S}^1$ is the orientation of the agent, and $\xi_k \in [\xi_l, \xi_u]$ and $\phi_k \in [\phi_l, \phi_u]$ are preferred information gathering directions of the "passive" and "interactive" FOVs, respectively. In addition, ξ_l, ξ_u and ϕ_l, ϕ_u bound the preferred information gathering directions for $\mathcal{S}_P$ and $\mathcal{S}_I$ with respect to the information-gathering agent body. This definition enables the preferred information gathering directions of both FOVs to not be fixed to the orientation of the agent itself. The corresponding configuration of the information-gathering agent body is defined as $\mathbf{t}_k = [\mathbf{s}_k^T \quad \theta_k]^T$. Let $\mathcal{C}$ represent all possible configurations of the information-gathering agent. The space of configurations that cause collisions between the agent and $\mathcal{B}_j$ is defined as C-obstacles $\mathcal{CB}_j = \{\mathbf{t} \in \mathcal{C} | \mathcal{I}(\mathbf{t}) \cap \mathcal{B}_j \neq \emptyset\}$. Then, the agent must travel in free configuration space $\mathcal{C}_{\text{free}} = \{\mathcal{C} \setminus \bigcup_{j=1}^{q} \mathcal{CB}_j\}$ [21]. While an agent is in free configuration space and a point of interest falls within the FOV. The information gathering process should also comply with line of sight visibility constraints, as in Definition 2.0.2.

Definition 2.0.2 (Line of sight) *An opaque object $\mathcal{B} \subset \mathcal{W}$ blocks the line of sight between a point of interest at $\mathbf{x} \in \mathcal{W}$ and an information-gathering agent at $\mathbf{s} \in \mathcal{W}$ if and only if,*

$$P(\mathbf{s}, \mathbf{x}) \cap \mathcal{B} \neq \emptyset \tag{2.1}$$

where $P(\mathbf{s}, \mathbf{x}) = \{(1-\gamma)\mathbf{s} + \gamma\mathbf{x} | \gamma \in [0, 1]\}$

Therefore, an information-gathering agent at $\mathbf{s} \in \mathcal{W}$ can capture the environmental information of a point of interest $\mathbf{x} \in \mathcal{W}$ if $\mathbf{x} \in \mathcal{S}_P(\mathbf{q})$ and $P(\mathbf{s}, \mathbf{x}) \cap \mathcal{B}_j = \emptyset,\ 1 \leq j \leq q$.

According to visibility theory developed in [26], the region that enables visibility of target at $\mathbf{x}_i$ with respect to $\mathcal{S}_P$, is defined in Definition 2.0.3.

Definition 2.0.3 (Target Visibility Region) *Given an information-gathering agent at $\mathbf{s} \in \mathcal{W}$, and FOV geometry $\mathcal{S}_P$, the visibility region of a target at $\mathbf{x}_i \in \mathcal{W}$ in the presence of q opaque obstacles $\mathcal{B}_j (j = 1, ..., q)$ is defined as the subset of $\mathcal{C}_{free}$ that simultaneously satisfies the FOV and LOS target visibility conditions,*

$$\mathcal{TV}_i = \{\mathbf{t} \in \mathcal{C}_{free} | \mathbf{x}_i \in \mathcal{S}_P, P(\mathbf{s}, \mathbf{x}_i) \cap \mathcal{B}_j = \emptyset, \forall j\} \tag{2.2}$$

where $P(\mathbf{s}, \mathbf{x}) = \{(1-\gamma)\mathbf{s} + \gamma\mathbf{x} | \gamma \in [0, 1]\}$

For multiple target visibility regions that are not mutually disjoint, every region of intersection can be associated with an index set that represents the indices of all targets visible from the corresponding set of configurations. Then, a one-to-one

correspondence can be established between target sets and visibility by introducing Definition 2.0.4.

Definition 2.0.4 (Set Visibility Region) *Given a set of target r visibility regions $\{\mathcal{TV}_i | i \in \{1, 2, ..., r\}\}$, let $S \subseteq \{1, 2, ..., r\}$ represent the set of target indices of two or more intersecting regions, $\bigcap_{i \in S} \mathcal{TV}_i \neq \emptyset$. Then, the set visibility region is defined as*

$$\mathcal{V}_S = \{\bigcap_{i \in S} \mathcal{TV}_i | S \subseteq \{1, 2, ..., r\}\} \tag{2.3}$$

If the ith point target satisfies $\mathbf{x}_i \in \mathcal{S}_I(\mathbf{q})$ and $P(\mathbf{s}, \mathbf{x}_i) \cap \mathcal{B}_j = \emptyset,\ 1 \leq j \leq q$, the agent is able to observe the target features and obtain a set of measurements, $M_i = \{m_{i,l} | 1 \leq l \leq \wp, l \in \mathbb{Z}\}$, for target features in F_i, where $\wp$ is the number of target feature measurements. The test variable $m_{i,l}$ is random and discrete, with the same finite range as $F_{i,l}$, and typically includes superimposed random noise that may induce measurement error, however, the outcome of a test variable always falls within the range of the corresponding target feature. Thus, the agent's goal is to identify the treasure(s) in the environment by inferring the hypothesis variable Y_i from M_i, by using a measurement model $P(Y_i, M_i)$, which is the casual relationship between measurements and classification variables [21]. The model could be, for example, a Bayesian network from expert knowledge or prior training data, and a mapping between measurements to classification variables in human memory from learning experience.

The target feature measurements are made through test decisions by an information-gathering agent. A test decision is a decision to look for more evidence to be entered into $P(Y, M)$ [41]. Define $u(t_k) \in \mathcal{U}_k$ as a test decision chosen

from the set $\mathcal{U}_k \subset \mathcal{U}$ of all admissible tests at t_k. The set $\mathcal{U} = \{\vartheta_c, \vartheta_s, \vartheta_{un}\}$ consists of all test decisions, where ϑ_c and ϑ_s represent the decisions whether to continue or stop measuring target features, and ϑ_{un} refers to not performing any tests on a target. The outcomes of the test decision $u(t_k)$ are a measurement variable $z(t_{k+1}) = m_{i,l}, 1 \leq i \leq r, 1 \leq l \leq n$, and information cost $J(t_k) \in \mathbb{Z}$, which is modeled as cumulative target feature measurement up to t_k. If a measurement budget J_b exists, then the cumulative information cost at final time t_T should not exceed J_b.

An action decision is a decision to change the state of the world and the decisions for an information-gathering agent [41]. Concretely, this decision determines the path of the information-gathering agent, the orientation of the agent, and the preferred information gathering directions of $\mathcal{S}_P$ and $\mathcal{S}_I$, as it navigates through the environment to observe the targets. Define $a(t_k) \in \mathcal{A}_k$ as an action decision chosen at time t_k from set $\mathcal{A}_k$ of all admissible actions. The agent motion can then be described by a causal model as the following difference equation,

$$\mathbf{q}_{k+1} = \mathbf{f}[\mathbf{q}_k, a(t_k), t_k] \tag{2.4}$$

where $\mathbf{f}[\cdot]$ is the known transition dynamics of an information-gathering agent with respect to time.

An active inferential strategy as a sequence of decision functions is then described as follows:

Definition 2.0.5 (Strategy) *An active inferential strategy is a class of admissi-*

ble policies that consists of a sequence of functions,

$$\sigma = \{\pi_0, \pi_1, ..., \pi_T\} \tag{2.5}$$

where π_k maps all past information-gathering agent states, test variables, action and test decisions into admissible action and test decisions,

$$\begin{aligned} \{a(t_k), u(t_k)\} = \pi_k[&\mathbf{q}_0, a(t_1), u(t_1), z(t_1), J(t_1), \mathbf{q}_1, \\ &..., a(t_{k-1}), u(t_{k-1}), z(t_{k-1}), J(t_{k-1}), \mathbf{q}_{k-1}] \end{aligned} \tag{2.6}$$

such that $\pi_k[\cdot] \in \{\mathcal{A}_k, \mathcal{U}_k\}$, for all $k = 1, 2, ..., T$.

Based on all the aforementioned definitions, the problem is formulated as follows:

Problem 1 (Satisficing Treasure Hunt) *Given the initial state* $\mathbf{q}_0$*, the satisficing objective of the problem is to develop a strategy within finite time horizon* $(0, T]$*, such that,*

$$\sum_{i=1}^{r} [\mathbb{1}(\exists k, \mathbf{x}_i \in \mathcal{S}_I(\mathbf{q}_k) \wedge P(\mathbf{s}_k, \mathbf{x}_i) \cap \mathcal{B}_j, \forall j) I(Y_i; M_i)] \geq \alpha$$

where

$$\mathbf{q}_{k+1} = \mathbf{f}[\mathbf{q}_k, a(t_k), t_k]$$

$$M_i = \{m_{i,l} | 1 \leq l \leq \wp_i, l \in \mathbb{Z}\}$$

$$\hat{y}_i = \underset{y \in \mathcal{Y}}{\operatorname{argmax}} P(Y_i = y, M_i)$$

$$I(Y_i; M_i) = H(Y_i) - H(Y_i | M_i)$$

$$J(t_T) \leq J_b$$

$$i = 1, 2, ..., r, \; 1 \leq k \leq T$$

$$j = 1, 2, ..., q$$

The satisficing search objective requires the summation of the information value of all visited targets to be no less than an aspiration level α. A feasible search strategy may use all or part of the available models of the environment and targets, or knowledge of prior states and decisions to produce a sequence of action and test decisions that satisfy the objective within t_T.

CHAPTER 3

HUMAN SATISFICING STUDIES

In order to investigate human satisficing strategies in the treasure hunt problem, this dissertation presents two classes of experiments: passive and active satisficing. Both classes share the same underlying mathematical formulation known as "treasure hunt" (Chapter 2) and serve as a behavioral paradigm for humans and autonomous robots. Passive satisficing experiments focus on treasure hunt problems in which information is presented to the decision maker who passively observes cues needed to make inferential decisions. Active satisficing experiments allow the decision maker to control the amount of information gathered in support of inferential decisions. Another important distinction is between static and dynamic treasure hunts in which the decision maker remains stationary or moves through the environment, respectively, in order to find the hidden treasures.

Previous studies showed that the urgency to respond [16] and the need for fast decision-making significantly affect human decision evidence accumulation, thus leading to the use of heuristics in solving such tasks [64]. Passive satisficing experiments focus on test decisions, which determine the evidence accumulation of the agent based on partial information under "urgency". Inspired by satisficing searches for Spanish treasures with feature ordering constraints [83], active satisficing includes both test and action decisions, which change not only the agent's knowledge and information about the world but also its physical state. In humans, knowledge of the world/agent is acquired through the six senses, i.e. assumed to be vision in this dissertation, whereas in robots it is acquired through onboard sensors, i.e. cameras. Because information gathering by a physical agent such as a human or robot is a causal process [21], feature ordering constraints are necessary

in order to describe the temporal nature of information discovery.

Both passive and active satisficing human experiments comprise a training phase and a test phase that are also similarly applied in the robot experiments in Chapters 5-7. During the training phase, human participants learn the validity of target features in determining the outcome of the hypothesis variable. They receive feedback on their inferential decisions to aid in their learning process. During the test phase, pressures are introduced, and action decisions are added for active tasks. Importantly, during the test phase, no performance feedback or ground truth is provided to human participants (or robots).

3.1 Passive Satisficing Task

The passive satisficing experiments presented in this dissertation adopted the passive treasure hunt problem, shown in Fig.3.1 and related to the well-known weather prediction task [34, 45, 86]. The problem was first proposed in [63] to investigate the cognitive processes involved in human test decisions under pressure. In view of its passive nature, the experimental platform of choice consisted of a desktop computer used to emulate the high-paced decision scenarios, and to encourage the human participants to focus on cue combination rather than memorization [63, 46].

The stimuli presented on a screen were precisely controlled, ensuring consistency across participants and minimizing distractions from irrelevant objects or external factors [24, 50]. In each task, participants were presented with two different stimuli from which to select the "treasure" before the total time, t_T, at one's disposal has elapsed (time pressure). The treasures are hidden but correlated with the visual appearance of the stimulus, and the underlying probabilities must be learned by trial and error during the training phase. Each stimulus is

characterized by four binary cues or "features", namely color (F_1), shape (F_2), contour (F_3), and line orientation (F_4), illustrated in the table in Fig. 3.1. The goal of this passive satisficing task is to find all treasures among stimuli that are presented on the screen or, in other words, to infer a binary hypothesis variable Y, with range $\mathcal{Y} = \{y_1, y_2\}$, where y_1 = "treasure" and y_2 = "not treasure". The task is passive by design because the participant cannot control the information displayed in order to aid his/her decisions.

During the training phase, each (human) participant performed 240 trials to learn the relationship between cues, $F = \{F_1, F_2, F_3, F_4\}$, and the hypothesis variable Y. After the training phase, participants were divided into two groups. The first group underwent a moderate time pressure (TP) experiment and was tested against two datasets, each consisting of 120 trials. Participants were required to make decisions within a response time $t_T = 750$ ms, which allowed ample time to ponder on the cues presented and how they related to the treasure. The second group underwent an intense TP experiment, with a response time of only $t_T = 500$ ms. Participants in this group also encountered two datasets, each containing 120 trials. A more detailed description of the experiment, including redundant cues and human subject procedures, can be found in [63]. Subsequently, the task was modified to develop a number of active satisficing treasure hunts in which information about the treasures had to be obtained by navigating a complex environment, as explained in the next section.

Feature dimension	Stimulus	Feature state 1	2
Color		Blue	Red
Shape		Circle	Square
Contour		White	Black
Line Orientation		Horizontal	Vertical

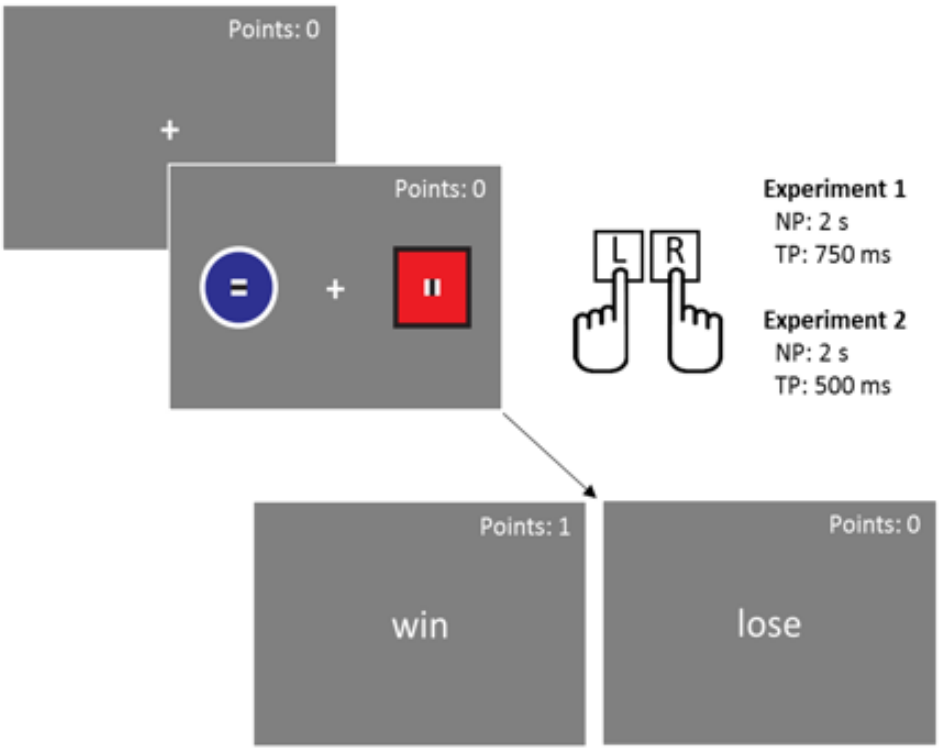
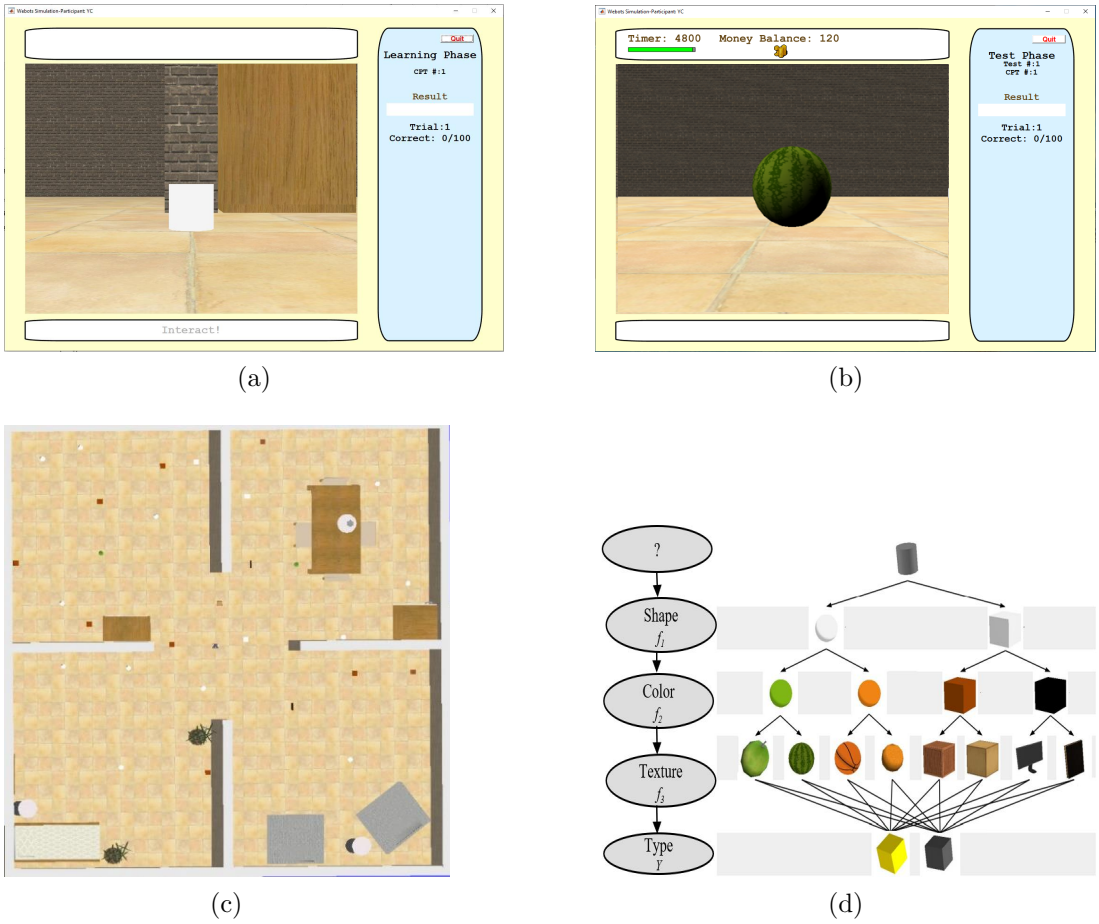


Figure 3.1: Cues and human display used for the passive satisficing experiment, where the result of "win" or "lose" was displayed only during the training phase.

3.2 Active Satisficing Treasure Hunt Task

The satisficing treasure hunt task is an ambulatory study in which participants must navigate a complex environment populated with a number of obstacles and objects in order to first find a set of targets (stimuli) and, then, determine which are the treasures. Additionally, once the targets are inside the participant's FOV, cues are displayed sequentially to him/her only after paying a price for the information requested. The ordering constraints (illustrated in Fig. 3.2d) allow for the study of information cost and its role in the decision making process by which the task is to be performed not only under time pressure but also a fixed budget. Thus, the satisficing treasure hunt allows not only to investigate how information about a hidden variable (treasure) is leveraged, but also how humans mediate between multiple objectives such as obstacle avoidance, limited sensing resources, and time constraints. Participants must, therefore, search and locate the treasures without any prior information on initial target features, target positions, or workspace and obstacle layout.

In order to utilize a controlled environment that can be easily changed to study all combinations of cues, target/obstacle distributions, and underlying probabilities, the active satisficing treasure hunt task was developed and conducted in a virtual reality environment known as the DiVE [18]. By this approach different experiments were designed and easily modified so as to investigate different difficulty levels and provide the human participants repeatable, well-controlled, and immersive experience of acquiring and processing information to generate behavior [90, 65, 78]. The DiVE consists of a 3m x 3m x 3m stereoscopic rear projected room with head and hand tracking, allowing participants to interact with a virtual environment in real-time [18]. By developing a new interface between the DiVE and



(a) (b)

(c) (d)

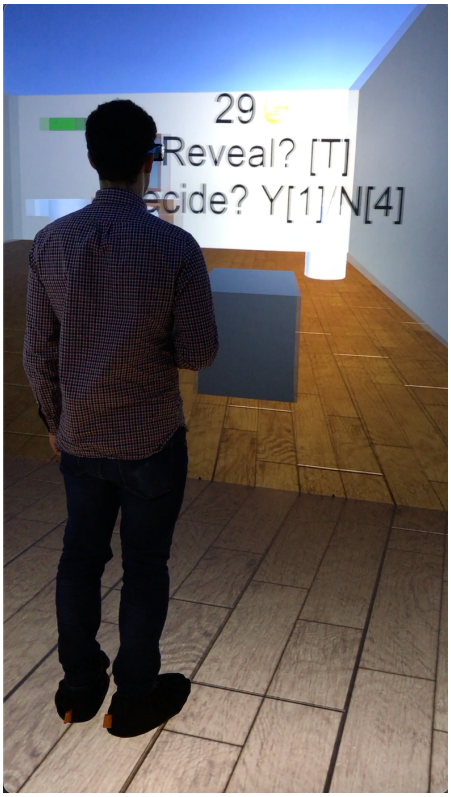
Figure 3.2: First-person view in training phase without prior target feature revealed (a) and with feature revealed by a participant (b) in the Webots® workspace (c) and target cues encoded in a BN structure with ordering constraints (d).

the robotic software Webots®, this research was able to readily introduce humans within the same environments designed for humans, and vice versa, according to the Bayesian network (BN) model of the desired treasure hunt task. The structure of the BN model capturing the relationship between the target variables in this treasure hunt perception task is plotted in Fig. 3.2d. The BN parameters, not shown for brevity, were varied across trials to obtain a representative dataset from the human study from which mathematical models of human decision strategies could be learned and validated.

Six human participants were trained and given access to the DiVE for a total of fifty-four trials with the objective to model aspects of human intelligence that outperform existing robot strategies. The number of trials and participants is adequate to the scope of the study which was not to learn from a representative sample of the human population, but to extract inferential decision making strategies generalizable to treasure hunt robot problems. Besides manageable in view of the high costs and logistical challenges associated with running DiVE experiments, the size of the resulting dataset was also found to be adequate to varying all of the workspace and target characteristics across experiments, similarly to the studies in [95, 52]. Moreover, through the VR googles and environment, it was possible to have precise and controllable ground truth not only about the workspace, but also about the human FOV, $\mathcal{S}_P$, within which the human could observe critical information such as targets, cues, and obstacles.

A mental model of the relationship between target features and classification was first learned by the human participants during 100 stationary training sessions (Fig. 3.2a and Fig. 3.2b) in which the target features (visual cues), comprised of shape (F_1), color (F_2), and texture (F_3), followed by the target classification Y, where $\mathcal{Y} = \{y_1, y_2\}$, were displayed on a computer screen, through the Webots® simulation shown in Fig. 3.2. Participants were then instructed to search for treasures inside an unknown 10m x 10m Webots® workspace with $r = 30$ targets (Fig. 3.2c), by paying a fee to see the features, $F = \{F_{i,1}, F_{i,2}, F_{i,3}\}$, of every target i inside their FOV sequentially over time (test phase). Based on the features observed, which may have included one or more features in the set F, participants then had to decide which targets were treasures ($Y = y_1$) or not ($Y = y_2$). No feedback about their decisions was provided and, as explained in Chapter 2, the task had to be performed within a limited budget J_b and time period t_T.

(a)



(b)

Figure 3.3: Test phase in active satisficing experiment in DiVE.

Mobility and ordering cue constraints are critical to causal systems, such as autonomous sensors and robots, because they are intrinsic to how these physical systems are able to gather information from the environment as they move and/or interact with the world around them. Thanks to the simulation environments and human experiment design presented in this section, we were able to engage participants in a series of classification tasks in which target features were revealed only after paying both a monetary and time cost, similarly to artificial sensors that require both computing and time resources to process visual data. Participants were able to build a mental model built for decision making with the inclusion of temporal constraints during the training phase, according to the BN conditional probabilities (parameters) of each study. By sampling the Webots® environments from each BN model, selected by the experiment designer to encompass the full range of inference problem difficulty, and by transferring them automatically into VR (Fig. 3.3) the data collected was guaranteed ideally suited for the modeling and generalization of human strategies to robots (Chapter 6). As explained in the next section, the test phase was conducted under three pressure conditions: no pressure, money pressure, and sensory deprivation (fog).

3.3 External Pressures Inducing Satisficing

Previous work on human satisficing strategies and heuristics illustrated that most humans resort to these approaches for two main reasons, one is computational feasibility and the other is the "less-can-be-more" effect [31]. When the search for information and computation costs become impractical for making a truly "rational" decision, satisficing strategies adaptively drop information sources or partially explore decision tree branches, thus accommodating the limitations of computa-

tional capacity. In situations in which models have significant deviations from the ground truth, external uncertainties are substantial, or closed-form mathematical descriptions are lacking, optimization on potentially inaccurate models can be risky. As a result, satisficing strategies and heuristics often outperform classical models by utilizing less information. This effect can be explained in two ways. Firstly, the success of heuristics is often dependent on the environment. For example, empirical evidence suggests that strategies such as "take-the-best," which rely on a single good reason, perform better than classical approaches under high uncertainty [39]. Secondly, decision-making systems should consider trade-offs between bias and variance, which is determined by model complexity[4]. Simple heuristics with fewer free parameters have smaller variance than complex statistical models, thus avoiding overfitting to noisy or unrepresentative data, and generalizable across a wider range of datasets [4, 6, 30].

Motivated by the situations where robots' mission goals can be severely hindered or completely compromised due to inaccurate environment or sensing models caused by pressures, the dissertation seeks to emulate aspects of human intelligence under the pressures and study their influence on decisions. The environment pressures include, for example, time pressure [66], information cost [17, 7], cue redundancy [17, 73], sensory deprivation, and high risks [85, 67]. Cue redundancy and high risk have been investigated extensively in statistics and economics, particularly in the context of inferential decisions [44, 59]. In the treasure hunt problem, sensory deprivation and information cost directly and indirectly influence action decisions, which brings insight how these pressures impact agents' motion. However, the effects of sensory deprivation on human decisions have not been thoroughly investigated compared to other pressures. Time pressure is ubiquitous in the real world, yet heuristic strategies derived from human behavior are still lack-

ing. Thus, this dissertation aims to fill this research gap by examining the time pressure, information cost pressure, and sensory deprivation and their effects on decision outcomes.

3.3.1 Time Pressure

Assume that a fixed time interval t_c is needed to integrate one additional cue into the inference decision-making process. In the meantime, each decision must be made within t_b, and $\wp_i$ is the number of measurements for the ith target. The satisficing strategies must adaptively select a subset of the cues such that a decision is made within the time constraint

$$\wp_i t_c < t_b, i = 1, 2, ..., r \tag{3.1}$$

According to the human studies in [63], the response time of participants in the passive satisficing tasks was measured during the pilot work. The average response time in these tasks was found to be approximately 700 ms. Based on this finding, three time windows were designed to represent different time pressure levels: a two-second time window was considered without any time pressure; a 750 ms time window was considered moderate time pressure; and a 500 ms time window was considered intense time pressure.

3.3.2 Information Cost Pressure

The cost of acquiring new information intrinsically makes an agent use fewer cues to reach a decision. In Chapter 2, new information for the ith target is collected through a sequence of $\wp_i$ measurements of target features. Thus, for all r targets, the information cost is mathematically described as the total number of target feature measurements not exceeding a preset budget J_b

$$\sum_{i=1}^{r} \wp_i \leq J_b \tag{3.2}$$

In Chapter 3, the human studies introduce information cost pressure using the parameter $J_b = 30$. In the context of the treasure hunt problem, J_b represents the measurement budget, which limits the number of features that a participant can measure for the targets. In this specific experiment, there are a total of $r = 30$ targets, and each target has multiple features that can be measured. The value of $J_b = 30$ means that the human participants are constrained to measure, on average, one feature per target. This budget is not sufficient to measure all available features for all the targets but allows the participants to gather limited information about each target.

3.3.3 Sensory Deprivation Pressure

Many robot applications face sensory deprivation due to environmental conditions, such as occlusion of sight by fog, clouds, or other adverse weather, as well as unexpected sensor damages or interference. This dissertation considers a scenario in which the environment contains steady, dense, and almost uniform particles

(e.g., fog or clouds) that strictly limit the effective distance of vision perception capabilities. Consider the assumption in Chapter 2 that the map is not known *a priori*. The information on the target and obstacle positions and geometries relies on information from FOV $\mathcal{S}_P$ shown in Fig. 3.4. Consider the concept of set visibility region (Definition 2.0.4). Let $S \subseteq \{1, 2, ..., r\}$ represent a subset of target indices, and the set visibility region $\mathcal{V}_S \subseteq \mathcal{C}_{\text{free}}$ enables the visibility to all the targets in S with respect to $\mathcal{S}_P$. A globally optimal solution to treasure hunt problem with respect to a subset of target S is feasible only if $\mathcal{V}_S \neq \emptyset$. This means that the agent has full visibility to all targets in the set S, allowing the agent to start with complete information and compute optimal action and test decisions. Unfortunately, under sensory deprivation, the perception capability for $\mathcal{S}_P$ is severely restricted. As in human studies in Section 3.2, the sensory deprivation is introduced by constraining the effective distance of vision to 1m in an environment of the size 20m by 20m. This situation often leads $V_S = \emptyset$ even for a set $|S| = 2$ (i.e., a set with two targets), which indicates that an globally optimal solution of the 2 targets is infeasible. Consequently, long-horizon optimal planning becomes extremely challenging due to lack of target information under sensor deprivation. In such situations, satisficing strategies are aimed at overcoming this difficulty, and use local information to explore the environment and visit targets. A fog environment is visually represented in Webots®, as shown in Fig. 3.4. The figure illustrates the camera FOV as an example of $\mathcal{S}_P$, and the measurement FOV as an example of $\mathcal{S}_I$.

(a)

(b)

Figure 3.4: Top view visibility conditions of the workspace (a) and first-person view visibility condition (b) under fog pressure

CHAPTER 4

MATHEMATICAL MODELING OF HUMAN PASSIVE SATISFICING STRATEGIES

Assume the probabilistic sensor model is a known a-priori under this pressure condition. The probabilistic classification tasks, as described in the author and the collaborators' previous work [63], is mathematically described as follows.

Two objects on the left(L) and right(R) are displayed on the screen simultaneously, each object has four cues (color, shape, contour, line orientation). For every binary cue, each cue state associates with a weight, where $w_{i,j}$ denotes the weight of the ith cue in state j, $i = 1, 2, 3, 4,\ j = 1, 2$. For example, denote the target feature/cue sets of the left and right objects as $F_L = \{F_{1,L}, F_{2,L}, F_{3,L}, F_{4,L}\}$ and $F_R = \{F_{1,R}, F_{2,R}, F_{3,R}, F_{4,R}\}$, respectively. and the associated weights for the cues of left and right objects are denoted as the sets $W_L = \{w_{1,L}, w_{2,L}, w_{3,L}, w_{4,L}\}$, and $W_R = \{w_{1,R}, w_{2,R}, w_{3,R}, w_{4,R}\}$. The outcomes of the hypothesis variable are denoted as $\mathcal{Y} = \{y_L, y_R\}$, where y_L denotes the left object "wins", while y_R denotes the right object "wins". The difference in the cue weights determines the probability of wining, as shown in Eq. 4.1.

$$p(y_L|F_L, F_R) = \frac{10^{\sum_{i=1}^{4}(w_{i,L}-w_{i,R})}}{1 + 10^{\sum_{i=1}^{4}(w_{i,L}-w_{i,R})}} \tag{4.1}$$
$$p(y_R|F_L, F_R) = 1 - p(y_L|F_L, F_R)$$

Denote the set of cue indices that are used for inference is M, which is an element of the power set of all cue indices and M is not an empty set, as represented in Eq. 4.2,

$$M \in \mathcal{P}(\{1,2,3,4\}) \backslash \emptyset \tag{4.2}$$

Given the set M, consider M is a set of binary state random variables, and denote $\hat{m}$ as an instantiation of M, the probability for the left object to "win" presented in Eq. 4.1 is modified as,

$$p(y_L|F_L, F_R, \hat{m}) = \frac{10^{\sum_{i\in\hat{m}}(w_{i,L}-w_{i,R})}}{1+10^{\sum_{i\in\hat{m}}(w_{i,L}-w_{i,R})}} \tag{4.3}$$

Each set M corresponds to one of fifteen decision models as shown in Fig. 4.1c that uses a particular subset of cues. Then the information gain is encoded as the expected entropy reduction. If no cue is used, then the probability for the left object to "win" is,

$$p(y_L|F_L, F_R, \emptyset) = \frac{10^0}{1+10^0} = \frac{1}{2}$$

Therefore, the initial entropy is,

$$H(Y) = \mathbb{E}[-\log p(y_L|F_L, F_R, \emptyset)]$$

The information gain is represented as,

$$\begin{aligned} \Delta H(Y, M) &= H(Y) - H(Y|M) \\ &= H(Y) - \sum H(Y|\hat{m})p(\hat{m}) \end{aligned} \tag{4.4}$$

where $H(Y|\hat{m}) = \mathbb{E}[-\log p(y_L|F_L, F_R, \hat{m})]$.

4.1 Human Data Analysis

Previous work by the authors in [63] showed that human participants drop less informative cues to meet pressing time deadlines that do not allow them to complete the tasks optimally. The analysis of data obtained from the moderate TP experiment (Fig. 4.1a) and intense TP experiment (Fig. 4.1b) reveals similar interesting findings regarding human decision-making under different time pressure conditions. Under the no TP condition, the most probable decision model selected by human participants (indicated by yellow boxes in Fig. 4.1a and Fig. 4.1b) utilizes all four cues and aims at maximizing information gain. However, under moderate TP, the most probable decision model selected by human participants (indicated by a red box in Fig. 4.1a) uses only three cues and has lower information gain than the no TP condition. As time pressure becomes the most stringent in the intense TP, the most probable decision model selected by human participants (indicated by a dark blue box in Fig. 4.1b) uses only two cues and exhibits even lower information gain than observed in the previous two time pressure conditions. These results demonstrate the trade-offs made by human participants among time pressure, model complexity, and information gain. As time pressure increases, individuals adaptively opt for simpler decision models with fewer cues, and sacrificed information gain to meet the decision deadline, thus reflecting the cognitive adaptation of human participants in response to time constraints.

4.2 Satisficing Strategy Modeling under Time Pressure

Inspired by human participants' satisficing behavior indicated by the data analysis above, this dissertation develops three heuristic decision models, which accommo-

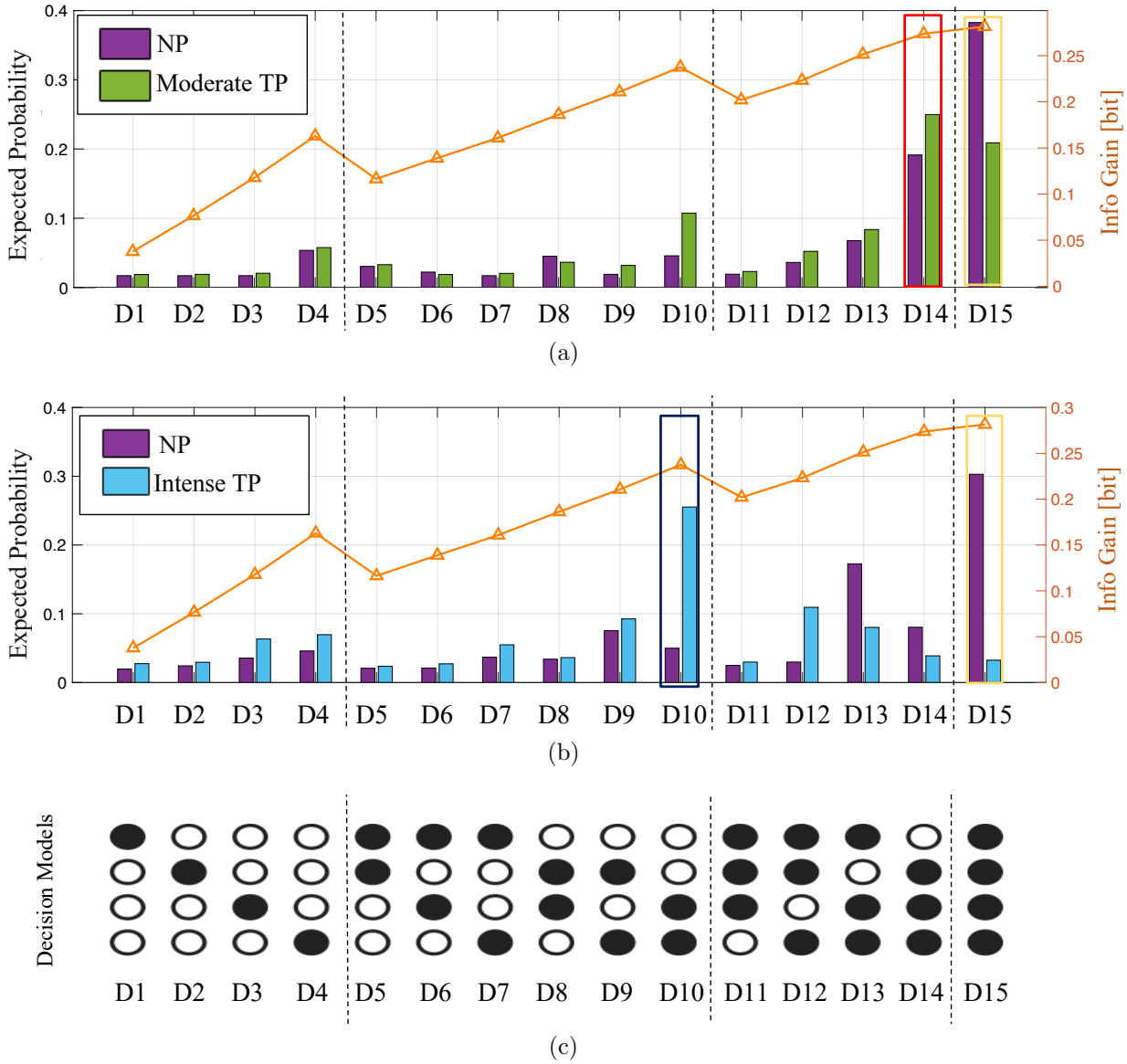


Figure 4.1: Human data analysis results for (a) the moderate TP experiment and (b) the intense TP experiment with (c) the enumeration of decision models.

date varying levels of time pressure and adaptively select a subset of information-significant cues to solve the inferential decision making problems. The heuristics assume that the measurement of cues is exact without any error.

4.2.1 Discounted Cumulative Probability Gain (ProbGain)

The heuristic is designed to incorporate two aspects of cue selection. First, the heuristic assesses the information value of individual cues and encourages the use of cues that provide valuable information for decision-making. By summing up the information value of each cue, the heuristic prioritizes the cues that contribute the most to evidence accumulation. Second, the heuristic also considers the cost of using multiple cues in terms of processing time. By applying a higher discount to models with more cues, the heuristic discourages excessive cost on time that might lead to violation of time constraints.

For an inferential decision-making problem with **sorted** cue measurements $\{m_i\}_{i=1}^{\wp}$ according to the information gain $v_I(m_i)$ in descending order, where $v_I(m_i)$ representing the increase in information value with respect to the maximum a-posterior rule

$$v_I(m_i) = \max_{y \in \mathcal{Y}} p(Y = y|m_i) - \max_{y \in \mathcal{Y}} p(Y = y) \tag{4.5}$$

Let $M_i = \{m_1, m_2, ..., m_i\}$ represent a subset of cue measurements that contains the first (i) most informative cues with respect to $v_I(m_i)$, where t_b is the allowable time to make a classification decision, and the discount factor $\gamma(t_b) \in (0, 1)$ is defined to be a function of t_b to represent the penalty induced by time pressure. Then, the heuristic strategy can be modeled as follows,

$$H_{\text{ProbGain}}(t_b, \{m_i\}_{i=1}^{\wp}) = \underset{i}{\text{argmax}}\{\gamma(t_b)^i \sum_{j=1}^{i} v_I(m_j)\} \tag{4.6}$$

where

$$\gamma(t_b) = e^{-\frac{\lambda}{t_b}} \tag{4.7}$$

4.2.2 Discounted Log-odds Ratio (LogOdds)

Log odds ratio plays a central role in classical algorithms like logistic regression [4], and represents the "confidence" of making a inferential decision. The update of log odds ratio with respect to a "new cue" is through direct summation, thus taking advantage of the cue independence and arriving at fast evidence accumulation. Furthermore, the use of log odds ratio in the context of time pressure is slightly modified such that a discount is applied with inclusion of an additional cue to penalize the cue usage because of time pressure. By combining the benefits of direct summation for fast evidence accumulation and the discount for time pressure, the heuristic based on log odds ratio can make efficient decisions by considering the most relevant cues under time constraints.

For an inferential decision-making problem with **sorted** cue measurements $\{m_i\}_{i=1}^{\wp}$ according to the information gain $|v_I(m_i)|$ in descending order, where $|v_I(m_i)|$ represents the log odds ratio of cue measurement m_i. Then, the heuristic strategy can be modeled as follows,

$$H_{\text{LogOdds}}(t_b, \{m_i\}_{i=1}^{\wp}) = \underset{i}{\text{argmax}}\{\gamma(t_b)^i | v_0 + \sum_{j=1}^{i} v_I(m_j)|\} \tag{4.8}$$

where

$$v_I(m_i) = log(p(m_i|y_1)) - log(p(m_i|y_2))$$

$$v_0 = log(p(Y = y_1)) - log(p(Y = y_2))$$

4.2.3 Information Free Cue Number Discounting (InfoFree)

The previous two cue selection heuristics are both based on comparison: multiple candidate sets of cues are evaluated and compared, and the heuristics select the one with the best trade-off between information value and processing time cost. However, because of comparison, the heuristics consume storage and time in processing the candidates. Thus, a simpler heuristic is propose to avoid comparisons and reduce the computation burden.

Sort the cues according to the information value $v_I(m_i)$ in descending order as $m_1, m_2, , ..., m_\wp$, and a subset of the first i most informative cues refers to as $M_i = \{m_1, m_2, ..., m_i\}$. The heuristic strategy is as follows,

$$H_{\text{InfoFree}}(t_b) = \left\lceil \wp \exp(-\frac{\lambda}{t_b}) \right\rceil \tag{4.9}$$

The outputs of the three heuristics are the numbers of cues to be fed into the model $P(Y, M)$ to make an inference decision. Some mathematical properties of the three heuristics strategies are presented in Appendix A.

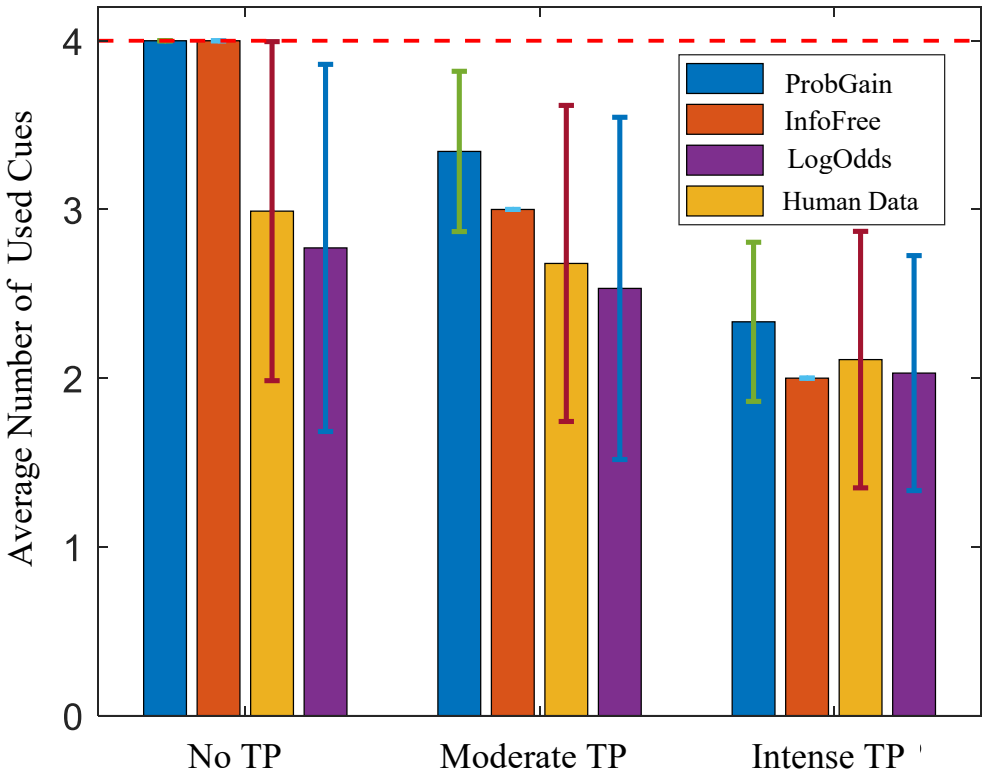


Figure 4.2: Average number of used cues and standard deviation of three heuristic strategies and human participants under different time pressure conditions.

4.3 Model Fit Test Against Human Data

The model fit tests against human data of the three proposed time-adaptive heuristics are under three time pressure levels, with the time constraints scaled to ensure comparability between human experiments and heuristic tests. The results, as shown in Fig. 4.2, indicate two major observations. First, as time pressure increases, all three strategies utilize fewer cues, thus demonstrating their adaptability to time constraints and mirroring the behavior observed in human participants. Second, among the three strategies, H_{LogOdds} exhibits the closest average number of cues and standard deviation to the human data across all time pressure conditions. Consequently, H_{LogOdds} is the heuristic strategy that best matches the human data among the three proposed strategies.

CHAPTER 5

AUTONOMOUS ROBOT APPLICATIONS OF PASSIVE SATISFICING STRATEGIES

To validate the effectiveness of the proposed passive satisficing strategies, the three heuristics learned from the human studies in Chapter 4 (H_{ProbGain}, H_{LogOdds}, H_{InfoFree}) are tested on an autonomous robot making inferential decisions on the well-established database known as car evaluation dataset [19]. This dataset contains 1728 samples, thus not requiring a complex statistics model to represent and having the right size for heuristics to model. Also, the dataset has six attributes or cues, which is big enough to reflect the heuristics' characteristics of adaptively selecting a subset of cues to make inferential decisions. The performance of the three heuristics is compared against that of a naïve Bayes classifier, referred to as "Bayes optimal" herein, which utilizes all available cues for decision-making.

The car evaluation dataset records the cars' acceptability, on the basis of six cues and originally four classes. The four classes are merged into two. A training set of 1228 samples is used to learn the conditional probability tables (CPTs), ensuring equal priors for both classes. After learning the CPTs, 500 samples are used to test the classification performance of the heuristics and the naïve Bayes classifier. The tests are conducted under three conditions: no TP, moderate TP, and intense TP.

The experiments are performed on a digital computer using MATLAB R2019b on an AMD Ryzen 9 3900X processor. The processing times of the strategies are depicted in Fig. 5.1. If a heuristic's processing time falls within the time pressure envelope (blue area), the time constraints are considered satisfied. The no TP condition provides sufficient time for all heuristics to utilize all cues for decision-

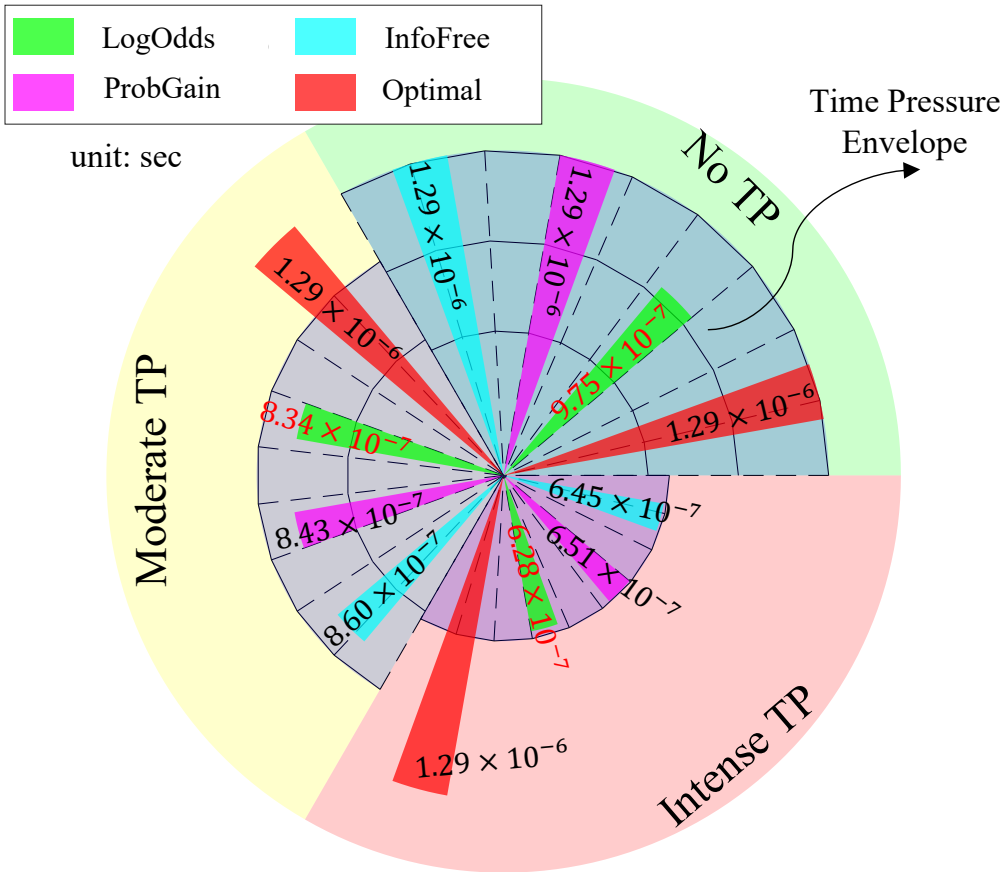


Figure 5.1: Processing time (unit: sec) of three time-adaptive heuristics and the "Bayes optimal" strategy.

making. The moderate TP condition allows for 75% of the time available in the no TP condition, whereas the intense TP condition allows for 50% of the time available in the no TP condition. All three heuristics are observed to satisfy the time constraints across all time pressure conditions.

Fig. 5.2 illustrates the classification performance and efficiency of the three time-adaptive strategies. H_{LogOdds} outperforms the other three strategies on this dataset, and its performance deteriorates as time pressure increases. Under moderate TP, the three time-adaptive strategies use fewer cues but achieve better classification performance than Bayes optimal. This finding exemplifies the less-can-be-more effect [31]. The classification efficiency measures the average contribution

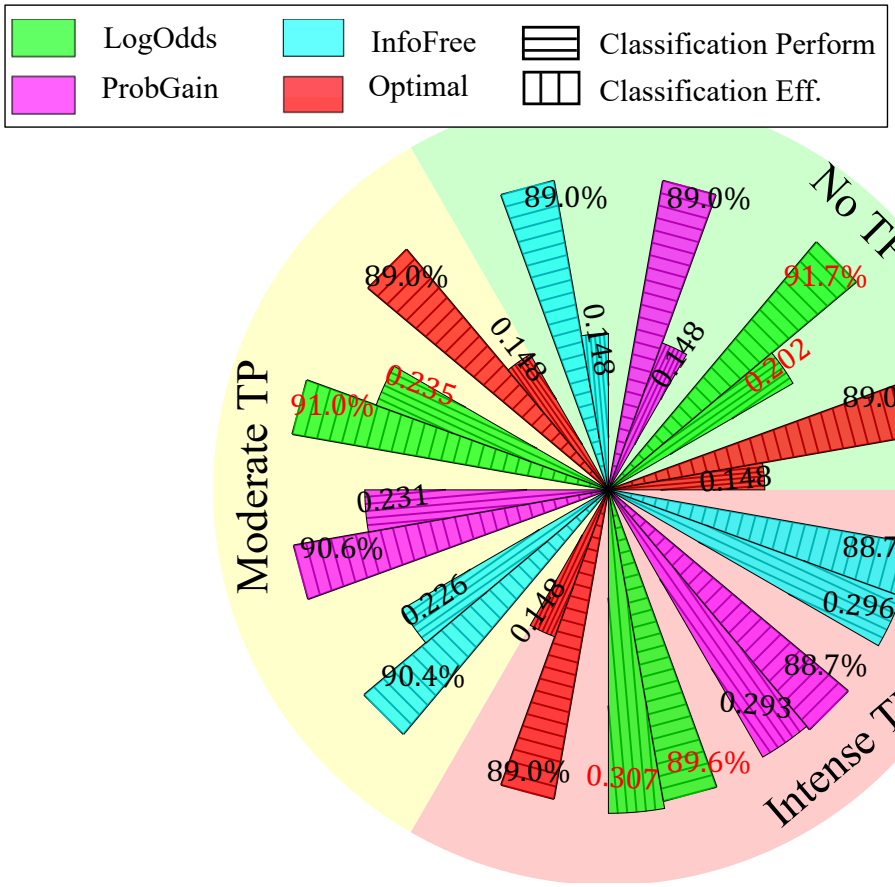


Figure 5.2: Classification performance and efficiency of three time-adaptive heuristics under three time pressure conditions.

of each cue to the classification performance. Bayes optimal has the lowest efficiency, because it utilizes all cues for all time pressure conditions, whereas H_{LogOdds} exhibits the highest efficiency among the three heuristics across all time pressure conditions.

CHAPTER 6

MATHEMATICAL MODELING OF HUMAN ACTIVE SATISFICING STRATEGIES

In the active satisficing tasks, the human participants in the experiments face pressures due to information cost (money) and sensory deprivation (fog pressure). These pressures prevent the participants from performing the test and action decisions optimally. The data analysis results for the information cost pressure, as described in Section 6.1, reveal that the test decisions and action decisions are coupled. The pressure on test decisions affect the action decisions made by the participants. The data analysis of the sensory deprivation (fog pressure) does not incorporate existing decision-making models, such as [95, 52, 27, 69], because the human participants perceive very limited information, thus violating the assumptions underlying these models. Instead, a set of decision rules are extracted in the form of heuristics from the human participants data from inspection. These heuristics capture the decision-making strategies used by the participants under sensory deprivation (fog pressure).

6.1 Information Cost (Money) Pressure

The author analyzes the human decision data with money pressure under two potential assumptions on human's decision-making incentives.

The author first assumes that the human participants make action and test decisions in order to maximize a cumulative objective function as in Eq. 6.1,

$$V = \sum_{k=0}^{T} \omega_B B(t_k) - \omega_D D(t_k) - \omega_J J(t_k) \tag{6.1}$$

where $\omega_B, \omega_D, \omega_J > 0$ are weights for three respective objectives. As target feature measurement sensor reduces the uncertainty of the hypothesis variables through making test decisions, the information gain $B(t_k)$ makes positive contributions to the objective at the cost of sensor measurements $J(t_k)$ and robot travel distance cost $D(t_k)$.

6.1.1 Human Data Analysis: Inverse Reinforcement Learning Algorithm

Under the assumption that the human participants are trying to maximize a cumulative objective function Eq. 6.1, it is natural to model to human participants' sequential decisions (action and test decisions) in Markov Decision Process (MDP) [69, 23, 75] framework. Then the problem of understanding the human decision incentives under money pressure translates to recover the mathematical representation of the human reward function given the human decision data. By studying the mathematical structure of the reward function, the author aims to find robust statistical evidence of how the pressure impact the human decision behaviors. In this dissertation, a class of algorithms called inverse reinforcement learning (IRL) [95, 52] is considered because the algorithms investigate MDPs that miss reward functions with known optimal policies, and the objective is to recover the reward functions. Two broad categories of IRL algorithms have been developed. "Max-Margin" [1, 61] is the very original version of IRL algorithms. With the assumption that the reward function is a linear combination of multiple pre-determined fea-

tures, the method aims to find a set of reward function parameters such that the cumulative reward of human/expert decision sequences are no less than that of any other policies. However, the problems of this category of methods are that 1). the parameter search process is easy to degenerate, because apparently $\underline{\mathbf{0}}$ satisfy all the inequalities; 2). the methods of this category is computationally intractable because in theory, it assumes that all possible policies can be enumerated to compare with the human/expert data. The second category is "feature expectation matching". With the assumption of linear form of reward function as well, the category of methods incorporates probabilistic thinking, and aims to find the parameters such that the likelihood against human decision data is maximized, and the recovered reward function "matches" the human data to the most extent. This dissertation adopts the second category and the algorithm is derived as follows.

A Sampling-based Algorithm Derviation

Suppose a trial of experiment data is in the form of configuration-decision trajectories $\tau_j = \{< \mathbf{q}_k, a_k, u_k >\}_{k=1}^{N_j}$ that record a human participant's configuration and decision history, where k is the time step, j is the index of the data trial and N_j is the number of entries in the jth data trial. The full data set is denoted as $D_{hum} = \{\tau_j\}_{j=1}^{N}$. Consider the objective is the linear combination of three separate objectives parameterized by a weight vector $\boldsymbol{\omega} = [\omega_B \quad \omega_D \quad \omega_J]^T$ as shown in Eq. 6.1. In this dissertation, it is decided to adopt the method described in [95, 42] to estimate the weight vector $\boldsymbol{\omega}$ based on the data D_{hum}.

As mentioned earlier, suppose the weight vector is $\boldsymbol{\omega} = [\omega_B \quad \omega_D \quad \omega_J]^T$ and the *feature reward vector* at time k is $\mathbf{f}_k = [B(t_k) \quad D(t_k) \quad J(t_k)]^T$. Thus, the total reward along jth data sequence is:

$$R(\tau_j) = \boldsymbol{\omega}^T \mathbf{f}_{\tau_j}$$

where $\mathbf{f}_{\tau_j} = \sum_{k=1}^{N_j} \mathbf{f}_k$, i.e. the sum of the reward features along the jth sequence.

The MaxEnt approach assumes that a data sequence τ samples from the distribution

$$p(\tau) = \frac{\exp(\boldsymbol{\omega}^T \mathbf{f}_\tau)}{Z}$$

where $Z = \int \exp(\boldsymbol{\omega}^T \mathbf{f}_\xi) d\xi$.

Given the data set D_{hum}, the objective is to estimate the weight vector $\boldsymbol{\omega}$ to maximize the log likelihood. Concretely,

$$\begin{aligned} lik(\boldsymbol{\omega}) &= \log \prod_{j=1}^{N} \frac{\exp(\boldsymbol{\omega}^T \mathbf{f}_{\tau_j})}{Z} \\ &= \boldsymbol{\omega}^T \sum_{j=1}^{N} \mathbf{f}_{\tau_j} - N \log Z \end{aligned}$$

Without generality, the problem reduces to the maximum likelihood problem:

$$\begin{aligned} \max_{\boldsymbol{\omega}} L(\boldsymbol{\omega}) &= \frac{1}{N} lik(\boldsymbol{\omega}) \\ &= \boldsymbol{\omega}^T \sum_{j=1}^{N} \mathbf{f}_{\tau_j} - \log Z \\ &= \boldsymbol{\omega}^T \hat{\mathbf{f}} - \log Z \end{aligned} \tag{6.2}$$

Consider the partition function Z [4] is an integral of all possible configuration-decision sequences, which is analytically intractable to compute. A common approach to approximate Z is importance sampling. A proposal distribution of the configuration-decision history τ is denoted as $q(\tau)$, then draw H sample sequences $D_{samp} = \{\tau_i\}_{i=1}^{H}$ from the distribution and use Monte-Carlo method to approximate the expectation.

$$\begin{aligned} Z &= \int \exp(\boldsymbol{\omega}^T \mathbf{f}_\xi) d\xi \\ &= \int \frac{\exp(\boldsymbol{\omega}^T \mathbf{f}_\xi) q(\xi)}{q(\xi)} d\xi \\ &= \mathbb{E}_q[\frac{\exp(\boldsymbol{\omega}^T \mathbf{f}_\xi)}{q(\xi)}] \\ &\approx \frac{1}{H} \sum_{i=1}^{H} \frac{\exp(\boldsymbol{\omega}^T \mathbf{f}_{\tau_i})}{q(\tau_i)} \end{aligned}$$

The existing methods that generate the data set $D_{samp} = \{\tau_i\}_{i=1}^{H}$, according to Boularias [5] and Kalakrishnan [42], are to uniformly sample states and actions or to sample state-action sequences from high dimensional Gaussian around the demonstrated sequences. These methods are not applicable to this problem because there exists obstacles and constrained situations to make test decisions, which make the samples from uniform distribution or Gaussian distribution be possibly infeasible to the workspace.

Therefore, in this dissertation, a simple stochastic policy $\pi(a_k, u_k|\mathbf{q}_k)$ is designed to generate a set of sample sequences D_{samp}. Then, the distribution of a sequence τ represented by the stochastic policy $\pi(a_k, u_k|\mathbf{q}_k)$ is:

$$q(\tau) = p(\mathbf{q}_0) \prod_{k=0}^{H_i-1} \pi(a_k, u_k|\mathbf{q}_k)$$

where $p(\mathbf{q}_0)$ is the probability value that initial configuration is $\mathbf{q}_0$, H_i is the length of the ith sampled sequence.

Then the gradient of the objective function Eq. 6.2 can be approximated as,

$$\frac{\partial L}{\partial \boldsymbol{\omega}} = \hat{\mathbf{f}} - \frac{M}{Z} \sum_{i=1}^{H} \frac{\exp(\boldsymbol{\omega}^T \mathbf{f}_{\tau_i})}{q(\tau_i)} \mathbf{f}_{\tau_i}$$

At maxima, the reward feature vector of the human strategy from human data and the policy under the reward function parameterized by $\boldsymbol{\omega}$ match, and thus the corresponding reward function reflects the decision preferences of human participants.

Parameter Learning Results

The averaged weights utilized by human participants are estimated using the Maximum Entropy Inverse Reinforcement Learning algorithm, adopted from [95], in order to understand the effects of money pressure on human decision behaviors. The two indices, $I_{IG} = \omega_B/\omega_D$ and $I_{IC} = \omega_B/\omega_J$, are designed by calculating the ratios of the three averaged weights, and they reflect the incentives underlying human decision behaviors. I_{IG}, the information gain attempt index, measures the willingness of human participants to trade travel distance for information gain. I_{IC}, the information cost parsimony index, measures the willingness of human participants to spend "money" (i.e., incur costs) for information gain. The results of the analysis (shown in Fig. 6.1) indicate that under the information cost (money) pressure condition, human participants are more willing to travel longer distances to acquire information gain (higher I_{IG}). However, they are less willing to incur costs (lower I_{IC}) for information gain, thus suggesting a tendency to be

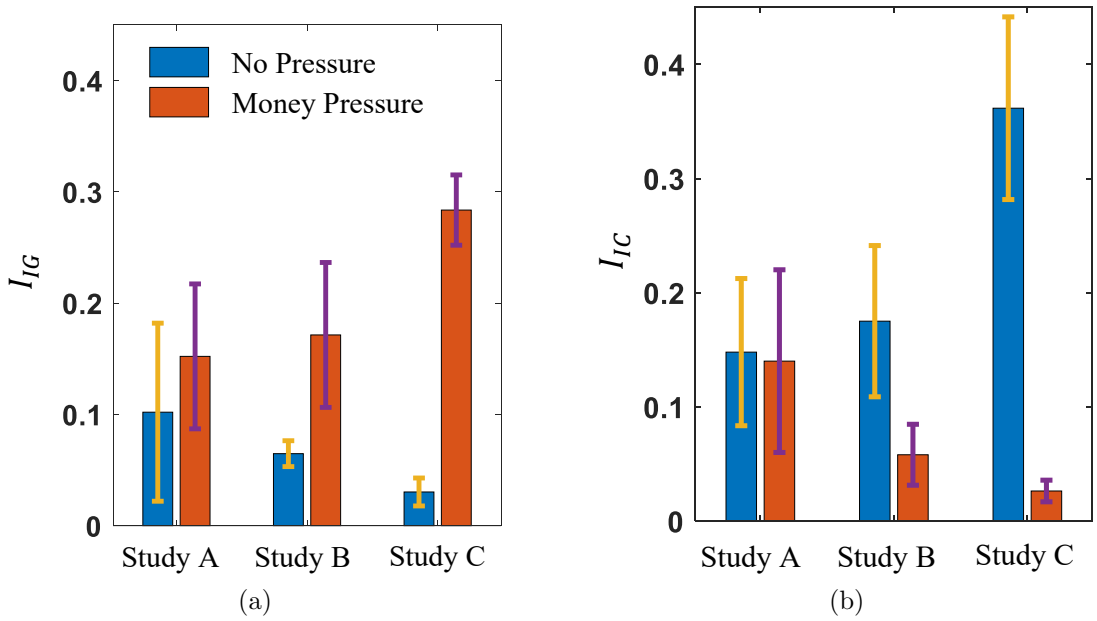


Figure 6.1: (a) Information gain attempt index I_{IG} and (b) information cost parsimony index I_{IC} of high performance human participants under two pressure conditions.

more conservative in spending resources for information acquisition.

6.1.2 Human Data Analysis: Dynamic Bayesian Network (DBN) Structure Learning

Secondly, it is assumed that no utility is associated with states or actions in human decision-making. Instead, decisions are made on this basis of causal relationships. To model human decision behavior under this assumption, this dissertation uses dynamic Bayesian networks (DBNs). The DBN intra-slice structure, as shown in Fig. 6.2, includes variables such as the human participants' states $\mathbf{q}_k$, the action decision $a(t_k)$, the test decision $u(t_k)$, the set of visible targets $o(t_k)$ at time t_k, and the "money" already spent $J(t_k)$.

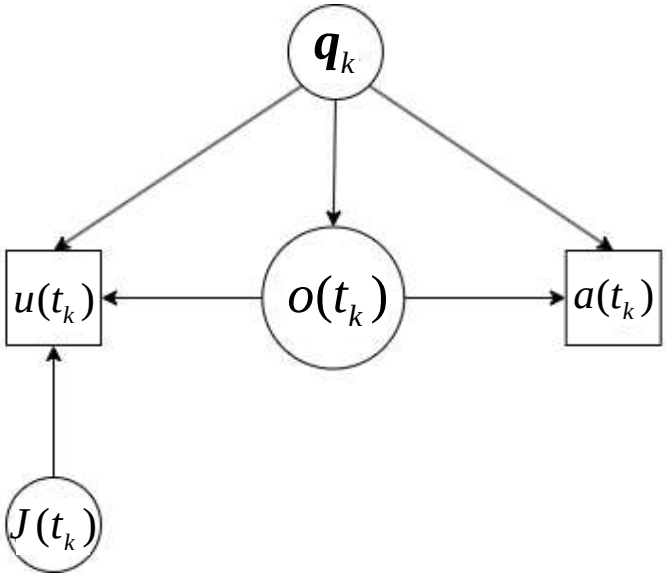


Figure 6.2: The intra-slice DBN that models human decision behavior

The intra-slice variables capture the relevant information for decision-making at a specific time slice. This dissertation investigates the inter-slice structure to understand how observations influence subsequent action and test decisions. The key question is: in how many time slices does an observation $o(t_k)$ influence decision-making? To determine the appropriate inter-slice structure, this dissertation conducts a series of hypothesis tests to assess the conformity of various models against the human decision data. Concretely, a F-test [14, 91] is conducted on two models: fit two models to human decision data, let the two models generate their decision predictions respectively, and check if one model is significant better than the other in term of a F-statistic metric. The algorithm is shown as Alg. 1.

Fig. 6.3 presents the results of these hypothesis tests. Each data point represents a p-value that evaluates the null hypothesis: "model $i+1$ does not fit the human data significantly better than model i". The models are defined according to the number of time slices in which an observation influences decisions. If the p-value is smaller than the significance level α, the null hypothesis is rejected, thus

Algorithm 1 InterSliceStructLearn

Input: D : the human data; α : the significance level

$l \leftarrow 0$

repeat

  $l \leftarrow l + 1$

  $N \leftarrow \text{countSamples}(D)$

  $p(a(t_k), u(t_k)|o(t_k), ..., o(t_{k-l+1}), \mathbf{q}_k) \leftarrow \text{learnCPT}(D)$

  $p(a(t_k), u(t_k)|o(t_k), ..., o(t_{k-l+1}), o(t_{k-l}), \mathbf{q}_k) \leftarrow \text{learnCPT}(D)$

  Initialize $\hat{b}[1, ..., N]$

  Initialize $\widetilde{b}[1, ..., N]$

  for $k := 1$ to N **do**

    $\hat{a}(t_k), \hat{u}(t_k) \leftarrow \text{argmax}\ \{p(a(t_k), u(t_k)|o(t_k), ..., o(t_{k-l+1}), \mathbf{q}_k)\}$

    $\widetilde{a}(t_k), \widetilde{u}(t_k) \leftarrow \text{argmax}\ \{p(a(t_k), u(t_k)|o(t_k), ..., o(t_{k-l+1}), o(t_{k-l}), \mathbf{q}_k)\}$

  end for

  $\hat{df} = N - df(a(t_k)) - df(u(t_k)) - l \times df(o(t_k))$

  $\widetilde{df} = N - df(a(t_k)) - df(u(t_k)) - (l+1) \times df(o(t_k))$

  $F = \frac{(\sum_{k=1}^{k=N} \mathbb{1}\{a(t_k)=\hat{a}(t_k), u(t_k)=\hat{u}(t_k)\} - \sum_{k=1}^{k=N} \mathbb{1}\{a(t_k)=\widetilde{a}(t_k), u(t_k)=\widetilde{u}(t_k)\})/(\hat{df}-\widetilde{df})}{\sum_{k=1}^{k=N} \mathbb{1}\{a(t_k)=\widetilde{a}(t_k), u(t_k)=\widetilde{u}(t_k)\}/\widetilde{df}}$

  $F^* = \text{F distribution}(\alpha, \hat{df}, \widetilde{df})$

until $F < F^*$

return l

indicating that the subsequent model fits the data better than the previous one. According to the results in Fig. 6.3, under the no pressure condition, an observation $o(t_k)$ influences one subsequent decision. However, under the money pressure, an observation $o(t_k)$ influences nine subsequent decisions. This finding suggests that the influence of observations extends over a longer time horizon under money pressure than in the no pressure condition. The final BN model is presented as Fig. 6.4.

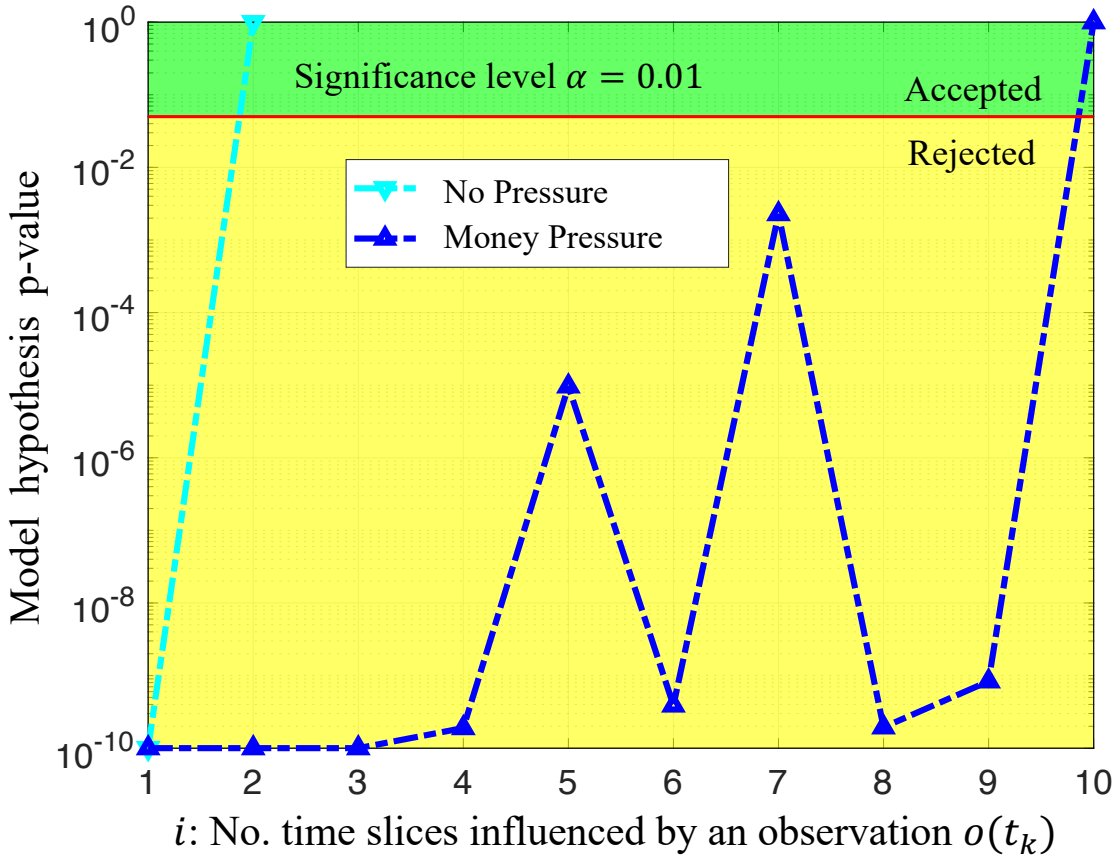


Figure 6.3: DBN inter-slice structure hypothesis testing results

6.2 Sensory Deprivation (Fog Pressure)

The introduction of sensory deprivation (fog pressure) in the environment poses two main difficulties for human participants during navigation. First, fog limits the visibility range, thus hindering human participants' capability of locating targets and being aware of obstacles. Second, fog impairs spatial awareness, thus hindering human participants' ability to accurately perceive their own position within the workspace.

In situations in which target and obstacle information is scarcely accessible, and the uncertainties are difficult to model, the human participants are believed to use local information to navigate in the environment, obtain target feature

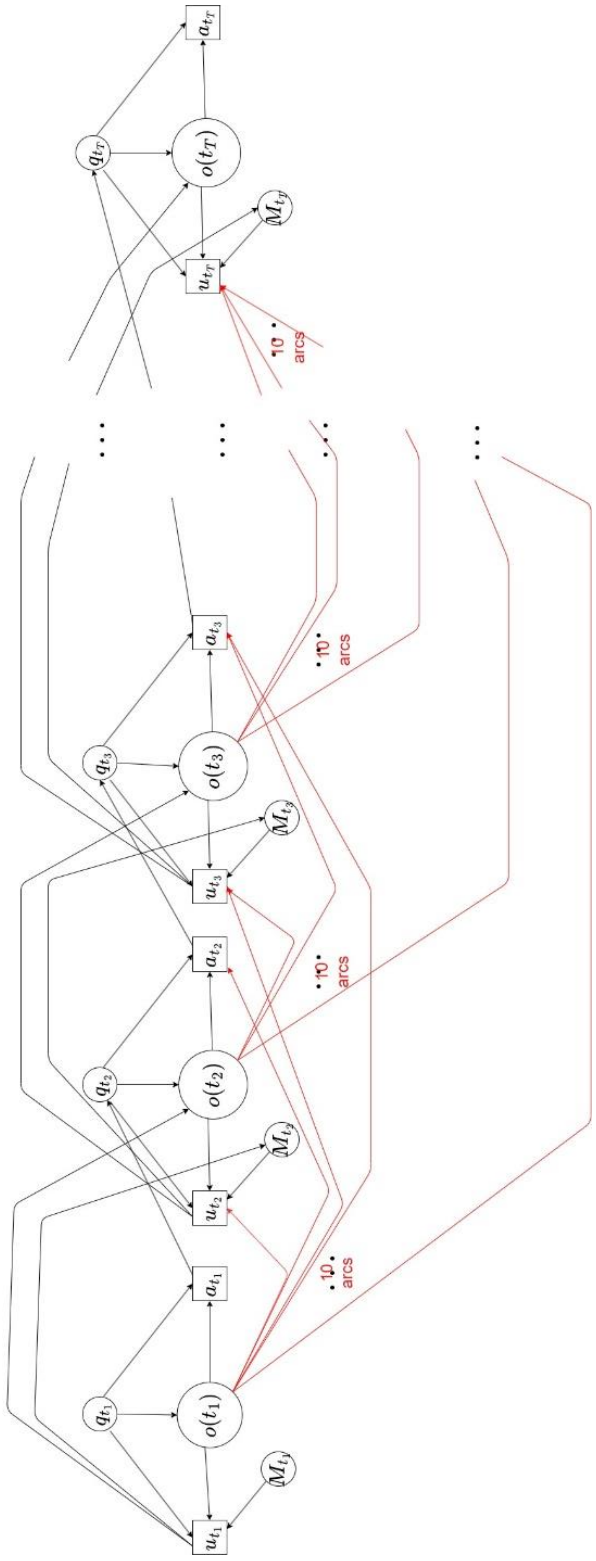
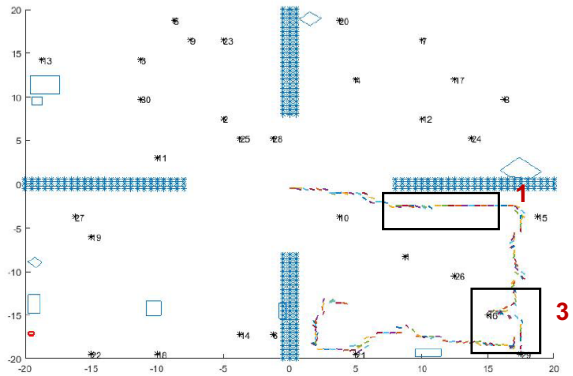
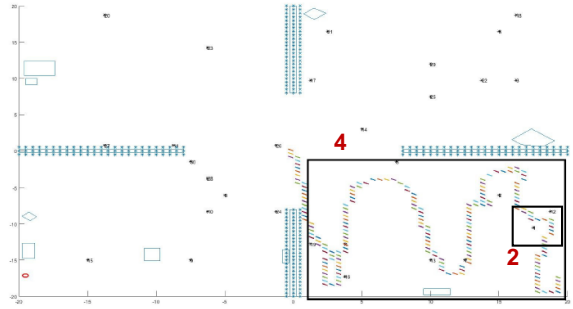


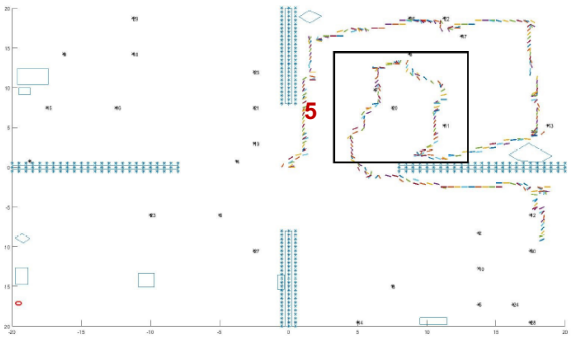
Figure 6.4: Human DBN decision model under money pressure. The observation of visible targets at time t_k will influence the subsequent action and test decisions over 10 time slices.



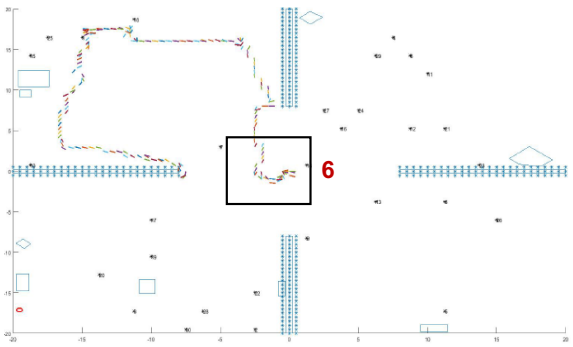
(a)



(b)



(c)



(d)

Figure 6.5: The human behavior patterns in a fog environment.

measurements, and classify the targets [31, 17]. This dissertation analyzes the human decision data collected in the active satisficing experiment and inspect the significant behavioral patterns shared by the human participants. The results are shown in Fig. 6.5, in which the following six behavioral patterns are observed:

1. When participants enter an area and no targets are immediately visible, they tend to walk along the walls or obstacles present in the environment. This behavior is depicted in Fig. 6.5a.
2. When a participant spot multiple targets, they tend to pursue the targets one by one, according to their proximity. This behavior is depicted in Fig. 6.5b.
3. While walking along a wall or obstacle, if participants spot a target, they will deviate from their original path and pursue the target, and may then return to their previous "wall/obstacle follow" path. This behavior is depicted in Fig. 6.5a.
4. Upon entering an area, participants may engage in a strategy of covering the entire room. This behavior is depicted in Fig. 6.5b.
5. After walking along a wall or obstacle for some time without encountering any targets, participants are likely to change their exploratory strategy. This behavior is depicted in Fig. 6.5c.
6. In the absence of visible targets, participants may exhibit random walking behavior. This behavior is depicted in Fig. 6.5d.

Analysis of the six behavioral patterns observed in the active satisficing experiment reveals three underlying incentives that drive human participants' behaviors in the presence of fog pressure, as follows:

- **Frugal**: Human participants exhibit tendencies to avoid repeated visitations. Navigating along walls or obstacles help participants localize themselves with respect to the surrounding environment, by using the walls or obstacles as reference points.
- **Greedy**: Human participants demonstrate a strong motivation to find targets and engage with them. After a target is detected, participants pursue it and interact with it immediately.
- **Adaptive**: Human participants display adaptability by using multiple strategies for exploring the workspace. These strategies include "wall/obstacle following," "area coverage," and "random walk." Participants can switch among these strategies according to the effectiveness of their current approach in finding targets.

On the basis of the observations above, this dissertation develops an algorithm called **AdaptiveSwitch** (Algorithm 2) to model human strategies in a fog environment. This algorithm captures the adaptive nature of human participants' strategies and allows for transitions among the simple heuristics under specific conditions.

The algorithm uses three exploratory heuristics: wall/obstacle following (π_1), area coverage (π_2), and random walk (π_3). The probability of executing each heuristic is referred to as $\Pi = [b_1, b_2, b_3]^T$, where b_i represents the probability of executing π_i. The index J indicates the exploratory policy being executed, and k represents the number of steps taken while executing a policy. The maximum number of steps before updating the distribution Π is K. The policy for interacting with targets is $\pi_I(u(t_k)|\mathbf{q}_k, o(t_k))$, and the policy for pursuing a target is $\pi_P(a(t_k)|\mathbf{q}_k, o(t_k))$.

Algorithm 2 AdaptiveSwitch

1: $\Pi = [b_1, b_2, b_3]^T$
2: $k = 0, J = 0$
3: **while** ($t_k \leq t_T \vee$ not all targets are classified) **do**
4: **if** $\exists \mathbf{x}_j \in \mathcal{S}_I(\mathbf{q}_k)$ **then**
5: $\pi_I(u(t_k)|\mathbf{q}_k, o(t_k))$
6: **else**
7: **if** $o(t_k) \neq \emptyset$ **then**
8: $\pi_P(a(t_k)|\mathbf{q}_k, o(t_k))$
9: $k = 0, J = 0$
10: **else**
11: **if** $J > 0 \ \wedge \ k \leq K$ **then**
12: $\pi_J(a(t_k)|\mathbf{q}_k, o(t_k))$
13: $k = k + 1$
14: **else**
15: **if** $k \geq K$ **then**
16: $\Pi[J] = \gamma * \Pi[J]$
17: **else**
18: **if** not closed to wall **then**
19: $\Pi[1] = 0$
20: **else**
21: $\Pi[1] = \beta(b_1 + b_2 + b_3)$
22: **end if**
23: **end if**
24: $\Pi =$normalize(Π)
25: $J \sim \Pi$
26: **end if**
27: **end if**
28: **end if**
29: **end while**

In Algorithm 2, the switch logic among the heuristics is described. The algorithm demonstrates the greediness of the heuristic strategy (lines 4 - 9), in which participants interact with targets if possible (line 4) and pursue a target if it is visible (line 7). If no targets are visible and the maximum exploratory step K is not exceeded, the current exploratory heuristic continues to be executed (lines 11-13). The adaptiveness of the three exploratory heuristics is shown in lines 15 - 22. If the current exploratory heuristic is executed for more than K steps, its probability of

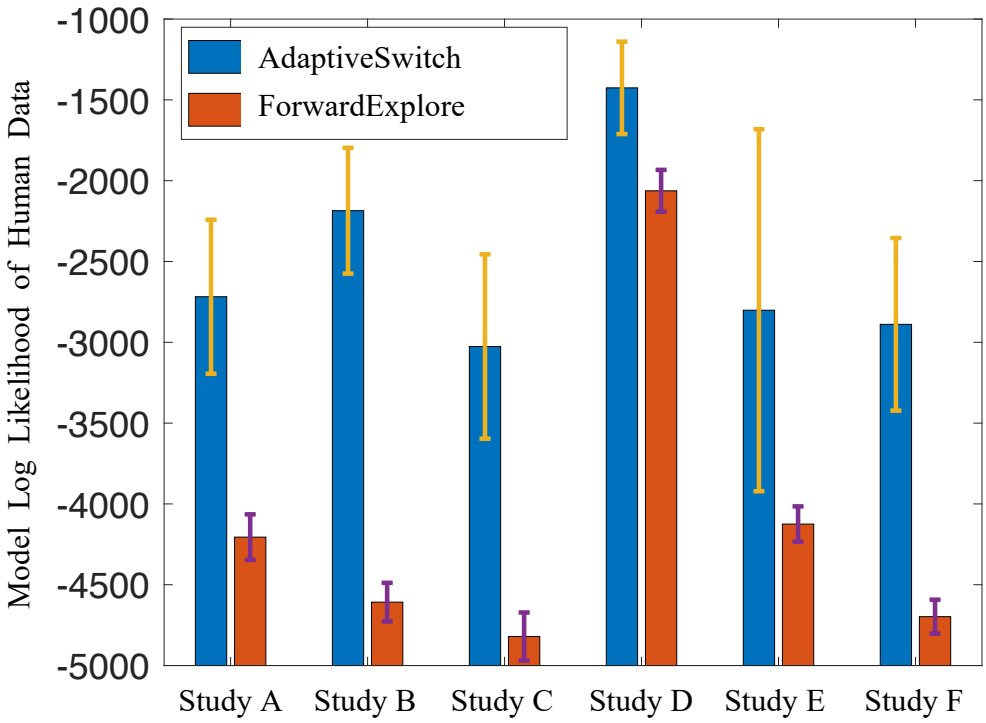


Figure 6.6: Averaged model log likelihood of AdaptiveSwitch and ForwardExplore in six human studies.

execution is discounted (line 16). The probability of executing the "wall/obstacle following" heuristic increases if the participant is close to a wall/obstacle; otherwise this heuristic is disabled (lines 19 - 21). Additionally, **ForwardExplore** heuristic is proposed, where participants predominantly move forward with a high probability and turn with a small probability or when encountering an obstacle.

This dissertation evaluates the log likelihood of AdaptiveSwitch and ForwardExplore against the human data from the active satisficing experiment involving six participants. The results in Fig. 6.6 show that the log likelihood of AdaptiveSwitch is greater than that of ForwardExplore across all human experiment trials. This finding suggests that AdaptiveSwitch aligns more closely with the observed human strategies than ForwardExplore.

CHAPTER 7

AUTONOMOUS ROBOT APPLICATIONS OF ACTIVE SATISFICING STRATEGIES

Two key contributions of this dissertation are the applications of the modeled human strategies on a robot, and the comparison of optimal strategies and the modeled human strategies in pressure conditions, under which optimization is infeasible. For simplicity, the preferred sensing directions of $\mathcal{S}_P$ and $\mathcal{S}_I$ are assumed to be fixed with respect to the robot platform. Therefore, the state vector for a robot reduces to $\mathbf{q} = [x \quad y \quad \theta]^T$, where the orientation of the robot platform θ also represents the preferred sensing directions. Both sensor FOVs are modeled by sectors with angle-of-view $\alpha_1, \alpha_2 \in [0, 2\pi)$ and radii $r_1, r_2 > 0$. The two FOVs share the same apex and their bisectors coincide with each other.

7.1 Information Cost (Money) Pressure

The introduction of information cost increases the complexity of planning test decisions. In the absence of information cost, a greedy policy that measures all available features for any target is considered "optimal", because it collects all information gain without any cost. However, when information cost is taken into account, a longer planning horizon for test decisions becomes crucial to effectively allocate the budget for measuring features of all targets. This dissertation proposes two planners, a probabilistic roadmap (PRM) based planner and a cell decomposition based planner, to solve the robot treasure hunt problem, which has identical workspace, initial conditions, and target layouts to the problem faced by human participants in the active satisficing experiment. The objective function Eq. 6.1 is maximized by using these methods. Unlike existing approaches [11, 20, 93]

that solve the original version of the treasure hunt problem as described in [21], the developed planners handle the problem without pre-specification of the final robot configuration. Consequently, the search space increases exponentially, thus rendering label-correcting algorithms [3] no longer applicable. Additionally, unlike previous methods that solely optimize the objective with respect to the path, the developed planners consider the constraint on the number of target feature measurements due to information cost (money) pressure. The number of measurements thus becomes a decision variable with a long planning horizon. To solve the problem, the developed planners use PRM and cell decomposition techniques to generate graphs representing the workspace [21]. The Dijkstra algorithm is used to compute the shortest path between targets. Furthermore, an MINLP algorithm is used to determine the optimal number of measurements and the visitation sequence of the targets. The detailed MINLP based algorithm is described as follows:

Consider the constraint brought by money pressure, as discussed in Section 3.3, can be very well described mathematically. Thus, the authors decided to solve the problem under money pressure within the optimization framework.

The methodology assumes that the sensor model $P(Y, M)$ is known and the "vision" sensor complies with the pin-hole camera model such that the working distance is infinity within the opening angle of the sensor FOV $\mathcal{S}_1$, the problem described in Chapter 2 can be transformed to an optimization problem with a clear objective as in Eq. 6.1.

The proposed strategy under money pressure is a MINLP based optimal sensor planning methodology on a path planning graph with observations [11] $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ generated via methods like PRM ad cell decomposition as described in [11, 56], where $\mathcal{N}$ denotes the set of nodes and $\mathcal{E}$ denotes the set of edges. Two types of

nodes in $\mathcal{N}$, the observation nodes that support target feature sensor measurements and void nodes that don't. Denote a node that supports measurement with $\overline{(\cdot)}$ and the set of all nodes that support target feature measurements on the ith target as $\overline{\mathcal{K}}_i$.

A path in graph $\mathcal{G}$ is defined as a sequence of connected nodes in $\mathcal{G}$, as a channel C [48]:

$$C = \{N_1, ..., N_f\}, \text{where}(N_l, N_{l+1}) \in \mathcal{E} \tag{7.1}$$

The mobile robot's path follows the configurations specified by the sequence of nodes in channel C.

Given only initial configuration $\mathbf{q}_0$, label-correcting/A*-type algorithms [3] are not applicable to search the optimal path/channel directly. Therefore, a two-step algorithm is applied to solve the problem:

1. Sample one node from $\overline{\mathcal{K}}_i, i = 1, ...r$ that support target feature measurements on ith target and add the node of initial configuration $\mathbf{q}_0$. Use Dijkstra algorithm to compute shortest path/channel C^*_{ij} with distance cost D^*_{ij} between every pair of the $r+1$ nodes. For the optimal channel from $\overline{N}_i$ to $\overline{N}_j$:

$$R(C^*_{ij}) = \max_{\wp_j} \{\omega_B B(\wp_j) - \omega_D D^*_{ij} - \omega_J \wp_j\} \tag{7.2}$$

   where $\wp_j$ is the number of target feature measurements made on jth target, $B(\wp_j)$ is the information gain collected on jth target from $\wp_j$ measurements,

and information cost is encoded as number of target feature measurements directly.

2. Use as $\mathbf{q}_0$ the initial configuration, apply MINLP to determine which channel C^*_{ij} should be included in the final optimal path to maximize Eq. 6.1.

Note the channel $C^*_{i,j}$ only considers the information gain on node $\overline{N}_j$ but not $\overline{N}_i$. Therefore, the reward along $C^*_{i,j}$ is not necessarily equal to that along $C^*_{j,i}$.

After applying Dijkstra to compute $C^*_{ij}, 1 \leq i, j \leq r+1$, the sensor planning problem is formulated as a MINLP:

The decision variable $\mathbf{V}$ is a matrix of $(r+1) \times (r+1)$ binary variables, where

$$\mathbf{V}_{ij} = \begin{cases} 1 \text{ if channel } C^*_{ij} \text{ is in the optimal path} \\ 0 \text{ otherwise} \end{cases} \tag{7.3}$$

The node of initial configuration is represented at the 1^{st} column and 1^{st} row in the matrix $\mathbf{V}$.

Then the objective is as follows:

$$\max_{\mathbf{V}} \sum_{i,j} R(C^*_{i,j}) \mathbf{V}_{i,j} \tag{7.4}$$

The constraints are described as follows:

1. The initial configuration is specified and the path should start at $\mathbf{q}_0$:

$$\begin{aligned} \sum_j \mathbf{V}_{1j} &= 1 \\ \sum_{i,,i\neq 1} \mathbf{V}_{i1} &= 0 \end{aligned} \tag{7.5}$$

2. Each target should be visited/classified at most once:

$$\begin{aligned} \sum_j \mathbf{V}_{ij} &\leq 1 \quad \forall i \\ \sum_i \mathbf{V}_{ij} &\leq 1 \quad \forall j \end{aligned} \tag{7.6}$$

3. The path should be connected:

   If $\mathbf{V}_{ij} = 1$, meaning that channel C^*_{ij} is in the optimal path, then there should be a channel that connects to C^*_{ij}, which leads to $\sum_k \mathbf{V}_{ki} = 1$(except that i is the initial configuration). Otherwise, if $\mathbf{V}_{ij} = 0$, $\sum_k \mathbf{V}_{ki}$ can be either 0 or 1.

$$\mathbf{V}_{ij}(\sum_k \mathbf{V}_{ki} - 1) = 0 \quad \forall i, j, i \neq 1 \tag{7.7}$$

4. There should be no circle in the path:

   Let the row vector $\boldsymbol{\lambda} = \overbrace{[0...\underbrace{1}_{j^{th}}...0]}^{r+1}$ represent node j in the path represented by $\mathbf{V}$, then $\boldsymbol{\lambda} \times \mathbf{V}$ is the next node in the path after j. If $\boldsymbol{\lambda}$ is the final node in the path then $\boldsymbol{\lambda} \times \mathbf{V} = \mathbf{0}_{1\times(r+1)}$ because the final node doesn't point to any other node. Thus, if there is no circle in the path, then

$$\forall \boldsymbol{\lambda},\ \boldsymbol{\lambda} \times \mathbf{V}^{r+1} = \mathbf{0}_{1\times(r+1)}. \tag{7.8}$$

   However, if there exists a circle in the path $\mathbf{V}$, $\exists \boldsymbol{\lambda}$ such that Eq. 7.8 doesn't hold because the final node points to the initial node in the circle. Since Eq.

7.8 is valid for arbitrary node represented by $\boldsymbol{\lambda}$, the Eq. 7.8 is equivalent to [1]:

$$\mathbf{V}^{r+1} = \mathbf{0}_{(r+1)\times(r+1)} \tag{7.9}$$

5. The target feature measurement budget should not be exceeded

$$\sum_{j=1}^{r} \wp_j \leq J_b$$

Note that the optimization problem formulated above doesn't impose a hard constraint that all targets has to be visited. This is due to the observation of the human decision data that under money pressure, many human participants don't complete visiting all targets while under no pressure condition, almost all participants do. In order to construct a fair comparison with human strategy, the optimization problem moves the "constraint" to the objective such that the planner aims to classify more targets to obtain information gain at the cost of travel distance and information cost. However, without the hard constraint, the search space (on action and test decisions) becomes exponentially large and the exact optimal search becomes intractable. Thus, in this dissertation, genetic algorithm is used to approximate the optimal solution. It is worth to mention that the 4th constraint of this optimization problem enforces there is no circle in the path. In theory, there is no need to explicitly enforce such requirement as the "global" optimal route shouldn't include a circle. However, in the implementation, as the exact global optimal solution is intractable and the genetic algorithm based optimizer is very likely to stop at a local optimal and the "no circle" constraint can't be

[1] Note that $\mathbf{V}_{1,1} = 1$ with all other entries zero is a corner case, which violates constraint (4) but represents a valid path with only the initial configuration.

automatically guaranteed. Therefore, the author explicitly enforces the "no cirle" condition.

7.1.1 Performance Comparison with Human Strategies

The performance (Fig. B.1) of two optimal strategies (PRM and cell decomposition) is compared with the human strategy from human data. Under information cost (money) pressure, the path and the number of measurements for each target are optimized with respect to a linear combination of three objectives. Definingτ as a planned continuous path [49], this dissertation focuses on four performance metrics: path efficiency $\eta_P = 1/D(\tau)$ $[m^{-1}]$; information gathering efficiency $\eta_B = B(\tau)/D(\tau)$ [bit/m]; measurement productivity $\eta_J = B(\tau)/J(\tau)$ [bit]; and classification performance $N = N(\tau)$. Higher numbers are better for all metrics. Six case studies are examined. One case study comprises of three different experiment layouts. The optimal strategies and the human participants have no prior knowledge of the target positions and initial features, and all environmental information is obtained from FOV $\mathcal{S}_P$. The results, shown in Fig. B.1, indicate that the two optimal strategies consistently outperform the human strategy across all four performance metrics. The performance envelopes of the optimal strategies are outside of the performance of the human strategy, thus indicating their superiority.

The finding that the optimal strategies outperform human strategies is unsurprising, because information cost (money) pressure imposes a constraint on only the expenditure of measurement resources, which can be effectively modeled mathematically. The finding suggests that under information cost (money) pressure, near-optimal strategies can make better decisions than human strategies.

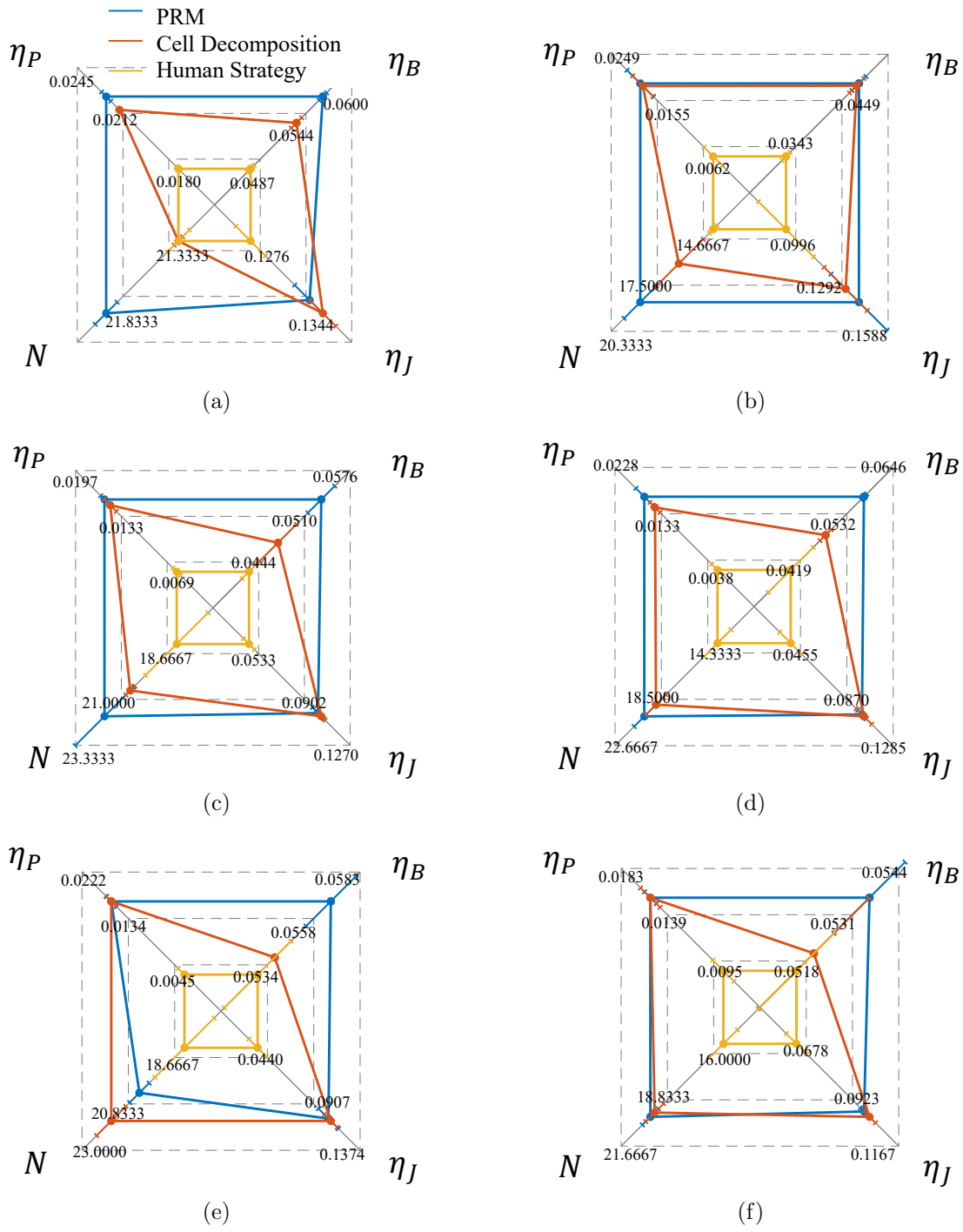


Figure 7.1: Performance comparison of two optimal strategies and human strategy over six case studies (a)-(f).

7.2 Sensory Deprivation (Fog) Pressure

An extensive series of tests are conducted to evaluate the effectiveness of AdaptiveSwitch (Chapter 6) under fog conditions and compare it with other strategies. These tests comprise of 118 simulations and physical experiments, encompassing various levels of uncertainty. The challenges posed by fog in robot planning are twofold. First, fog obstructs the robot's ability to detect targets and obstacles by using onboard sensors such as cameras, thus making long-horizon optimization-based planning nearly impossible. Second, fog complicates the task of self-localization for the robot with respect to the entire map, although short-term localization can rely on inertial measurement units. Three test groups are described as follows:

7.2.1 Performance Tests in the Human Experiment Workspace

AdaptiveSwitch is applied to the workspace and target layouts in the active satisficing experiment workspace described in Fig. 3.3 and Fig. 3.4. The experiment involves six human participants, each of whom completes three trials with different target layouts, thus resulting in a total of eighteen different target layouts with a uniform obstacle layout.

The performance of AdaptiveSwitch is compared with that of optimal strategies and the human strategy. One important metric used to evaluate a strategy's capability to search for targets in fog conditions is the number of classified targets: N_v. Under fog pressure, as shown in Fig. 7.2a and Fig. 7.2b, the optimal strategies

face difficulties in moving and classifying targets, because of the lack of prior information on the target and obstacle layouts. In contrast, both the human strategies and the AdaptiveSwitch are able to explore the unknown environment, and even at times do not capture target information through $\mathcal{S}_P$. AdaptiveSwitch, in particular, achieves slightly higher target classification rates and shorter travel distances than the human strategy.

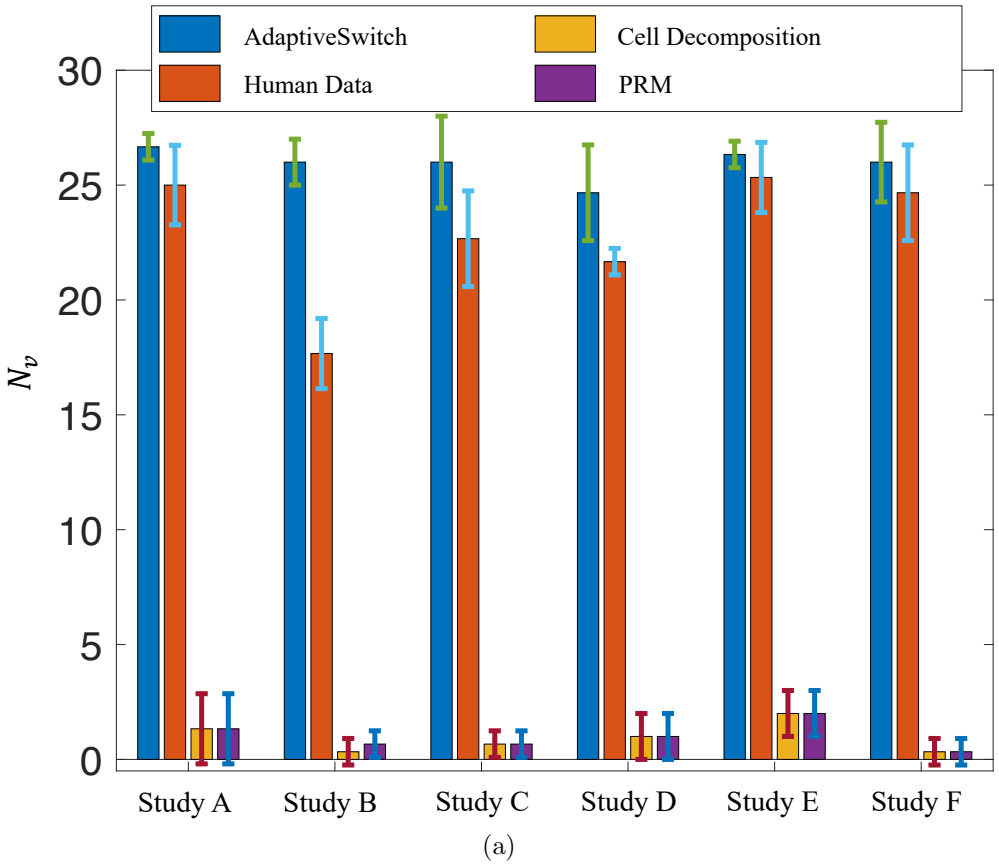
7.2.2 Extended Performance Tests in Simulations

This dissertation also presents new workspaces and target layouts beyond those used in the active satisficing experiment. These new layouts are used to assess the performance of AdaptiveSwitch in different environments and to determine its applicability beyond the specific experimental settings.

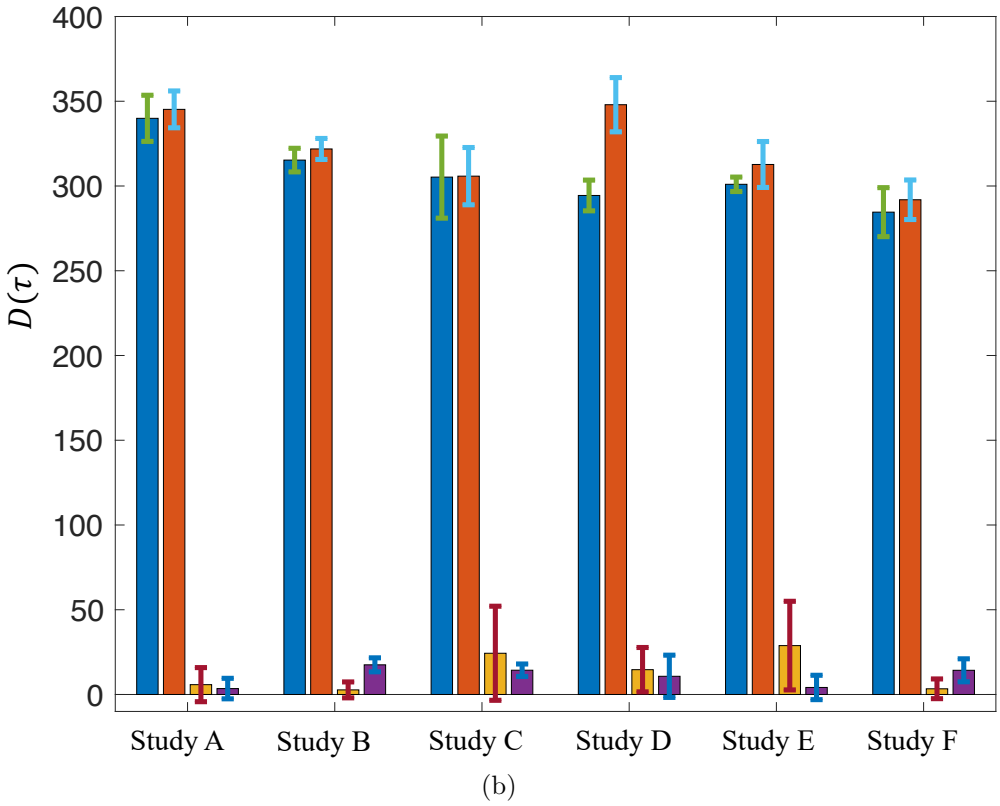
Simulations with Fixed Truncated Sensing Range

The evaluation of AdaptiveSwitch involves conducting simulations in MATLAB®, by using four newly designed workspaces and corresponding target layouts. The effect of fog is emulated by imposing a fixed truncated sensing range for $\mathcal{S}_P$, and the trajectory of AdaptiveSwitch is superimposed on each workspace to observe its behavior (Fig. 7.4). The simulations consider fixed geometries for the FOV of the onboard sensors, assume no target miss detection or false alarms, and assume perfect target feature recognition. ForwardExplore and two optimal strategies are also implemented for comparison.

The results of the simulations demonstrate that the optimal strategies perform poorly in terms of travel distance $D(\tau)$ and the number of classified targets (N_v)



(a)



(b)

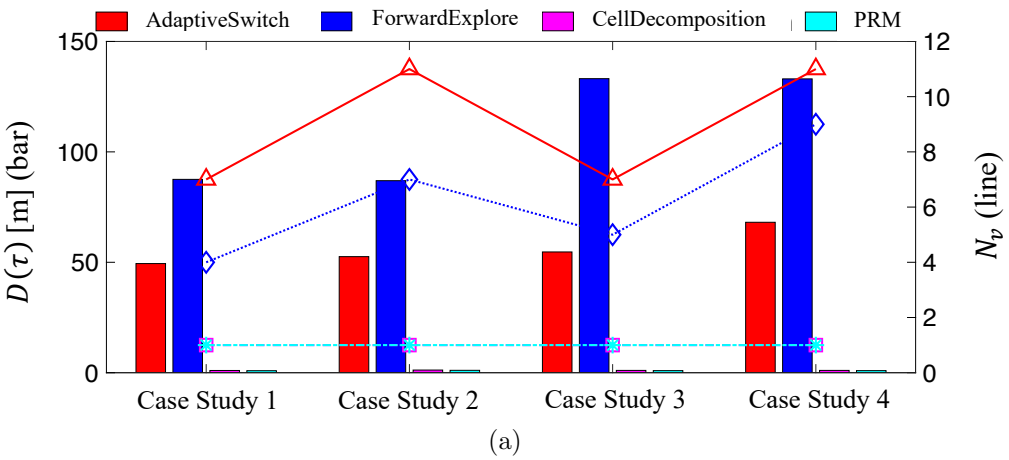
Figure 7.2: (a) Number of classified targets and (b) travel distance of AdaptiveSwitch optimal strategies and the human strategy.

(Fig. 7.3a), owing to the challenges posed by fog and limited sensing capabilities. In contrast, the proposed heuristics (AdaptiveSwitch and ForwardExplore) outperform the optimal strategies in terms of N_v, because they are able to explore the workspace even when no targets were visible.

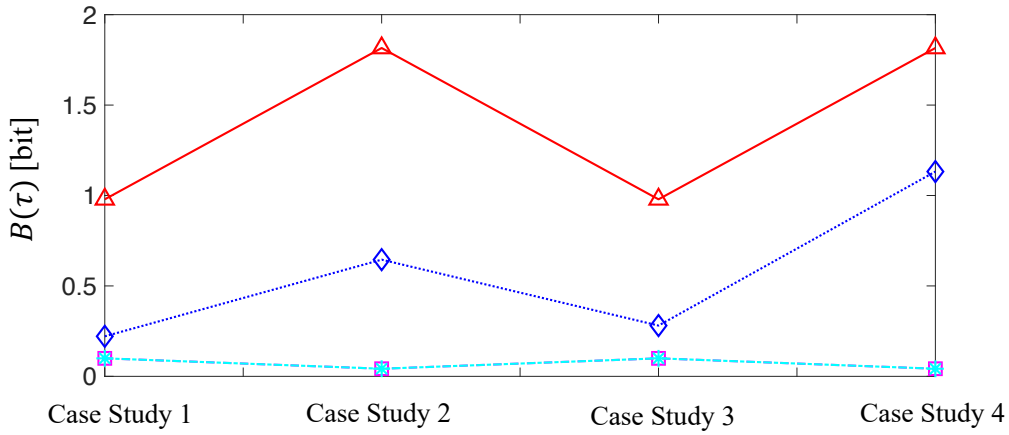
Additionally, AdaptiveSwitch is more efficient than ForwardExplore in terms of travel distance. By adapting its exploration strategy and leveraging the combination of three simple heuristics, AdaptiveSwitch is able to classify more targets while traveling shorter distances. Consequently, higher information gain $B(\tau)$ than that with both ForwardExplore and the optimal strategies is observed across all four case studies (Fig. 7.3b). These findings highlight the effectiveness of the AdaptiveSwitch in navigating foggy environments and its superiority to the optimal strategies and the ForwardExplore in terms of information gathering and travel efficiency.

Simulations with Artificial Fog

Two new workspaces are designed in Webots® as shown in Fig. 7.5. The performance of AdaptiveSwitch and its standalone heuristics for the two workspaces is shown in Table 7.1 and Table 7.2. The comparison reveals the substantial advantage of AdaptiveSwitch. In both workspace scenarios, as shown in Table 7.1 and 7.2, AdaptiveSwitch outperforms its standalone heuristics by successfully finding and classifying all targets within the given simulation time upper bound. In contrast, the standalone heuristics are unable to achieve this level of performance. AdaptiveSwitch not only visits and classifies all targets, but also accomplishes the tasks within shorter travel distances than the standalone heuristics. Therefore, AdaptiveSwitch exhibits higher target visitation efficiency (η_v) which is calculated



(a)



(b)

Figure 7.3: (a). Number classified targets and travel distance (b) information gain for two heuristic strategies and two optimal strategies in four case studies.

as the ratio of the number of classified targets to the travel distance ($N_v/D(\tau)$). The target visitation efficiency of AdaptiveSwitch is at least twice higher than that of the standalone heuristics.

These results highlight the strength of combination used by AdaptiveSwitch. By integrating multiple simple heuristics, AdaptiveSwitch demonstrates a greater ability to explore the entire environment in the presence of fog. In contrast, the standalone heuristics tend to be less flexible and may become trapped in certain "moving patterns"; therefore, although they can explore some areas effectively,

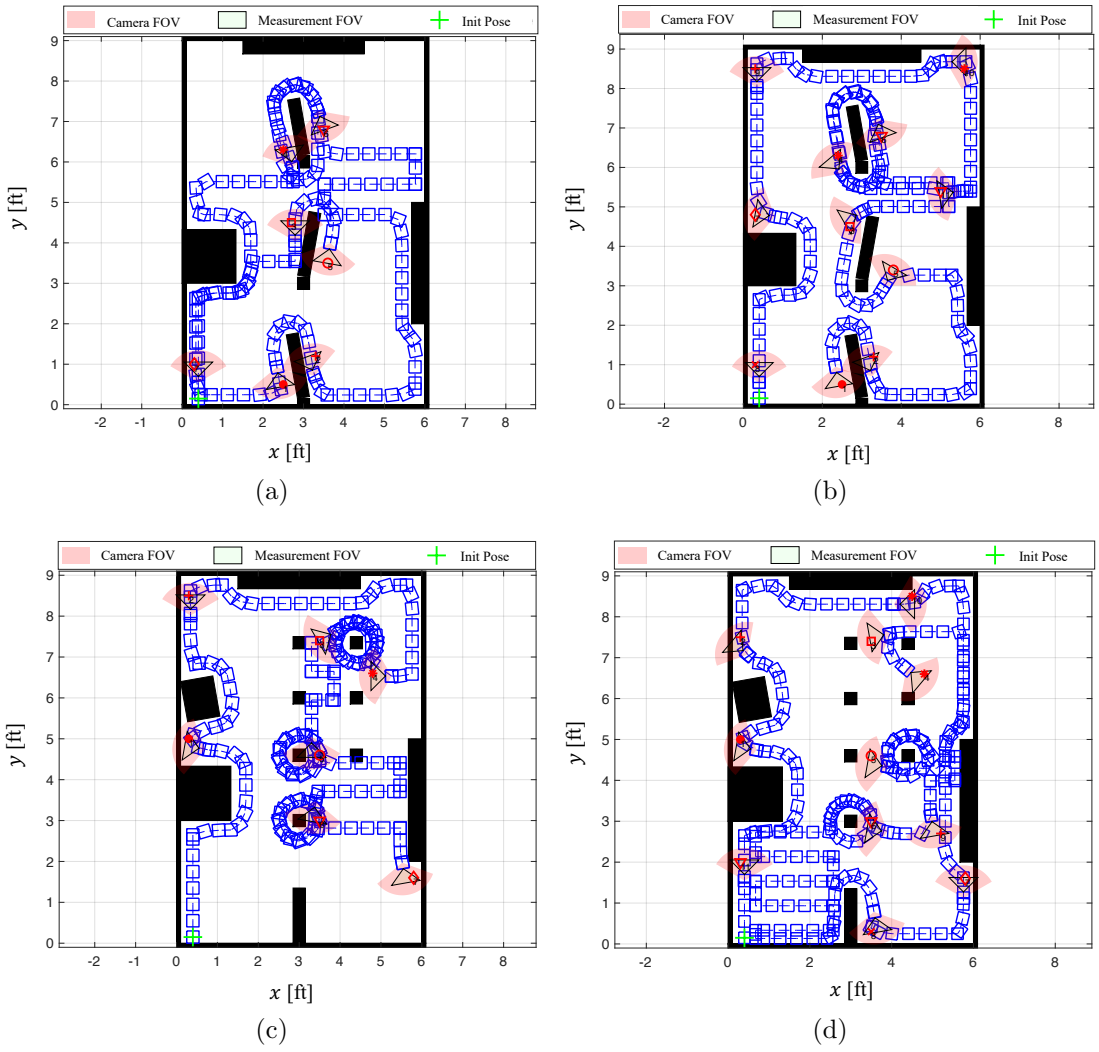
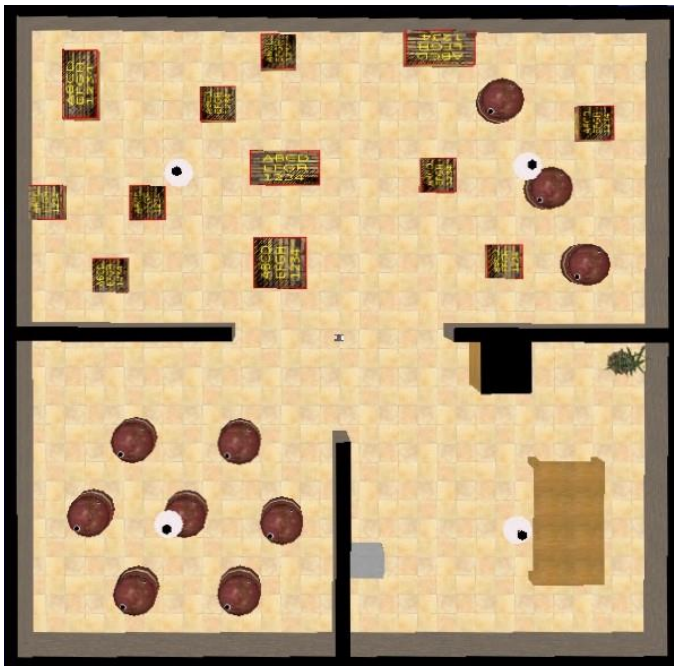
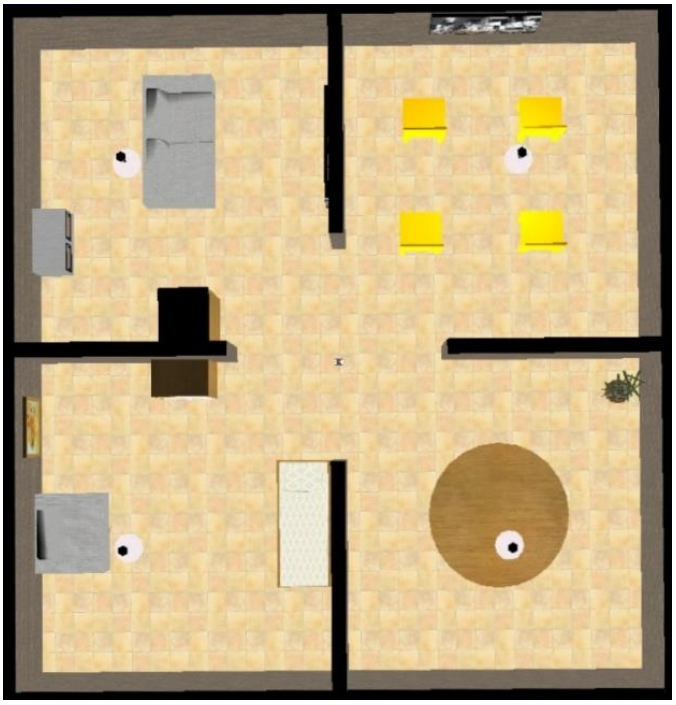


Figure 7.4: Four workspace in MATLAB® simulations and AdaptiveSwitch trajectories for case studies (a)-(d).

they might struggle to reach other areas.



(a)



(b)

Figure 7.5: New designs of workspace for heuristic strategy tests.

Table 7.1: Performance comparison of AdaptiveSwitch and Standalone heuristics in Webots®: Workspace A

Performance Metrics	Heuristic Strategies		
	AdaptiveSwitch	RandomWalk	AreaCoverage
Travel distance, $D(\tau)$ [m]	**86.19**	164.87	224.18
Number of classified targets, N_v	**7**/7	7/7	3/7
Target visitation efficiency, η_v [m^{-1}]	**0.0812**	0.0425	0.0134
Travel distance, $D(\tau)$ [m]	**148.98**	291.69	246.38
Number of classified targets, N_v	**13**/13	11/13	6/13
Target visitation efficiency, η_v [m^{-1}]	**0.0873**	0.0377	0.0244
Travel distance, $D(\tau)$ [m]	**159.97**	236.86	205.78
Number of classified targets, N_v	**15**/15	11/15	8/15
Target visitation efficiency, η_v [m^{-1}]	**0.0938**	0.0464	0.0389

7.2.3 Physical Experiments in Fog Experiment

Sensing Interruption and Classification Performance Degradation

To handle real-world uncertainties that are not adequately modeled in simulations, this dissertation conducts physical experiments to test the AdaptiveSwitch. These uncertainties include factors such as the robot's initial position and orientation, target miss detection and false alarms, depth measurement errors, and control disturbances. In addition, the fog models available in Webots®, are relatively simple and do not provide a wide range of possibilities for simulating the degrading effects of fog on target detection and classification performance. Consequently, this dissertation performs physical experiments to better capture the complexities and uncertainties associated with real-world conditions.

The physical experiments use the ROSbot2.0 robot equipped with an RGB-

Table 7.2: Performance comparison of AdaptiveSwitch and Standalone heuristics in Webots®: Workspace B

Performance Metrics	Heuristic Strategies		
	AdaptiveSwitch	RandomWalk	AreaCoverage
Travel distance, $D(\tau)$ [m]	**122.86**	218.72	265.49
Number of classified targets, N_v	**7**/7	5/7	5/7
Target visitation efficiency, η_v [m^{-1}]	**0.0570**	0.0229	0.0188
Travel distance, $D(\tau)$ [m]	**122.57**	219.49	234.70
Number of classified targets, N_v	**13**/13	10/13	7/13
Target visitation efficiency, η_v [m^{-1}]	**0.0873**	0.0456	0.0298
Travel distance: $D(\tau)$ [m]	**129.19**	226.57	216.25
Number of classified targets, N_v	**15**/15	12/15	8/15
Target visitation efficiency, η_v [m^{-1}]	**0.1161**	0.0530	0.0370

D camera as the primary sensor. The YOLOv3 object detection algorithm is employed to detect the targets of interest (e.g., an apple, watermelon, orange, basketball, computer, book, cardboard box, and wooden box) identical to those in human experiments. The training images for the YOLOv3 are captured in a clear environment.

As depicted in Fig. 7.6, the YOLOv3 [72] algorithm successfully detects the existence of the target "computer" when the environment is clear, as shown in Fig. 7.6a. However, when fog is present, as illustrated in Fig. 7.6b, the algorithm fails to detect the target. This result demonstrates the degrading effect on the performance of target detection algorithms.

The YOLOv3 was trained with customized datasets of targets of interest (apple, watermelon, orange, basketball, computer, book, cardboardbox, woodenbox) as in the human experiments. The training images are all captured in non-fog

environment. The perception pipeline is shown as Fig. 7.7. As shown in Fig. 7.6, the YOLOv3 algorithm is able to detect the existence of the target "computer" when the environment is clear as in Fig. 7.6.A. However, when the fog exists, the algorithm fails to detect the target as in Fig. 7.6.B.

The author also develops SVM based target feature (shape, color, texture) classifiers. The impact of fog on the target feature classification correct is shown in Fig. 7.8a. Under non-fog condition, the classification correct drops significantly beyond $d = 0.61m$. For the fog condition, the classification performance drops significantly beyond $d = 0.30m$.

The localization of a target with respect to the robot's body frame in the physical experiment relies on a successful detection of the target from the RGB frame and the measurement of the relative distance from the target to camera from depth frame (RGB-D camera). The depth measurement of our device is through Structured Light [43]. Our analysis shows that the existence of fog not only influences the detection of the target, but also make the depth measurement nosier, which is confirmed by Quintana et al. in [70].

In the physical experiments conducted with ROSbot2.0 [40], AdaptiveSwitch and ForwardExplore are implemented to test their performance in an environment with fog. A plastic box is constructed with dimensions 10'0" x 6'0" x 1'8" in order to create the foggy environment. The box is designed to contain different layouts of obstacles and targets, capturing various aspects of a "treasure hunt" scenario, such as target density and target view angles. Each heuristic strategy is tested five times in each layout, considering all the uncertainties described earlier. The travel distances in the physical experiments are measured in inertial measurement unit.

The first layout (Fig. 7.9) comprises of six targets: a watermelon, wooden box, basketball, book, apple, and computer. The target visitation sequences of AdaptiveSwitch along the path are depicted in Fig. 7.10, showing the robot's trajectory and the order in which the targets are visited. The performance of the two strategies is summarized in Table 7.3, as evaluated according to three aspects: travel distance $D(\tau)$, correct target feature classifications, and information gathering efficiency η_B. These metrics assess the quality of the strategies' action and test decisions.

Table 7.3: Performance Comparison of Heuristic Strategies in target layout 1

Performance Metrics	Heuristic Strategies	
	AdaptiveSwitch	ForwardExplore
Number of classified targets, N_v	6/6	6/6
Travel distance, $D(\tau)$ [m]	**6.43 ± 0.90**	8.38 ± 2.07
Correct target feature classifications	**13.40 ± 1.82**	12.40 ± 1.95
Info gathering efficiency, η_B [bit/m]	**0.155 ± 0.023**	0.090 ± 0.018

The second layout (Fig. 7.11) contains eight targets: a watermelon, wooden box, basketball, book, computer, cardboard box, and two apples. The obstacles layout is also changed with respect to the first layout: the cardboard box is placed in a "corner" and is visible from only one direction, thus increasing the difficulty of detecting this target. This layout enables a case study in which the targets are more crowded than in the first layout. The mobile robot first-person-views of AdaptiveSwitch along the path are demonstrated in Fig. 7.12, and the performance is shown in Table 7.4.

The third layout (Fig. 7.13) contains two targets: a cardboard box, and a watermelon. Note that having fewer targets does not necessarily make the problem easier, because the difficulty in target search in fog comes from how to navigate

Table 7.4: Performance Comparison of Heuristic Strategies in target layout 2

Performance Metrics	Heuristic Strategies	
	AdaptiveSwitch	ForwardExplore
Number of classified targets, N_v	8/8	8/8
Travel distance, $D(\tau)$ [m]	**8.41 ± 0.46**	13.45 ± 2.10
Correct target feature classifications	**17.80 ± 1.10**	15.20 ± 1.64
Info gathering efficiency, η_B [bit/m]	**0.151 ± 0.008**	0.091 ± 0.016

when no target is in the FOV. This layout intentionally makes the problem "difficult", because it "hides" two targets behind the walls. The mobile robot first-person-views of AdaptiveSwitch along the path are demonstrated in Fig. 7.14, and the performance is shown in Table 7.5. The videos for all physical experiments (AdaptiveSwitch and ForwardExplore in three layouts) are accessible through the link in [15].

Table 7.5: Performance Comparison of Heuristic Strategies in target layout 3

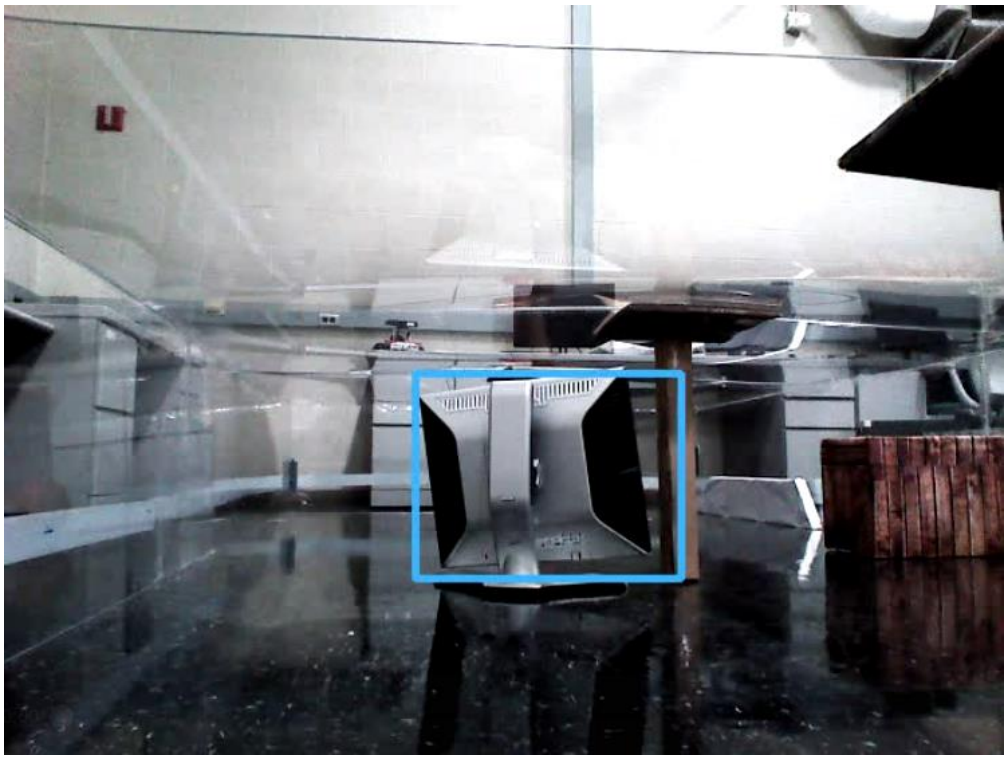
Performance Metrics	Heuristic Strategies	
	AdaptiveSwitch	ForwardExplore
Number of classified targets, N_v	2/2	2/2
Travel distance, $D(\tau)$ [m]	**7.48 ± 0.465**	11.67 ± 1.37
Correct target feature classifications	**5.00 ± 1.00**	4.80 ± 1.64
Info gathering efficiency, η_B [bit/m]	**0.033 ± 0.003**	0.021 ± 0.002

According to the performance summaries in Table 7.3, Table 7.4, and Table 7.5, both AdaptiveSwitch and ForwardExplore are capable of visiting and classifying all targets in the three layouts under real-world uncertainties. However, AdaptiveSwitch demonstrates several advantages over ForwardExplore:

1. The average travel distance of AdaptiveSwitch is **30.33%**, **59.93%**, and **56.02%** more efficient than ForwardExplore in the three workspaces, respec-

tively. This finding indicates that AdaptiveSwitch is able to search target with a shorter travel distance than ForwardExplore.

2. The target feature classification performance of AdaptiveSwitch is slightly better than that of ForwardExplore, with improvements of 8.06%, 17.11%, and 4.16% in the three workspace, respectively. One possible explanation for these results is that the "obstacle follow" and "area coverage" heuristics in AdaptiveSwitch cause the robot's body to be parallel to obstacles during classification of target features, thus ensuring that the targets are the major part of the robot's first-person view and make them relatively easier to classify. In contrast, ForwardExplore does not always lead the robot body to be parallel to obstacles during classification, thereby sometimes allowing obstacles to dominate the robot's first-person view and decreasing the target classification performance.



(a)

(b)

Figure 7.6: Object detection results (a) in clear and (b) fog conditions.

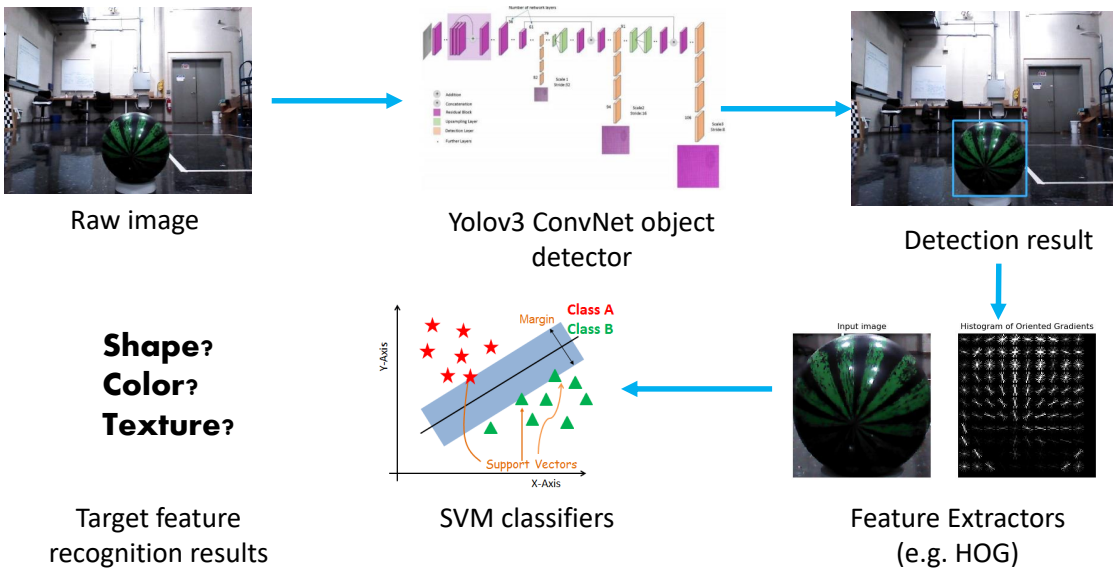
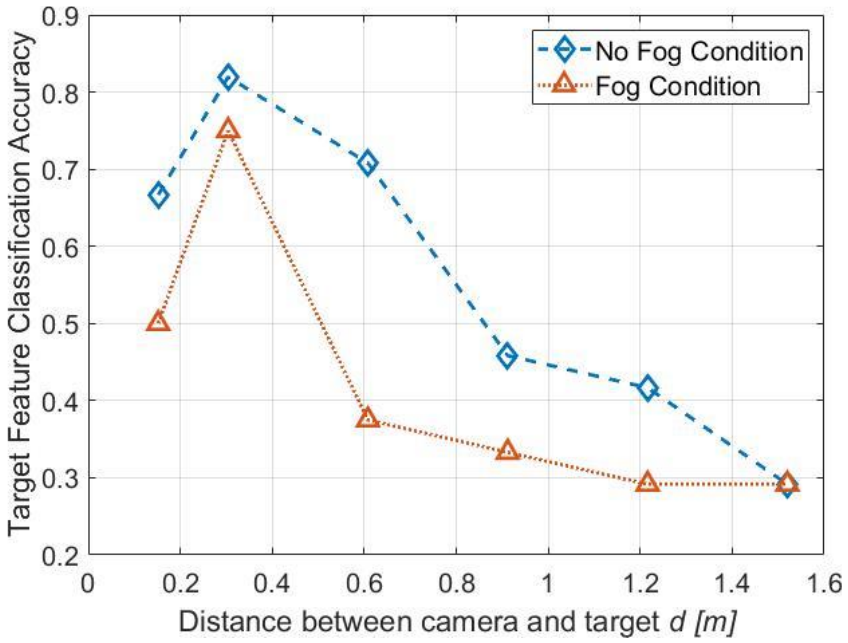
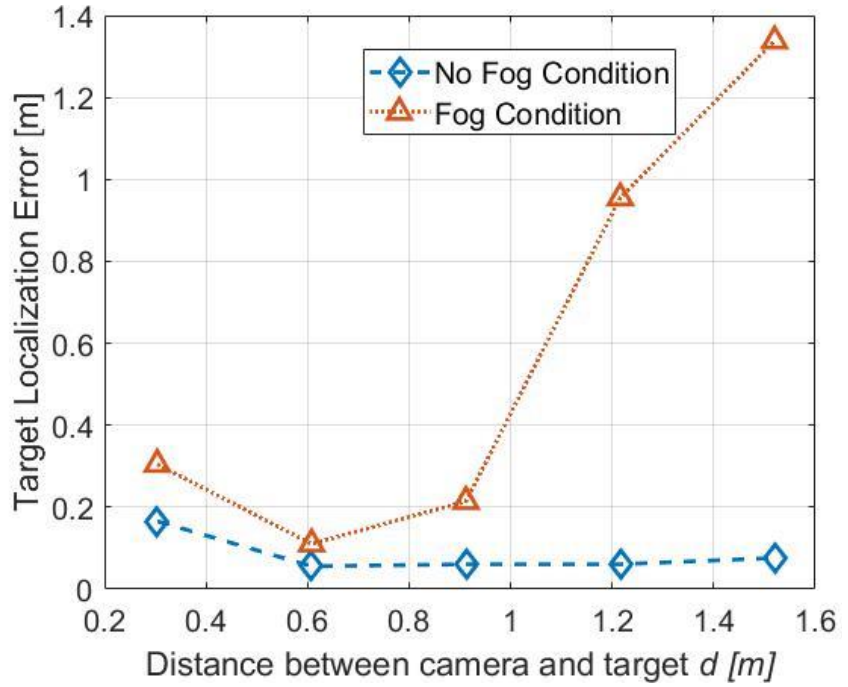


Figure 7.7: The CovNet based perception pipeline. The objective of the process is to use a object detector to identity the existence of the target of interest, and then use multiple SVM classifiers to sequentially recognize the target features: shape, color, texture.



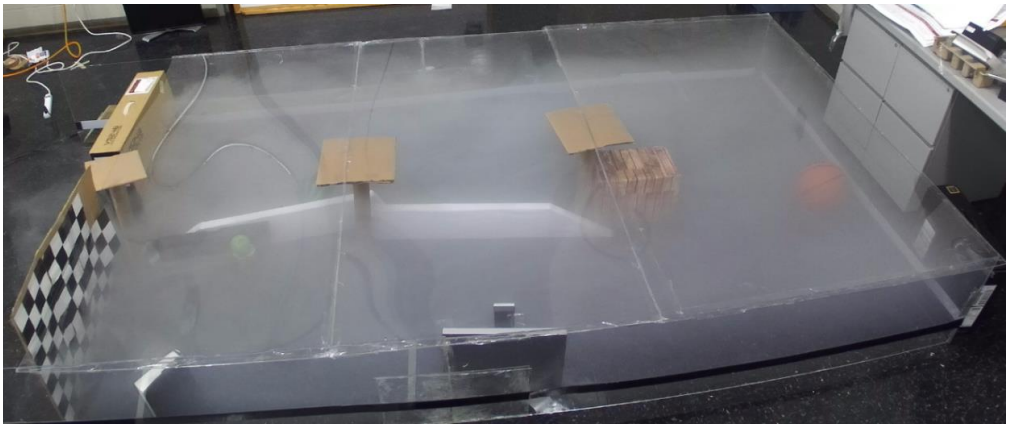
(a)



(b)

Figure 7.8: (a). Overall target feature (color, shape, texture) classification accuracy versus the measurement distance in physical experiment. (b). The target localization error with respective to the distance between a camera and a target. In fog environment, as distance increases, it is likely to fail detect a target and thus the localization error increases very quickly.

(a)



(b)

Figure 7.9: The first workspace and target layout for the physical experiment under (a) clear and (b) fog condition.

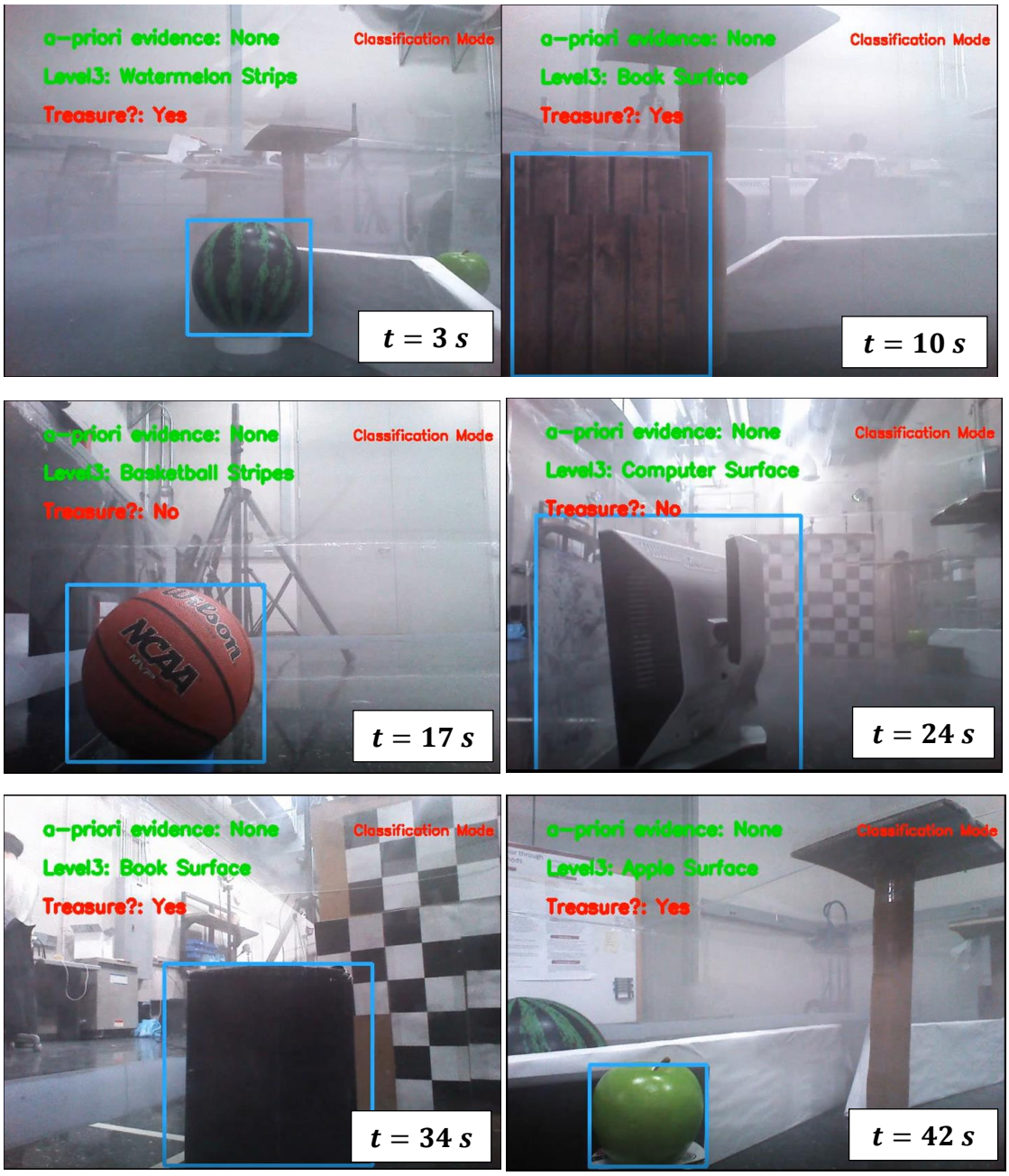
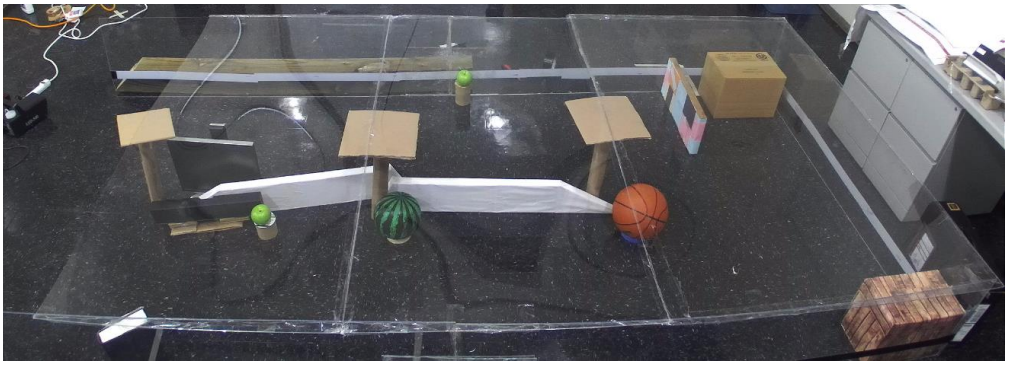
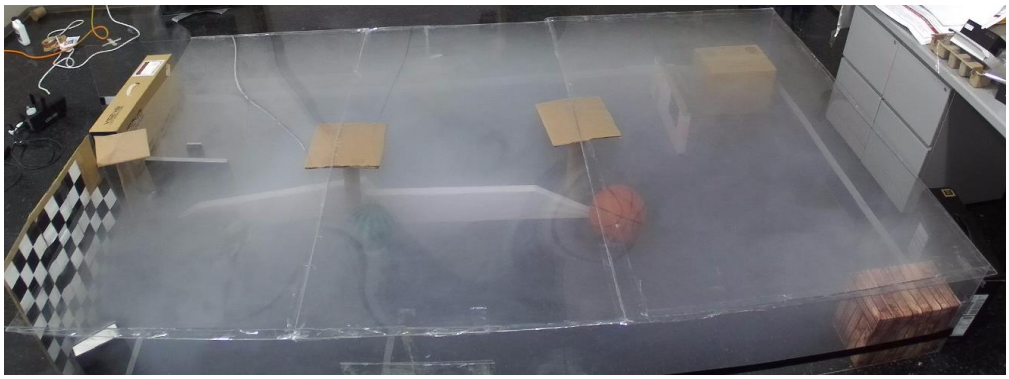


Figure 7.10: Target visitation sequence of AdaptiveSwitch in the first workspace.



(a)



(b)

Figure 7.11: The second workspace and target layout for the physical experiment under (a) clear and (b) fog condition.

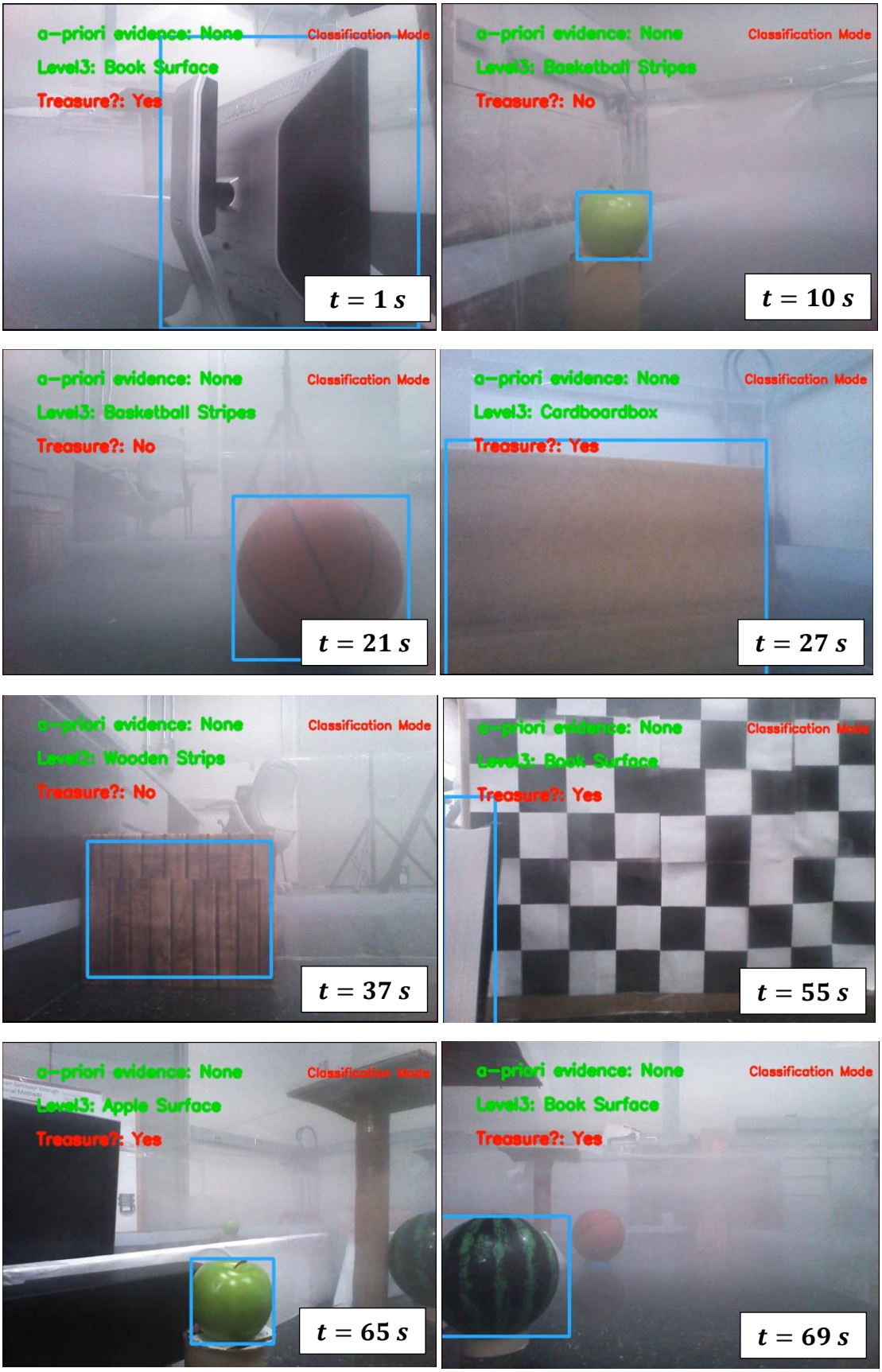
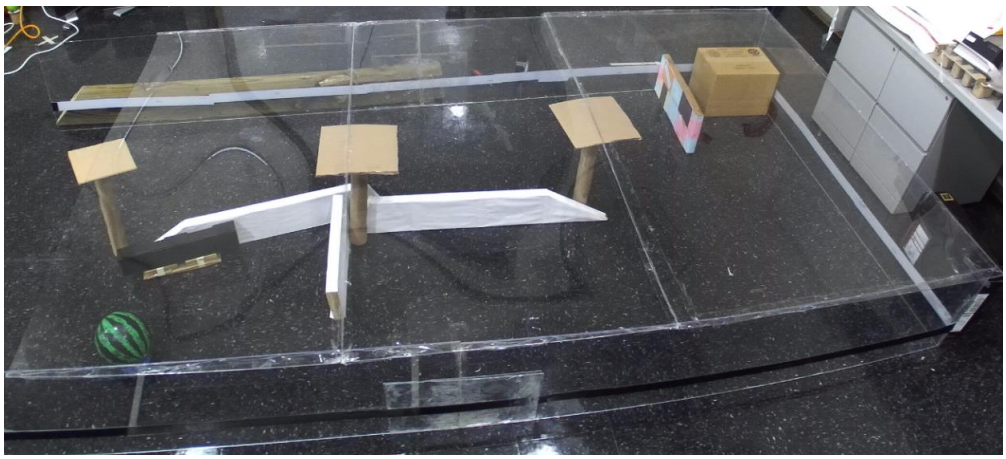
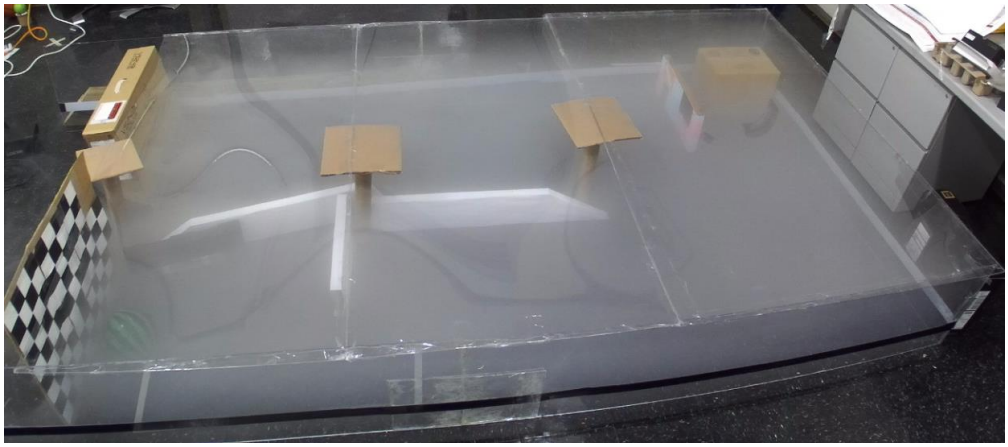


Figure 7.12: Target visitation sequence of AdaptiveSwitch in the second workspace.



(a)



(b)

Figure 7.13: The third workspace and target layout for the physical experiment under (a) clear and (b) fog condition.

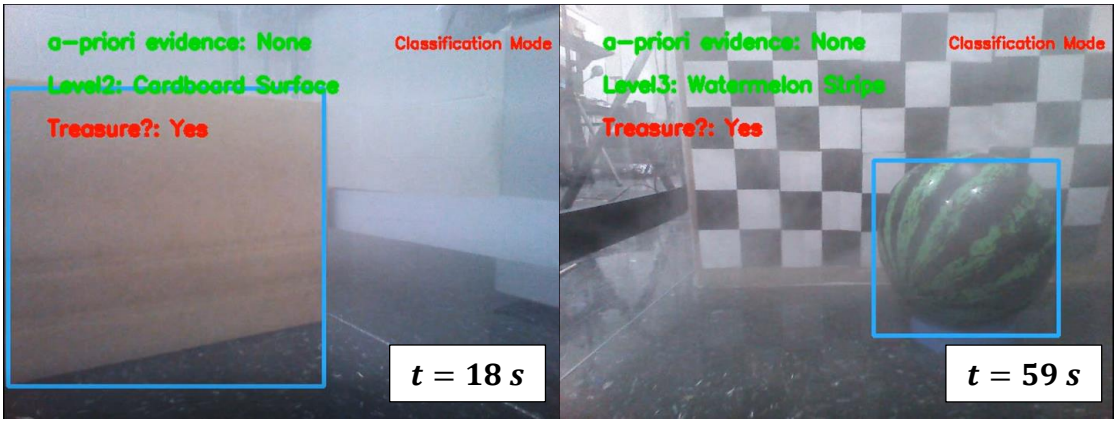


Figure 7.14: Target visitation sequence of AdaptiveSwitch in the third workspace.

CHAPTER 8

CONCLUSION

The research aims to develop a decision making toolbox that is able to handle different pressure conditions and make high quality decisions. In addition to the optimization based decision making approaches, the author embraces the idea of satisficing proposed by Herbert Simon, and tries to answer a question that Simon raised "How do human beings reason when the conditions for rationality postulated by the model of neoclassical economics are not met?"

To answer this question, the author and the collaborators considers a benchmark decision making problem that involves both action and test decision under four representative environmental pressures, which make the assumptions for the global rationality no longer hold. The four considered pressure conditions are: time pressure, information cost pressure, training data scarcity pressure and sensory deprivation pressure. The author designs passive and active satisficing experiments based on the proposed benchmark decision problem, and invites human participants to conduct the experiments and solve the decision problem under the aforementioned pressures. To understand the human decision mechanism under pressures and how the pressures influence the human participant's decision behavior, the author uses various statistical tools, such as hypothesis testings, dynamic Bayesian Network and inverse reinforcement learning algorithms to fit the data and observes how model parameters change due to the existence of the pressures. Specifically, for the pressure condition under which the decision data can't be well explained by current off-the-shelf statistical tools or algorithms, the raw decision data is manually inspected and the decision behavior pattern is summarized.

Based on the data analysis results, the author proposes the most plausible de-

cision models that human participants use under each pressure condition. Among those, the author manages to model the human decisions under time pressure, limited data pressure and sensory deprivation pressures as heuristics, which can not only complete the decision making tasks that optimization based strategies are not feasible, but also outperform optimization based strategies even if they work to some extent. Several mathematically properties of the heuristics under time pressure have been proved and show that as the pressure becomes more severe, the more cue are dropped to make a decision. The decision problem under information cost pressure, however, is solved under optimization framework because the pressure can be described well mathematically. Particularly, the human heuristics modelled under sensory deprivation pressure, as the most important heuristics in this dissertation, is extracted based on the inspection of the human decision data and tested extensively on robotic applications with different levels of uncertainties from simulations to physical experiments. The modeled strategy manages to complete the "treasure hunt" task with handling the uncertainties brought by physical world quite well: the heuristic strategy is able to guide a four-wheeled robot navigate and find treasure in a fog environment which makes the camera's visible range extremely short. It is also worth to note that the physical experiments on the robot brings an extra layer of uncertainty that the human experiments don't have, which is the target detection false alarm and miss detection. In this sense, the "treasure hunt" experiments for a robot is more challenging than for the human participants.

In terms of practical applications of this research, the author foresees potential applications of the inference heuristics on highly structured problems requiring quick decisions. For the heuristic strategies under interrupted sensory signal pressure, the author expects that a mobile robotic sensor with the algorithm can be applied in a realistic and unstructured environment, like victim rescue after disas-

ters or target search under adverse weather such as fog or heavy rain environments.

APPENDIX A

PROPERTIES OF HEURISTICS UNDER TIME PRESSURE AND PROOFS

A.1 Discounted Cumulative Probability Gain (ProbGain)

The output of $H_{\text{ProbGain}}(t_b, \{m_i\}_{i=1}^{\wp})$ is the number of most informative cues to use given the discount due to time pressure. The discount factor $\gamma(t_b)$ decreases as the allowable time t_b decreases, λ can be understood as a parameter that human participants use to control the discount given a time pressure.

The idea of H_{ProbGain} is as follows: as more cues are used, the probability gain accumulates; in the meantime, as the probability gain of one more cue is added, one more discount factor due to time pressure is multiplied to the cumulative probability gain. Then, the the number of cues to use is determined by maximizing the product of discount factors and the cumulative probability gain to make the trade off between the time pressure and the cue probability gain.

Proposition 1. A sufficient condition for *ProbGain* to use all $\wp$ cues is: if the allowable time t_b to make a decision satisfies:

$$t_b \geq \frac{\lambda \wp}{\ln(1 + \frac{\alpha}{\wp})} \tag{A.1}$$

where $\alpha = \frac{v(m_{\mathbf{c}_\wp})}{v(m_{\mathbf{c}_1})}$ is ratio of information values between the least informative cue and the most informative cue.

Proof: For an arbitrary classification task with cues $\{m_i\}_{i=1}^{\wp}$, according the heuristic represented by Eq. 4.6, the allowable time to make a classification decision should satisfy following inequalities:

$$\gamma(t_b)^{\wp}\sum_{i=1}^{\wp} v(m_{\mathbf{c}_i}) \geq \gamma(t_b)^m \sum_{i=1}^{m} v(m_{\mathbf{c}_i}), \quad 1 \leq m \leq \wp - 1 \tag{A.2}$$

Simplify the inequalities and we have:

$$\gamma(t_b) \geq (\frac{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})}{\sum_{i=1}^{\wp} v(m_{\mathbf{c}_i})})^{\frac{1}{\wp - m}}, \quad 1 \leq m \leq \wp - 1 \tag{A.3}$$

Plug Eq. 4.7 in Eq.A.3, then we have:

$$e^{-\frac{\lambda}{t_b}} \geq (\frac{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})}{\sum_{i=1}^{\wp} v(m_{\mathbf{c}_i})})^{\frac{1}{\wp - m}}, \quad 1 \leq m \leq \wp - 1 \tag{A.4}$$

Compute the log of both sides of Eq. A.4 and simplify the equation, we have

$$t_b \geq \frac{\lambda}{\ln(1 + \frac{\sum_{i=m+1}^{\wp} v(m_{\mathbf{c}_i})}{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})})^{\frac{1}{\wp - m}}}, \quad 1 \leq m \leq \wp - 1 \tag{A.5}$$

Consider the cues $m_{\mathbf{c}_1}, m_{\mathbf{c}_2}, , ..., m_{\mathbf{c}_\wp}$ are sorted according to their information values, i.e. $v(m_{\mathbf{c}_1}) \geq v(m_{\mathbf{c}_2}) \geq ... \geq v(m_{\mathbf{c}_\wp})$.

For $m = 1, 2, ..., \wp - 1$

$$\begin{aligned} \ln(1 + \frac{\sum_{i=m+1}^{\wp} v(m_{\mathbf{c}_i})}{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})})^{\frac{1}{\wp - m}} &\geq \ln(1 + \frac{v(m_{\mathbf{c}_\wp})}{\wp v(m_{\mathbf{c}_1})})^{\frac{1}{\wp - m}} \\ &\geq \ln(1 + \frac{v(m_{\mathbf{c}_\wp})}{\wp v(m_{\mathbf{c}_1})})^{\frac{1}{\wp}} \\ &= \frac{1}{\wp}\ln(1 + \frac{v(m_{\mathbf{c}_\wp})}{\wp v(m_{\mathbf{c}_1})}) \end{aligned} \tag{A.6}$$

Plug Eq. A.6 in Eq. A.5, and we have a sufficient condition for the heuristic to cue all cues:

$$t_b \geq \frac{\lambda \wp}{\ln(1 + \frac{\alpha}{\wp})} \tag{A.7}$$

where $\alpha = v(m_{\mathbf{c}_\wp})/v(m_{\mathbf{c}_1})$.

Proposition 2. A sufficient condition for *ProbGain* to use 1 (the least possible number of cues to use) cue is: if the allowable time to make a decision t_b satisfies

$$t_b \leq \frac{\lambda}{\ln(\wp)} \tag{A.8}$$

Proof: For an arbitrary classification task with cues $\{m_i\}_{i=1}^{\wp}$, according the heuristic represented by Eq.4.6, the allowable time t_b to make a classification decision should satisfy following inequalities:

$$\gamma(t_b)v(m_{\mathbf{c}_1}) \geq \gamma(t_b)^m \sum_{i=1}^{m} v(m_{\mathbf{c}_i}), \quad 2 \leq m \leq \wp \tag{A.9}$$

Simplify the inequalities and we have:

$$\gamma(t_b) \leq (\frac{v(m_{\mathbf{c}_1})}{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})})^{\frac{1}{m-1}}, \quad 2 \leq m \leq \wp \tag{A.10}$$

Plug Eq. 4.7 in Eq. A.10, then we have:

$$e^{-\frac{\lambda}{t_b}} \leq (\frac{v(m_{\mathbf{c}_1})}{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})})^{\frac{1}{m-1}}, \quad 2 \leq m \leq \wp \tag{A.11}$$

Compute the log of both sides of Eq. 14 and simplify the equation, we have

$$t_b \leq \frac{\lambda}{\frac{1}{m-1}\ln(\frac{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})}{v(m_{\mathbf{c}_1})})} \tag{A.12}$$

Consider the cues $m_{\mathbf{c}_1}, m_{\mathbf{c}_2}, , ..., m_{\mathbf{c}_M}$ are sorted according to their information values, i.e. $v(m_{\mathbf{c}_1}) \geq v(m_{\mathbf{c}_2}) \geq ... \geq v(m_{\mathbf{c}_\wp})$.

For $m = 2, 3, ..., \wp$

$$\begin{aligned} \frac{1}{m-1}\ln(\frac{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})}{v(m_{\mathbf{c}_1})}) &\leq \ln(\frac{\sum_{i=1}^{m} v(m_{\mathbf{c}_i})}{v(m_{\mathbf{c}_1})}) \\ &\leq \ln(\frac{\wp v(m_{\mathbf{c}_1})}{v(m_{\mathbf{c}_1})}) \\ &= \ln(\wp) \end{aligned} \tag{A.13}$$

Plug Eq.A.13 in Eq. A.12, and we have a sufficient condition for the heuristic to cue 1 cue:

$$t_b \leq \frac{\lambda}{\ln(\wp)} \tag{A.14}$$

Proposition 3. Monotonicity with respect to allowable time t_b: for a classification task with cues $\{m_i\}_{i=1}^{\wp}$, $H_{\text{ProbGain}}(t_b, \{m_i\}_{i=1}^{\wp})$ satisfies:

$$H_{\text{ProbGain}}(t_{b,2}, \{m_i\}_{i=1}^{\wp}) \geq H_{\text{ProbGain}}(t_{b,1}, \{m_i\}_{i=1}^{\wp}) \tag{A.15}$$

for $\forall t_{b,1}, t_{b,2},\ t_{b,2} > t_{b,1}$

Proof by contradiction: Denote $n_1 = H_{\text{ProbGain}}(t_{b,1}, \{m_i\}_{i=1}^{\wp})$, $n_2 = H_{\text{ProbGain}}(t_{b,2}, \{m_i\}_{i=1}^{\wp})$. Suppose $n_1 > n_2$, then according to the heuristic, we have the following two inequalities:

$$\gamma(t_{b,2})^{n_2} \sum_{i=1}^{n_2} v(m_{\mathbf{c}_i}) \geq \gamma(t_{b,2})^{n_1} \sum_{i=1}^{n_1} v(m_{\mathbf{c}_i}) \tag{A.16}$$

$$\gamma(t_{b,1})^{n_2} \sum_{i=1}^{n_2} v(m_{\mathbf{c}_i}) \leq \gamma(t_{b,1})^{n_1} \sum_{i=1}^{n_1} v(m_{\mathbf{c}_i}) \tag{A.17}$$

Simplify Eq. A.16, A.17, we have:

$$\gamma(t_{b,2})^{n_1-n_2} \leq \frac{\sum_{i=1}^{n_2} v(m_{\mathbf{c}_i})}{\sum_{i=1}^{n_1} v(m_{\mathbf{c}_i})} \tag{A.18}$$

$$\gamma(t_{b,1})^{n_1-n_2} \geq \frac{\sum_{i=1}^{n_2} v(m_{\mathbf{c}_i})}{\sum_{i=1}^{n_1} v(m_{\mathbf{c}_i})} \tag{A.19}$$

According to Eq. A.18, A.19, we have

$$\gamma(t_{b,1})^{n_1-n_2} \geq \gamma(t_{b,2})^{n_1-n_2} \tag{A.20}$$

Since we assume $n_1 > n_2$, we have:

$$\gamma(t_{b,1}) \geq \gamma(t_{b,2}) \tag{A.21}$$

This result contradicts to the fact that given $t_{b,2} > t_{b,1}$, according to the definition of $\gamma(t_b)$, $\gamma(t_{b,2}) > \gamma(t_{b,1})$. Then the assumption $n_1 > n_2$ is incorrect. Thus, for $\forall t_{b,1}, t_{b,2},\ t_{b,2} > t_{b,1}$, we have

$$H_{\text{ProbGain}}(t_{b,2}, \{m_i\}_{i=1}^{\wp}) \geq H_{\text{ProbGain}}(t_{b,1}, \{m_i\}_{i=1}^{\wp}) \tag{A.22}$$

Propositions 1 and 2 tell the behavior of *ProbGain* under "extreme" conditions. Particularly, proposition 1 shows that as the allowable time $t_b \geq \frac{\lambda\varphi}{\ln(1+\frac{\alpha}{\varphi})}$, the heuristic uses all cues to make the classification decision (i.e. converges to the "optimal strategy", which uses all cues to make a decision). Also, according to Proposition 2. when the allowable time is too little ($t_b \leq \frac{\lambda}{\ln(\varphi)}$), the heuristic only uses 1 cue (the least possible number of cues to use) to make the decision. Proposition 3 shows the monotonicity of the heuristic with respect to allowable time t_b, as allowable time increases, the heuristic uses monotonically more cues to make a classification decision.

A.2 Discounted Log-odds Ratio (LogOdds)

This heuristic regards log-odds ratio,

$$L_{M_i} = log\frac{p(Y = y_1|m_{\mathbf{c}_1}, ..., m_{\mathbf{c}_i})}{p(Y = y_2|m_{\mathbf{c}_1}, ..., m_{\mathbf{c}_i})}$$

based on cues in set M_i as the "confidence" of making the classification task. The greater the value $|L_{M_i}|$ is, the more confident is to make the classification decision. While one cue comes into consideration, an additional time-pressure dependent discount factor is imposed on the absolute value the log-odds ratio L_{M_i} of the cues in set M_i. The heuristic selects the cues under pressure based on the maximization of the product of the discount factors and the log-odds ratio, in this way, the less informative cues will be dropped due to the discount factor. As the time pressure increases, the heuristic has larger tendency to drop the cues.

Proposition 4. A sufficient condition for *LogOdds* to use 1 cue is if the allowable time t_b to make a decision satisfies

$$t_b \leq \frac{\lambda}{\ln(1 + \frac{\wp - 1}{|1+\beta|})} \tag{A.23}$$

where $\beta = v_0/v(m_{\mathbf{c}_1})$.

Proof: For an arbitrary classification task with cues $\{m_i\}_{i=1}^{\wp}$, according the heuristic represented by Eq. 4.8, the allowable time to make a classification decision should satisfy following

$$\gamma(t_b)|v_0 + v(m_{\mathbf{c}_1})| \geq \gamma(t_b)^m |v_0 + \sum_{i=1}^{m} v(m_{\mathbf{c}_i})|, \quad 2 \leq m \leq \wp \tag{A.24}$$

Simplify the inequalities and we have:

$$e^{-\frac{\lambda}{t_b}} \leq (\frac{|v_0 + v(m_{\mathbf{c}_1})|}{|v_0 + \sum_{i=1}^{m} v(m_{\mathbf{c}_i})|})^{\frac{1}{m-1}}, \quad 2 \leq m \leq \wp \tag{A.25}$$

If $\exists m$, s.t.

$$|v_0 + v(m_{\mathbf{c}_1})| > |v_0 + \sum_{i=1}^{m} v(m_{\mathbf{c}_i})| \tag{A.26}$$

then the Eq. A.25 is automatically satisfied. Otherwise, we have

$$t_b \leq \frac{\lambda}{\frac{1}{m-1} \ln(\frac{|v_0 + \sum_{i=1}^{m} v(m_{\mathbf{c}_i})|)}{|v_0 + v(m_{\mathbf{c}_1})|)})} \tag{A.27}$$

For $m = 2, 3, ..., \wp$

$$\begin{aligned}\frac{1}{m-1}\ln(\frac{|v_0+\sum_{i=1}^{m} v(m_{\mathbf{c}_i})|)}{|v_0+v(m_{\mathbf{c}_1})|)} &\leq \ln(\frac{|v_0+v(m_{\mathbf{c}_1})|+\sum_{i=2}^{m}|v(m_{\mathbf{c}_i})|)}{|v_0+v(m_{\mathbf{c}_1})|)} \\ &\leq \ln(1+\frac{(\wp-1)|v(m_{\mathbf{c}_1})|}{|v_0+v(m_{\mathbf{c}_1})|}) \\ &\leq \ln(1+\frac{\wp-1}{|1+\beta|})\end{aligned} \tag{A.28}$$

where $\beta = \frac{v_0}{v(m_{\mathbf{c}_1})}$.

Thus, a sufficient condition for *LogOdds* to use only 1 cue is that

$$t_b \leq \frac{\lambda}{\ln(1+\frac{\wp-1}{|1+\beta|})} \tag{A.29}$$

Proposition 5. Monotonicity with respect to allowable time t_b: for an object with cues $\{m_i\}_{i=1}^{\wp}$, $H_{\text{LogOdds}}(t_b, \{m_i\}_{i=1}^{\wp})$ satisfies:

$$H_{\text{LogOdds}}(t_{b,2}, \{m_i\}_{i=1}^{\wp}) \geq H_{\text{LogOdds}}(t_{b,1}, \{m_i\}_{i=1}^{\wp}) \tag{A.30}$$

for $\forall t_{b,1}, t_{b,2},\ t_{b,2} > t_{b,1}$.

Proof by Contradiction: Denote $n_1 = H_{\text{LogOdds}}(t_{b,1}, \{m_i\}_{i=1}^{\wp})$, $n_2 = H_{\text{LogOdds}}(t_{b,2}, \{m_i\}_{i=1}^{\wp})$. Suppose $n_1 > n_2$, then according to the heuristic, we have the following two inequalities:

$$\gamma(t_{b,2})^{n_2}|v_0+\sum_{i=1}^{n_2} v(m_{\mathbf{c}_i})| \geq \gamma(t_{b,2})^{n_1}|v_0+\sum_{i=1}^{n_1} v(m_{\mathbf{c}_i})| \tag{A.31}$$

$$\gamma(t_{b,1})^{n_2}|v_0+\sum_{i=1}^{n_2} v(m_{\mathbf{c}_i})| \leq \gamma(t_{b,1})^{n_1}|v_0+\sum_{i=1}^{n_1} v(m_{\mathbf{c}_i})| \tag{A.32}$$

Simplify Eq. A.31, A.32, we have:

$$\gamma(t_{b,2})^{n_1-n_2} \leq \frac{|v_0 + \sum_{i=1}^{n_2} v(m_{\mathbf{c}_i})|}{|v_0 + \sum_{i=1}^{n_1} v(m_{\mathbf{c}_i})|} \tag{A.33}$$

$$\gamma(t_{b,1})^{n_1-n_2} \geq \frac{|v_0 + \sum_{i=1}^{n_2} v(m_{\mathbf{c}_i})|}{|v_0 + \sum_{i=1}^{n_1} v(m_{\mathbf{c}_i})|} \tag{A.34}$$

According to Eq.A.33, A.34, we have

$$\gamma(t_{b,1})^{n_1-n_2} \geq \gamma(t_{b,2})^{n_1-n_2} \tag{A.35}$$

Since we assume $n_1 > n_2$, we have:

$$\gamma(t_{b,1}) \geq \gamma(t_{b,2}) \tag{A.36}$$

This result contradicts to the fact that given $t_{b,2} > t_{b,1}$, according to the definition of $\gamma(t_b)$, $\gamma(t_{b,2}) > \gamma(t_{b,1})$.

Then the assumption $n_1 > n_2$ is incorrect. Thus, for $\forall t_{b,1}, t_{b,2},\ t_{b,2} > t_{b,1}$, we have

$$H_{\text{LogOdds}}(t_{b,2}, \{m_i\}_{i=1}^{\wp}) \geq H_{\text{LogOdds}}(t_{b,1}, \{m_i\}_{i=1}^{\wp}) \tag{A.37}$$

Note that unlike H_{ProbGain}, although H_{LogOdds} tends to use more cues as time pressure release, H_{LogOdds} doesn't necessarily use all $\wp$ cues when the time available t_b is greater than a certain threshold. This is because the value metric used in

H_{LogOdds}: $|L_{M_i}| = |v_0 + \sum_{j=1}^{i} v(m_{\mathbf{c}_j})|$ is not monotonically increasing as the number of cues to use i increases.

A.3 Information Free Cue Number Discounting (InfoFree)

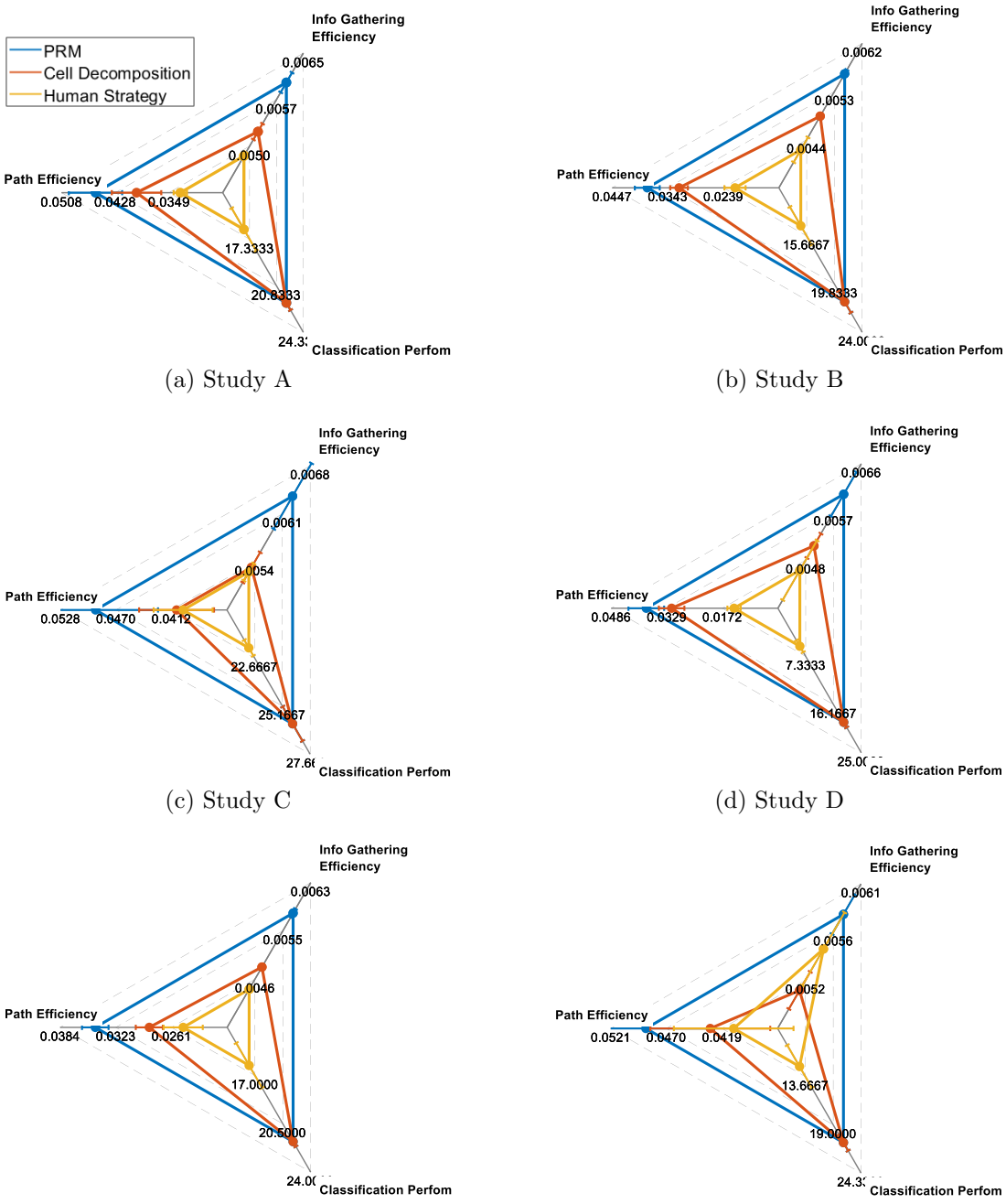
After sorting the cues in terms of the information gain, the cut-off criterion of this heuristic is no longer dependent on the information gain. Thus the allowable decision time t_b is the only argument for the heuristic. As $\exp(-\lambda/t_b) < 1, t_b \in (0, +\infty)$, the number of cues to use is always less than or equal to $\wp$ and decreases exponentially when time pressure increases, and the parameter $\lambda > 0$ controls how much a time pressure is discounted. Given the monotonicity of the exponential function, it is obvious that H_{InfoFree} uses more cues as time pressure releases and it uses all $\wp$ cues if the time available t_b is large enough and uses 1 cue if time available t_b is small enough.

APPENDIX B

NO PRESSURE PLANNING RESULTS

The Active Satisficing Experiment includes no pressure condition besides money pressure and fog pressure, which are deeply addressed in the dissertation. To compare with human performance under no pressure condition, two graph search based optimal strategies (PRM based and Cell Decomposition based) have been developed. When there is no pressure imposed, the optimal strategies reduce to Traveling Salesman Problem (TSP) [2, 54, 38, 37] that tries to find the shortest path that covers all targets while makes all target features measurements.

Thus, we focus on the path efficiency $1/D(\tau)$, the info gathering efficiency $B(\tau)/D(\tau)$ and, finally, classification performance $N(\tau)$. The optimal strategies outperform the human strategy in the aforementioned three aspects, except Study F, where human strategy has higher Info Gathering efficiency than Cell decomposition but not PRM. One study comprises of three different experiment trials. The optimal strategies and the human subjects don't have prior knowledge of the target positions and initial features, and the information is obtained from an onboard directional sensor.



(a) Study A

(b) Study B

(c) Study C

(d) Study D

(e) Study E

(f) Study F

Figure B.1: Performance comparison between human strategy and optimal strategies under no pressure condition.

APPENDIX C

FAST AND FRUGAL TREE AS INFERENCE STRATEGY

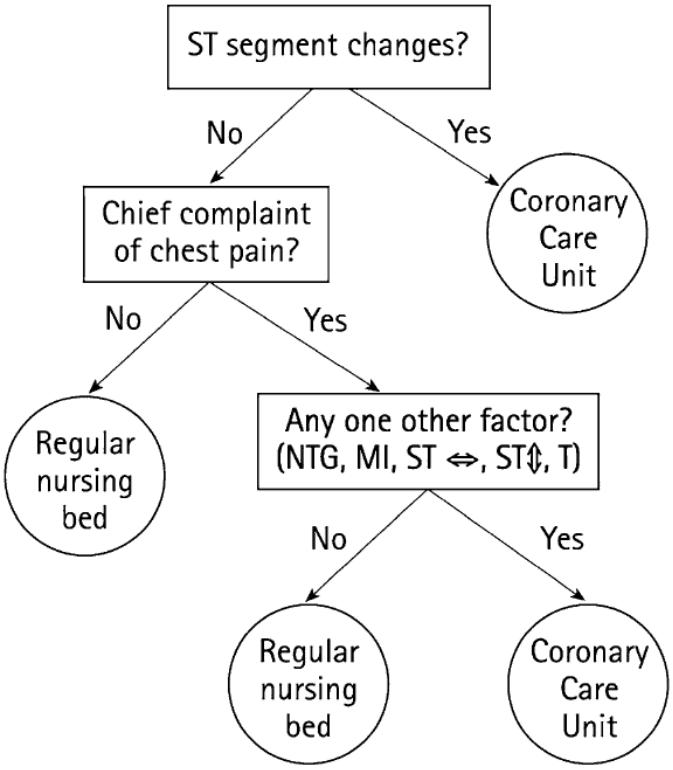


Figure C.1: A quick procedure for physicians to decide whether a patient has acute ischemic heart disease.

The target feature measurement strategy under fog pressure is modeled as a Fast Frugal Tree (FFT)[31, 57, 36, 58]. This is a inference decision heuristic model that is able to handle sequential cues. Suppose there is a stimulus with m binary cues. If a Bayesian model is used and incorporates all m cues, then the decision model has 2^m leaves in the representation of a tree. In contrast, FFTs have only $m+1$ leaves as the model doesn't wait until all cues are revealed to make a decision. As each cue is revealed, the decision is potentially to be made. If not, the next cue will be revealed to obtain more information. The simplicity of FFTs makes the likely to be robust to wider ranges of data sets due to Less-Can-Be-More effect. The construction of FFTs usually embedded with expert/human knowledge as

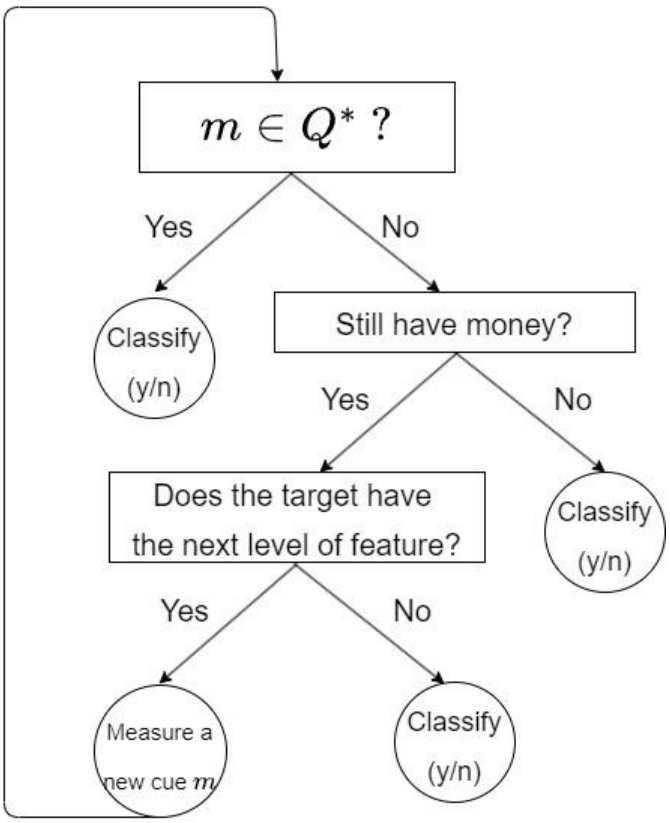


Figure C.2: The FFT structure that models human participants cue measurement strategy under fog environment.

the cue order in the tree can affect not only the computation complexity but also inference accuracy. A common example of FFT as shown in Fig. C.1 is constructed by Green et.al [36] to make a medical decision.

The structure of the FFT in the target feature measurement under fog pressure condition is shown as Fig. C.2. The set of cues that human participants deem to have high enough confidence to exit the cue measurement process is denoted as Q^*.

BIBLIOGRAPHY

[1] Pieter Abbeel and Andrew Y Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, page 1, 2004.

[2] Mandell Bellmore and George L Nemhauser. The traveling salesman problem: a survey. *Operations Research*, 16(3):538–558, 1968.

[3] Dimitri Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 1. Athena scientific, 2012.

[4] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[5] Abdeslam Boularias, Jens Kober, and Jan Peters. Relative entropy inverse reinforcement learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 182–189. JMLR Workshop and Conference Proceedings, 2011.

[6] Henry Brighton and Gerd Gigerenzer. Bayesian brains and cognitive mechanisms: Harmony or dissonance. *The probabilistic mind: Prospects for Bayesian cognitive science, ed. N. Chater & M. Oaksford*, pages 189–208, 2008.

[7] Arndt Bröder. Decision making with the" adaptive toolbox": influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4):611, 2003.

[8] Chenghui Cai and S. Ferrari. On the development of an intelligent computer player for clue: a case study on preposterior decision analysis. In *2006 American Control Conference*, pages 4350–4355, 2006.

[9] Chenghui Cai and Silvia Ferrari. Comparison of information-theoretic objective functions for decision support in sensor systems. In *2007 American Control Conference*, pages 3559–3564, 2007.

[10] Chenghui Cai and Silvia Ferrari. A q-learning approach to developing an automated neural computer player for the board game of clue®. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2346–2352. IEEE, 2008.

[11] Chenghui Cai and Silvia Ferrari. Information-driven sensor path planning by approximate cell decomposition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(3):672–689, 2009.

[12] Chenghui Cai, Silvia Ferrari, and Ming Qian. Bayesian network modeling of acoustic sensor measurements. In *SENSORS, 2007 IEEE*, pages 345–348, 2007.

[13] Andrew Caplin and Paul W Glimcher. Basic methods from neoclassical economics. In *Neuroeconomics*, pages 3–17. Elsevier, 2014.

[14] George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.

[15] Yucheng Chen. Navigation in fog, 2021.

[16] Paul Cisek, Geneviève Aude Puskas, and Stephany El-Murr. Decisions in changing conditions: the urgency-gating model. *Journal of Neuroscience*, 29(37):11560–11571, 2009.

[17] Anja Dieckmann and Jörg Rieskamp. The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, 35(7):1801–1813, 2007.

[18] DiVE. The duke immersive virtual environment (dive). `https://https://digitalhumanities.duke.edu/resource/duke-immersive-virtual-environment-dive`, 2006.

[19] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.

[20] Silvia Ferrari and Chenghui Cai. Information-driven search strategies in the board game of clue®. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(3):607–625, 2009.

[21] Silvia Ferrari and Thomas A Wettergren. *Information-Driven Planning and Control*. MIT Press, 2021.

[22] Peter C Fishburn. Subjective expected utility: A review of normative theories. *Theory and decision*, 13(2):139–199, 1981.

[23] Frédérick Garcia and Emmanuel Rachelson. Markov decision processes. *Markov Decision Processes in Artificial Intelligence*, pages 1–38, 2013.

[24] David Garlan, Daniel P Siewiorek, Asim Smailagic, and Peter Steenkiste. Project aura: Toward distraction-free pervasive computing. *IEEE Pervasive computing*, 1(2):22–31, 2002.

[25] Shuzhi Sam Ge, Qun Zhang, Aswin Thomas Abraham, and Brice Rebsamen. Simultaneous path planning and topological mapping (sp2atm) for environment exploration and goal oriented navigation. *Robotics and Autonomous Systems*, 59(3-4):228–242, 2011.

[26] Jake Gemerek, Bo Fu, Yucheng Chen, Zeyu Liu, Min Zheng, David van Wijk, and Silvia Ferrari. Directional sensor planning for occlusion avoidance. *IEEE Transactions on Robotics*, pages 1–21, 2022.

[27] Zoubin Ghahramani. Learning dynamic bayesian networks. *Adaptive Processing of Sequences and Data Structures: International Summer School on Neural Networks "ER Caianiello" Vietri sul Mare, Salerno, Italy September 6–13, 1997 Tutorial Lectures*, pages 168–197, 2006.

[28] Gerd Gigerenzer. From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological review*, 98(2):254, 1991.

[29] Gerd Gigerenzer. *Gut feelings: The intelligence of the unconscious*. Penguin, 2007.

[30] Gerd Gigerenzer and Henry Brighton. Homo heuristicus: Why biased minds make better inferences. *Topics in cognitive science*, 1(1):107–143, 2009.

[31] Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic decision making. *Annual review of psychology*, 62(1):451–482, 2011.

[32] Gerd Gigerenzer and Daniel G Goldstein. Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4):650, 1996.

[33] Gerd Gigerenzer and Peter M Todd. *Simple heuristics that make us smart*. Oxford University Press, USA, 1999.

[34] Mark A Gluck, Daphna Shohamy, and Catherine Myers. How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory*, 9(6):408–418, 2002.

[35] Daniel G Goldstein and Gerd Gigerenzer. Models of ecological rationality: the recognition heuristic. *Psychological review*, 109(1):75, 2002.

[36] Lee Green and David R Mehr. What alters physicians' decisions to admit to the coronary care unit? *Journal of family practice*, 45(3):219–227, 1997.

[37] Gregory Gutin and Abraham P Punnen. *The traveling salesman problem and its variations*, volume 12. Springer Science & Business Media, 2006.

[38] Karla L Hoffman, Manfred Padberg, Giovanni Rinaldi, et al. Traveling salesman problem. *Encyclopedia of operations research and management science*, 1:1573–1578, 2013.

[39] Robin M Hogarth and Natalia Karelaia. Heuristic and linear models of judgment: matching rules and environments. *Psychological review*, 114(3):733, 2007.

[40] Husarion. Rosbot autonomous mobile robot. `https://husarion.com/manuals/rosbot/`, 2018.

[41] Finn V Jensen and Thomas Dyhre Nielsen. *Bayesian networks and decision graphs*, volume 2. Springer, 2007.

[42] Mrinal Kalakrishnan, Peter Pastor, Ludovic Righetti, and Stefan Schaal. Learning objective functions for manipulation. In *2013 IEEE International Conference on Robotics and Automation*, pages 1331–1336. IEEE, 2013.

[43] Mike Knicker. 3d scanning basics: How structured light scanning works, 2014.

[44] John K Kruschke. Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(5):658–676, 2010.

[45] David A Lagnado, Ben R Newell, Steven Kahan, and David R Shanks. Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, 135(2):162, 2006.

[46] Koen Lamberts. Categorization under time pressure. *Journal of Experimental Psychology: General*, 124(2):161, 1995.

[47] Jean-Claude Latombe. *Robot motion planning*, volume 124. Springer Science & Business Media, 2012.

[48] Jean-Claude Latombe. *Robot motion planning*, volume 124. Springer Science & Business Media, 2012.

[49] Steven M LaValle. *Planning algorithms*. Cambridge university press, 2006.

[50] Nilli Lavie. Attention, distraction, and cognitive control under load. *Current directions in psychological science*, 19(3):143–148, 2010.

[51] Mikhail A Lebedev, Jose M Carmena, Joseph E O'Doherty, Miriam Zacksenhouse, Craig S Henriquez, Jose C Principe, and Miguel AL Nicolelis. Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface. *Journal of Neuroscience*, 25(19):4681–4693, 2005.

[52] Sergey Levine, Zoran Popovic, and Vladlen Koltun. Nonlinear inverse reinforcement learning with gaussian processes. *Advances in neural information processing systems*, 24, 2011.

[53] Allan J Lichtman. *The keys to the White House: a surefire guide to predicting the next president*. Rowman & Littlefield, 2008.

[54] Shen Lin. Computer solutions of the traveling salesman problem. *Bell System Technical Journal*, 44(10):2245–2269, 1965.

[55] Chun Liu, Shuhang Zhang, and Akram Akbar. Ground feature oriented path planning for unmanned aerial vehicle mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(4):1175–1187, 2019.

[56] Wenjie Lu, Guoxian Zhang, and Silvia Ferrari. An information potential approach to integrated sensor path planning and control. *IEEE Transactions on Robotics*, 30(4):919–934, 2014.

[57] Laura Martignon, Konstantinos V Katsikopoulos, and Jan K Woike. Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology*, 52(6):352–361, 2008.

[58] Laura Martignon, Oliver Vitouch, Masanori Takezawa, and Malcolm R Forster. Naive and yet enlightened: From natural frequencies to fast and frugal decision trees. *Thinking: Psychological perspective on reasoning, judgment, and decision making*, pages 189–211, 2003.

[59] Sendhil Mullainathan and Richard H Thaler. Behavioral economics, 2000.

[60] Ben R Newell and David R Shanks. Take the best or look at the rest? factors

influencing" one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1):53, 2003.

[61] Andrew Y Ng, Stuart Russell, et al. Algorithms for inverse reinforcement learning. In *Icml*, volume 1, page 2, 2000.

[62] Phedon Nicolaides. Limits to the expansion of neoclassical economics. *Cambridge Journal of Economics*, 12(3):313–328, 1988.

[63] Hanna Oh, Jeffrey M Beck, Pingping Zhu, Marc A Sommer, Silvia Ferrari, and Tobias Egner. Satisficing in split-second decision making is characterized by strategic cue discounting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(12):1937, 2016.

[64] Thorsten Pachur and Ralph Hertwig. On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5):983, 2006.

[65] Xueni Pan and Antonia F de C Hamilton. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3):395–417, 2018.

[66] John W Payne, James R Bettman, and Eric J Johnson. Adaptive strategy selection in decision making. *Journal of experimental psychology: Learning, Memory, and Cognition*, 14(3):534, 1988.

[67] Anthony J Porcelli and Mauricio R Delgado. Stress and decision making: effects on valuation, learning, and risk-taking. *Current opinion in behavioral sciences*, 14:33–39, 2017.

[68] Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.

[69] Martin L Puterman. Markov decision processes. *Handbooks in operations research and management science*, 2:331–434, 1990.

[70] Jaime Quintana Benito, Antonio Álvarez Fernández-Balbuena, Juan Carlos Martínez-Antón, and Daniel Vázquez Molini. Improvement of driver night vision in foggy environments by structured light projection. *Available at SSRN 4146366*.

[71] Roger Ratcliff and Gail McKoon. Similarity information versus relational

information: Differences in the time course of retrieval. *Cognitive Psychology*, 21(2):139–155, 1989.

[72] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *CoRR*, abs/1804.02767, 2018.

[73] Jörg Rieskamp and Philipp E Otto. Ssl: a theory of how people learn to select strategies. *Journal of experimental psychology: General*, 135(2):207, 2006.

[74] Nicolas Bono Rossello, Renzo Fabrizio Carpio, Andrea Gasparri, and Emanuele Garone. Information-driven path planning for uav with limited autonomy in large-scale field monitoring. *IEEE Transactions on Automation Science and Engineering*, 19(3):2450–2460, 2021.

[75] Stuart J Russell. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010.

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

[77] Stephen H Scott. Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, 5(7):532–545, 2004.

[78] Jean-Christophe Servotte, Manon Goosse, Suzanne Hetzell Campbell, Nadia Dardenne, Bruno Pilote, Ivan L Simoneau, Michèle Guillaume, Isabelle Bragard, and Alexandre Ghuysen. Virtual reality experience: Immersion, sense of presence, and cybersickness. *Clinical Simulation in Nursing*, 38:35–43, 2020.

[79] Jennie Si, Andrew G Barto, Warren B Powell, and Don Wunsch. *Handbook of learning and approximate dynamic programming*, volume 2. John Wiley & Sons, 2004.

[80] Herbert A Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99–118, 1955.

[81] Herbert A Simon. Rational decision making in business organizations. *The American economic review*, 69(4):493–513, 1979.

[82] Herbert A Simon. *The Sciences of the Artificial, reissue of the third edition with a new introduction by John Laird*. MIT press, 2019.

[83] Herbert A Simon and Joseph B Kadane. Optimal problem-solving search: All-or-none solutions. *Artificial Intelligence*, 6(3):235–247, 1975.

[84] Herbert Alexander Simon. *Models of bounded rationality: Empirically grounded economic reason*, volume 3. MIT press, 1997.

[85] Paul Slovic, Ellen Peters, Melissa L Finucane, and Donald G MacGregor. Affect, risk, and decision making. *Health psychology*, 24(4S):S35, 2005.

[86] Maarten Speekenbrink, David A Lagnado, Leonora Wilkinson, Marjan Jahanshahi, and David R Shanks. Models of probabilistic category learning in parkinson's disease: Strategy use and the effects of l-dopa. *Journal of Mathematical Psychology*, 54(1):123–136, 2010.

[87] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[88] Ashleigh Swingler and Silvia Ferrari. On the duality of robot and sensor path planning. In *52nd IEEE Conference on Decision and Control*, pages 984–989. IEEE, 2013.

[89] Emanuel Todorov and Michael I Jordan. Optimal feedback control as a theory of motor coordination. *Nature neuroscience*, 5(11):1226–1235, 2002.

[90] Hendrik AHC Van Veen, Hartwig K Distler, Stephan J Braun, and Heinrich H Bülthoff. Navigating through a virtual city: Using virtual reality technology to study human action and perception. *Future Generation Computer Systems*, 14(3-4):231–242, 1998.

[91] Alexey Vasilyev and Andrei Kapishnikov. Approximation of conditional probability function using artificial neural networks. In *Int. Conference on Modelling and Simulation of Business Systems*, pages 79–81. Citeseer, 2003.

[92] Marco A Wiering and Martijn Van Otterlo. Reinforcement learning. *Adaptation, learning, and optimization*, 12(3):729, 2012.

[93] Guoxian Zhang, Silvia Ferrari, and M Qian. An information roadmap method for robotic sensor path planning. *Journal of Intelligent and Robotic Systems*, 56(1):69–98, 2009.

[94] Pingping Zhu, Silvia Ferrari, Julian Morelli, Richard Linares, and Bryce Doerr. Scalable gas sensing, mapping, and path planning via decentralized hilbert maps. *Sensors*, 19(7), 2019.

[95] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al.

Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.