THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

IN-SILICO BEHAVIOR DISCOVERY SYSTEM

Haibo Wang

Master of Computer Science

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2015

School of Information Technology and Electrical Engineering

Abstract

It is now widely accepted that animal behavior optimizes certain performance criteria. The question is, which criteria are being optimized? To answer this question, today, ethologists start by conducting some exploratory experiments. After the preliminary data is gathered, they infer a set of potential hypotheses. In order to find out which hypothesis is more likely to be correct, they have to carry out real animal experiments for each and every one of the hypotheses. This process could be repetitive, tedious, time consuming and costly.

To alleviate all these difficulties, in this thesis, we propose an *in-silico* system – termed "*in-silico* behavior discovery" – to explore the feasibility of performing an *in silico* study on the underlying performance criteria. This study is distinct from animal dynamics and trajectory simulation as the focus is on the *decision making strategy* behind the observed motions, rather than the specific motions in space. Key to the system is the use of Partially Observable Markov Decision Processes (POMDPs) to generate an optimal strategy under a given hypothesis. POMDPs enable the system take into account imperfect information about the animals' dynamics and the operating environment. Given multiple hypotheses and a set of preliminary observational data, the system will compute the optimal strategy under each hypothesis, generate a set of synthesized data for each optimal strategy, and then rank the hypotheses based on the similarity between the set of synthesized data generated under each hypothesis and the provided observational data. Using this system, the ethologists can focus on the most promising hypotheses first, thus reduce the number of animal experiments.

We evaluate the system using 100 data sets of close encounters between two honeybees. The results are promising, indicating that the system independently identifies the same hypothesis as discovered by ethologists. Moreover, at the sub-criterion level, the system helps clarify the relative weights between contributing factors, even when the ethologists do not have much prior knowledge on the performance criteria. Last but not least, the system is able to discover subtle behaviors that are beyond the discernment of human observers, thereby providing new insights for better understanding animal behaviors.

This thesis is partially comprised of publications, which means some chapters of the thesis are copied from my published papers. In particular, Chapter 2 is copied from the paper [42] published on ACRA 2013; Chapter 3 is copied from the paper [43] published on ICAPS 2015, in which our paper won the Outstanding Student Paper Award.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Haibo Wang

July, 2015

Publications during candidature

- **H. Wang** and K. Liu. User Oriented Trajectory Similarity Search. *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, ACM, 2012.
- **H. Wang**, H. Kurniawati, S. Singh, and M. Srinivasan. Animal Locomotion In Silico: A POMDP-Based Tool to Study Mid-Air Collision Avoidance Strategies in Flying Animals. *Proc. Australian Conference on Robotics and Automation (ACRA)*, 2013.
- **H. Wang**, H. Kurniawati, S. Singh, and M. Srinivasan. In-silico Behavior Discovery System: An Application of Planning in Ethology. *Proc. International Conference on Automated Planning and Scheduling (ICAPS)*, 2015. (Outstanding Student Paper Award)

Publications included in this thesis

- **H. Wang**, H. Kurniawati, S. Singh, and M. Srinivasan. Animal Locomotion In Silico: A POMDP-Based Tool to Study Mid-Air Collision Avoidance Strategies in Flying Animals. *Proc. Australian Conference on Robotics and Automation (ACRA)*, 2013. – Incorporated as Chapter 2.

Contributor	Statement of contribution
Haibo Wang (Candidate)	experimental design (70%) implement the model (80%) perform the experiment (100%) analyse experimental results (70%) write the paper (60%)
Hanna Kurniawati	experimental design (15%) implement the model (20%) analyse experimental results (15%) write the paper (25%)
Surya Singh	experimental design (7.5%) analyse experimental results (7.5%) write the paper (7.5%)
Mandyam Srinivasan	experimental design (7.5%) provide data from animal experiment (100%) analyse experimental results (7.5%) write the paper (7.5%)

- **H. Wang**, H. Kurniawati, S. Singh, and M. Srinivasan. In-silico Behavior Discovery System: An Application of Planning in Ethology. *Proc. International Conference on Automated Planning and Scheduling (ICAPS)*, 2015. – Incorporated as Chapter 3.

Contributor	Statement of contribution
Haibo Wang (Candidate)	experimental design (70%) implement the model (80%) perform the experiment (100%) analyse experimental results (70%) write the paper (60%)
Hanna Kurniawati	experimental design (15%) implement the model (20%) analyse experimental results (15%) write the paper (25%)
Surya Singh	experimental design (7.5%) analyse experimental results (7.5%) write the paper (7.5%)
Mandyam Srinivasan	experimental design (7.5%) provide data from animal experiment (100%) analyse experimental results (7.5%) write the paper (7.5%)

Contributions by others to the thesis

Dr. Hanna Kurniawati, Dr. Surya Singh and Prof. Surya Singh guided me to formulate the problem in Chapter 4 and inspired me to interpret the experimental data.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

First, I want to express my special gratitude to my principal supervisor Dr. Hanna Kurniawati, for being patient and giving me tremendous room to grow. I would like to thank her for helping me so much, especially in those difficult times during my PhD study – without her help, I simply could not get through. I am deeply grateful for her time, dedication and support. I would also like to express my great appreciations to my co-supervisors Dr. Surya Singh and Prof. Mandyam Srinivasan for their professional guidance and insightful discussions. I always appreciate the fact that I can be a PhD student and get directed by them. This is an amazing supervisor team. Without the three wonderful supervisors, this work would never have been possible.

Second, I want to thank the China Scholarship Council for funding me three years on my PhD study in Australia. Also I am truly grateful to my supervisors for providing me scholarships, including top up scholarships.

Third, I want to thank my labmates for having numerous interesting discussions about research, work and life. It has been a great pleasure working with you all.

Last but not least, I want to thank my family, for their continuous support and unconditional love.

Keywords

in silico behaviour discovery, ethology, partially observable markov decision processes, pomdp, planning, optimal strategy, animal behaviour

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 080110, Simulation and Modelling, 75%
ANZSRC code: 069999, Biological Sciences not elsewhere classified, 25%

Fields of Research (FoR) Classification

FoR code: 0801, Artificial Intelligence and Image Processing, 75%
FoR code: 0699, Other Biological Sciences, 25%

Table of contents

List of figures

List of tables

List of abbreviations

POMDP	Partially Observable Markov Decision Process
UAV	Unmanned Aerial Vehicles
TCAS	Traffic Alert and Collision Avoidance System
MCVI	Monte Carlo Value Iteration
MED	Minimum Encounter Distance

Chapter 1

Introduction

1.1 Overview

It is now widely accepted that animal behavior optimizes certain performance criteria [5, 9]. The question is, which criteria are being optimized? Answers to this question may hold the key to significant technical advances. For instance, understanding the landing strategy of honeybees could inspire the design of a vision-based guidance system for the automatic landing of fixed-wing aircraft in unstructured outdoor terrain, just by using onboard video cameras [40], while understanding how honeybees navigate to avoid mid-air collisions could inspire the design of a guiding system that allows a robot progresses along a corridor without colliding with the walls [37].

To answer this question, conventionally, ethologists start by conducting some exploratory experiments. After the preliminary data is gathered, they infer a set of potential hypotheses. In order to find out which hypothesis is more likely to be correct, they have to carry out real animal experiments for each and every one of the hypotheses. Not only could this conventional approach be repetitive, tedious, time consuming and costly, it also faces other challenges such as the need of a large body of observational data to delineate the hypotheses, or the need of novel experimental designs for new hypotheses.

To alleviate all these difficulties, in this thesis, we propose an *in-silico* system – termed "*in-silico* behavior discovery" – to explore the *feasibility of performing an in silico study on the underlying performance criteria*. As a preliminary work, we focus the in silico study on pairwise head-on collision avoidance scenarios, and perform case studies on pairwise head-on collision avoidance encounters of honeybees.

This *in-silico* system is distinct from animal dynamics and trajectory simulation as the focus is on the *decision making strategy* behind the observed motions, rather than the specific motions in space. Key to the system is the use of Partially Observable Markov Decision Processes (POMDPs) to generate an optimal strategy under a given hypothesis. POMDPs enable the system to take into account imperfect information about the animals' dynamics and the operating environment. Given a set of preliminary observational data and *multiple hypotheses*, representing the corresponding *reward functions* in the POMDP model, the system will compute the optimal strategy under each hypothesis, generate a set of synthesized data for each optimal strategy, and then rank the hypotheses based on how well the set of synthesized data generated under each hypothesis *fits* the provided observational data. Using this system, the ethologists can focus on the most promising hypotheses first, thus reducing the number of animal experiments.

We evaluate the system using 100 data sets of close encounters between two honeybees. Experimental results are promising, indicating that the system independently identifies the same hypothesis as discovered by ethologists. Moreover, at a sub-criterion level, the system helps clarify the relative weights between contributing factors, even when the ethologists do not have much prior knowledge on the performance criteria. Last but not least, the system is capable of discovering subtle behaviors that are beyond the discernment of human observers, thereby providing new insights for better understanding animal behaviors.

1.2 Thesis Contributions

In this thesis, we have demonstrated that it is feasible to perform an in silico study of the underlying strategies toward the understanding of animal behaviors, and we have also constructed the *in-silico* behavior discovery system to help ethologists carry out such studies. Specifically, our contributions include:

- We have conceptually designed the *in-silico* behavior discovery system.
- We have carried out the feasibility study of the *in-silico* behavior discovery system. The preliminary results demonstrate that it is feasible to construct the *in-silico* behavior discovery system.
- We have constructed a prototyped *in-silico* behavior discovery system for pairwise head-on collision avoidance scenarios.
- We have developed a set of metrics to measure the similarities between the synthesized data generated by the *in-silico* system and the preliminary observational data of real honeybees.
- We have evaluated the *in-silico* system through real head-on encounter data of honeybees.
- We have verified that the *in-silico* system is capable of independently identifying the most promising hypotheses as discovered by biologists.
- The *in-silico* system can help clarify the relative importance of multiple objectives.
- The *in-silico* system can discover subtle behaviors that are beyond the discernment of human observers.

1.3 Thesis Outline

The following chapters are organized as follows. Chapter 2 performs an feasibility study of the *in-silico* system with a case study on honeybees head-on collision avoidance

scenarios. Chapter 3 builds a prototype of the *in-silico* system – termed "*in-silico* behavior discovery", and tests the system with a set of well understood hypotheses. Chapter 4 extensively evaluates the prototyped *in-silico* behavior discovery system with multiple sets of parameters. The evaluations mainly focus on two systems: the system modeling the *outgoing honeybee* as the rational agent and the system modeling the *incoming honeybee* as the rational agent. A comparative study on the ranking results of the two systems is also carried out.

Chapter 2

In-Silico Behavior Discovery: A Feasibility Study

In this chapter[1], we study the feasibility of an *in-silico* system for studying animal locomotions; specifically, we bring modelling techniques from robotics to enable biologists to perform an *in-silico* study of mid-air collision avoidance strategies of flying animals. This *in-silico* system is distinct from flying animal dynamics and trajectory simulation, as the focus is on the strategy behind the observed motions, rather than the specific motions in space. Our *in-silico* system consists of a model and a simulator. To handle limited data and variations in the flight dynamics and sensing parameters of the animals, we employ a Partially Observable Markov Decision Process (POMDP) framework —a general and principled approach for making decisions under uncertainty. Here, the solution to the POMDP problem is an optimal motion strategy to avoid mid-air collision with another animal. The system simulates the motion strategies in various head-on encounter scenarios. Preliminary results on comparing the simulated behaviours with 100 encounters from real honeybees are promising; the collision rate differs by less than 1%, while the difference in the minimum encounter distance between

[1]This chapter is copied from the follwing publication: H. Wang, H. Kurniawati, S. P. N. Singh, and M. V. Srinivasan. Animal Locomotion In-Silico: A POMDP-Based Tool to Study Mid-Air Collision Avoidance Strategies in Flying Animals. In *Proc. Australasian Conference on Robotics and Automation*, 2013.

two bees in 100 head-on encounters is on average around 12mm, which is roughly equivalent to the average wing span of the honeybees used to generate the data.

2.1 Introduction

Many robotic systems, from RoboBees [23] to BigDog [28] to RoboTuna [41], have benefited from a better understanding of animal motion. These theories and concepts of animal locomotion are developed based on observations over a large amount of data on the animals' motion. However, gathering such data is not always easy, especially when the manoeuvre under study seldom occurs, such as the mid-air collision avoidance of insects or the chase of a cheetah.

In this chapter, we present our preliminary work in bringing modelling techniques from robotics to enable biologists to perform an *in-silico* study of underlying motion strategies. This is distinct from animal dynamics and trajectory simulation as the focus is on the decision making strategy behind the observed motions, rather than the specific motions in space. In this work, we model the animal under study as a decision making agent, and generate the best motion strategy assuming the animal is a rational agent that tries to maximize a certain objective function, such as avoiding collision with minimal effort or catching its prey as fast as possible. We then use the motion strategy to generate simulated motions for the animal. Biologists can observe these simulated motions as if they are the motions of the animal being studied. This method of observation enables biologists to study various examples of motions that seldom occur. The model, the generated motion strategy, and the simulator make up the *in-silico* system for studying motion strategies of certain animal behaviour.

A principled approach for decision making in the presence of limited and uncertain data and varying parameters, is the Partially Observable Markov Decision Process (POMDP) framework. This framework is well-suited for our purpose. Aside from the limited data, no two animals are exactly alike even though they are of the same species. This uniqueness causes variations in parameters critical to generating rational motion

strategies. For instance, some honeybees have better vision than others, enabling them to perceive possible collisions more accurately and hence avoid collisions more often, different honeybees have different wing beat frequencies causing varying levels of manoeuvrability, etc. These variations, while complex, are not random; indeed, animal morphology provides additional information on these uncertainties and their mean effects. As has been shown in various robotics domains [13, 14], POMDP provides a robust way to incorporate and reason about these uncertainties.

We adopt the POMDP framework to model the collision avoidance strategies of flying animals such as birds, bats, and bees, who seem to avoid mid-air collisions effortlessly even in incredibly dense situations and apparently without the complex structure and communications of civil systems such as the Traffic Alert and Collision Avoidance System (TCAS). While the dynamics of a bird and a plane are different, a comparison of animal strategies with TCAS might better inform the ongoing development of next generation TCAS systems. This is an active research area especially due to the recent progress in Unmanned Aerial Vehicles (UAVs) that spurred the need for a more reliable and robust TCAS that can handle more traffic [34].

Interestingly, the POMDP framework is currently central to such efforts [16]. This coincidence is not accidental. Due to errors in sensing and control, an agent (e.g., pilot) may not know their exact state and the actions of the neighbouring entities. POMDP is designed to handle such types of uncertainty. Instead of finding the best action with respect to a single state, a POMDP solver finds the best action with respect to the set of states that are consistent with the available information so far. This set of states is represented as a probability distribution, called a belief b, and the set of all possible beliefs is called the belief space B. A POMDP solver calculates an optimal policy $\pi^* : B \to \mathcal{A}$ that maps a belief in B to an action in the set $\mathcal{A}$ of all possible actions the agent can perform, so as to maximize a given objective function. In TCAS, POMDP models the flying dynamics and sensing ability of an aircraft along with the errors and uncertainty of the system, to generate a robust collision-avoidance strategy for the aircraft.

Although solving a POMDP is computationally intractable in the worst case [26], recent developments of point-based POMDP approaches [3, 20, 27, 35] have drastically increased the speed of POMDP planning. Using sampling to trade optimality with approximate optimality for speed, point-based POMDP approaches have moved POMDP framework from solving a 12 states problem in days to solving non-trivial problems with millions of states and even problems with 10 dimensional continuous state space within seconds to minutes [2, 13, 14, 18, 19]. This progress in POMDP solving is key to its recent adoption in TCAS [39], and to the feasibility of our proposed *in-silico* system.

Leveraging this result, we adopt the POMDP model of TCAS and adjust the dynamics and sensing model to approximate those of flying animals. The solution to this POMDP problem is an optimal policy / motion strategy for the flying animal to avoid mid-air collisions. We also develop a simulator that simulates motion strategies of the animal that uses the policy to avoid mid-air collision in various encounter scenarios. The encounter scenarios are generated based on data and information about flight plans of the flying animal under study. Biologists can then use the simulator to generate and observe various motions on how the flying animal avoid mid-air collisions.

We have developed and tested our *in-silico* system for characterising mid-air collision avoidance for honeybees. We have also compared the simulated bee motion generated by our system and the motion of 100 actual honeybees in avoiding mid-air collisions. Preliminary results are promising, with less than 1% difference in the collision rate, and an average difference of approximately 12mm in the minimum distance between two bees in 100 head-on encounters, which corresponds roughly to the average wingspan of the bees in our data.

Of course an *in-silico* study of animal motion is no substitute for studying the motion of real animals. However, it may enable biologists to develop better initial hypotheses, and hence perform more focused and efficient studies on real animals, which can be much more costly and difficult compared to an *in-silico* study.

2.2 Related Work

2.2.1 Motion Strategies for Mid-Air Collision Avoidance

Motion is a defining characteristic of an animal. Its analysis, however, is typically focused on the dynamics and loadings that drive the motion [1]. The decision making strategies behind these motions are typically made by using the observed trajectories [25] to determine gait model parameters that are then compared to hypothesized models and strategies that minimise energy or forces, for example.

In the case of mid-air flight steering and collision avoidance, analysis has ranged from Ros et al. [29] who studied manoeuvrability in pigeons to Groening et al. [12] who studied pairwise collision avoidance behaviour in bees flying through narrow tunnels. They discovered that bees actively avoid mid-air collisions when they are flying. Discovering such behaviour requires a large amount of data, which is often difficult to get. This work propose to alleviate such difficulty by developing an *in-silico* system that generates trajectories similar to real animal trajectories, based on limited data and known information about the animal under study.

Mid-air collision avoidance is also of great interest to air traffic. Recent advancements in Unmanned Aerial Vehicles (UAVs) means heavier air traffic is expected in the near future, which spurred the need for more reliable and robust TCAS. One of the key issues in increasing TCAS' reliability and robustness is in taking into account the various uncertainty affecting pilots or UAVs in avoiding mid-air collision. Therefore, POMDP has been proposed [8] and successfully applied to improve the reliability and robustness of today's TCAS system [2, 39]. The POMDP model for TCAS provided a good starting point for our work.

2.2.2 Background on POMDP

Formally, a POMDP model is defined by a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, T, Z, R, \gamma, b_0 \rangle$, where $\mathcal{S}$ is a set of states, $\mathcal{A}$ is a set of actions, and $\mathcal{O}$ is a set of observations. At each time step, the POMDP agent is at a state $s \in \mathcal{S}$, performs an action $act \in \mathcal{A}$, and

perceives an observation $o \in \mathcal{O}$. Due to errors in its controller and the partially observed world dynamics, the next state the agent might be in after performing an action is uncertain. This uncertainty is modeled as a conditional probability function $T = f(s' \mid s, act)$, with $f(s' \mid s, act)$ representing the probability the agent moves from state s to s' after performing action act. Uncertainty in sensing is represented as a conditional probability function $Z = g(o \mid s', act)$, where $g(o \mid s', act)$ represents the probability the agent perceives observation $o \in \mathcal{O}$ after performing action act and ends at state s'.

Furthermore at each step, the agent receives a reward $R(s, act)$, if it takes action act from state s. The agent's goal is to choose a suitable sequence of actions that will maximize its expected total reward, while the agent's initial belief is denoted as b_0. When the sequence of actions may have infinite length, we specify a discount factor $\gamma \in (0, 1)$, so that the total reward is finite and the problem is well defined.

The solution of a POMDP problem is an *optimal policy* that maximizes the agent's expected total reward. A policy $\pi \colon B \to \mathcal{A}$ assigns an action act to each belief $b \in B$, and induces a value function $V(b, \pi)$ which specifies the expected total reward of executing policy π from belief b. The value function is computed as

$$V(b, \pi) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t, act_t) | b, \pi] \tag{2.1}$$

To execute a policy π, a POMDP agent executes action selection and belief update repeatedly. Suppose the agent's current belief is b. Then, it selects the action referred to by $act = \pi(b)$, performs action act and receives an observation o according to the observation function Z. Afterwards, the agent updates b to a new belief b' given by

$$\begin{aligned} b'(s') &= \tau(b, act, o) \\ &= \eta Z(s', act, o) \int_{s \in S} T(s, act, s') ds \end{aligned} \tag{2.2}$$

where η is a normalization constant.

2.3 The Model

In this work, our goal is to create a simulated flying animal that mimics the behaviour of the real animal in avoiding mid-air collisions with another flying animal. We refer to our simulated animal as the *outgoing* animal, while the other animal as the *incoming* animal. To take into account our lack of information about the behaviour and about the motion and sensing capabilities of the outgoing and incoming animals, we model the outgoing animal as a POMDP agent that needs to avoid colliding with the incoming animal whose flight plan is not perfectly known.

In particular, we adopt the POMDP model of TCAS [2]. This model is based on a very general and simplified flight dynamics and sensing model of airplanes, such that when we simplify the flying dynamics and sensing capabilities of the animals under study to a similar level of simplification used in [2], the set of parameters used to model the airplane dynamics and sensing in [2] are similar to those used for flying animals. Of course, the values of the parameters would be different, and need to be adjusted. We describe the model in this section, and discuss the required adjustments for honeybees in Section 2.5.

2.3.1 Flying Dynamics

The state space $\mathcal{S}$ of our POMDP model is a continuous space that represents the joint flight state spaces of the two animals. A flight state of each animal is specified as (x, y, z, θ, u, v), where (x, y, z) is the 3D position of the animal, θ is the animal's heading angle with respect to the positive direction of X axis, u is the animal's horizontal speed, and v is the animal's vertical speed (Figure 2.1 shows an illustration).

The action space $\mathcal{A}$ represents the control parameters of only the outgoing animal. It is a joint product of vertical acceleration a and turn rate ω. Considering the heavy computation cost of solving the POMDP model, we restrict a to be in discrete values $\{-a_m, 0, a_m\}$ and ω to be in discrete values $\{-\omega_m, 0, \omega_m\}$, where a_m and ω_m are the maximum vertical acceleration and the maximum turn rate, respectively. Although the

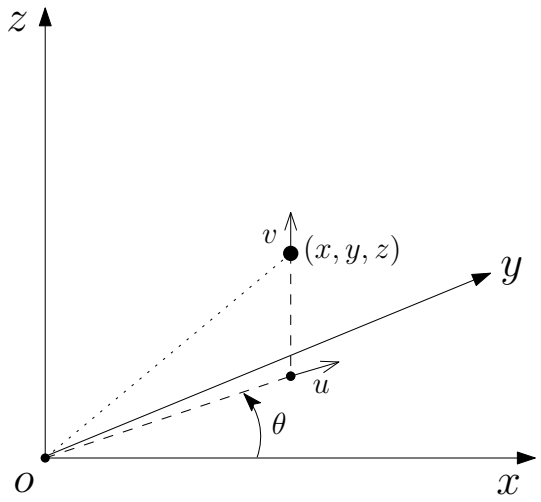


Figure 2.1 The flight state of one flying animal

control inputs are continuous, restricting their values to extreme cases is reasonable because when under the danger of near mid-air collisions, it is reasonable to assume that an animal will maximize its maneuvering in order to escape to a safe position. Figure 2.2 shows the 9 discrete actions. As the incoming animal's control inputs are unknown to the POMDP agent, we can either model them as uniformly randomized values, or as the controls of flying to some prescribed destinations, or based on information on the flight path of the animal under study.

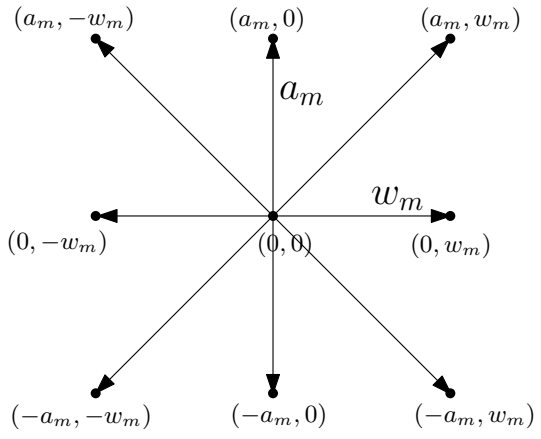


Figure 2.2 The action space contains 9 discrete actions.

We use a simplified model of flight dynamics in which each animal is treated as a point mass. Given a control (a, ω), the next flight state of an animal after a small time

duration Δt is given by

$$\begin{aligned} x_{t+1} &= x_t + u_t \Delta t cos\theta, \quad \theta_{t+1} &= \theta_t + \omega \Delta t, \\ y_{t+1} &= y_t + u_t \Delta t sin\theta, \quad u_{t+1} &= u_t, \\ z_{t+1} &= z_t + v_t \Delta t, \quad v_{t+1} &= v_t + a \Delta t. \end{aligned} \tag{2.3}$$

Figure 2.3 demonstrates this transition process. In this model, we assume that during the encounter process, the horizontal speed is a constant.

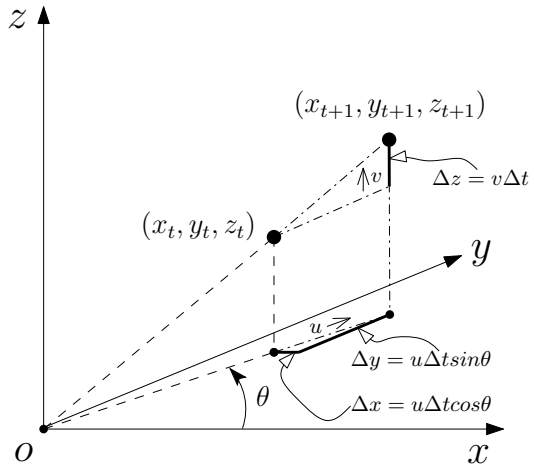


Figure 2.3 State transition from timestamp t to the next timestamp $t+1$, after a small time duration Δt.

2.3.2 Sensor Model

Although the outgoing animal has no prior information regarding the incoming animal's flight path, it can noisily sense the location of the incoming animal. Given this noisy sensor input, the outgoing animal (i.e., the agent) manoeuvres to prevent near mid-air collisions by keeping a safe separation distance from the incoming animal.

We assume the animal has a visibility sensor with limited field of view and limited range. The field of view is limited in the elevation direction (both up and down) with a maximum elevation angle of θ_e, and is limited in the horizontal direction (both left and right) with a maximum azimuth of θ_a. The range limit is denoted as D_R.

The observation space $\mathcal{O}$ is a discretization of the sensor's field of view. The discretization is done on the elevation and azimuth angles such that it results in n equally spaced bins along its elevation and azimuth angles. Figure 3.3 illustrates this discretization with $n = 16$. The observation space $\mathcal{O}$ is then these bins plus the observation `NO-DETECTION`, resulting in 17 observations in total.

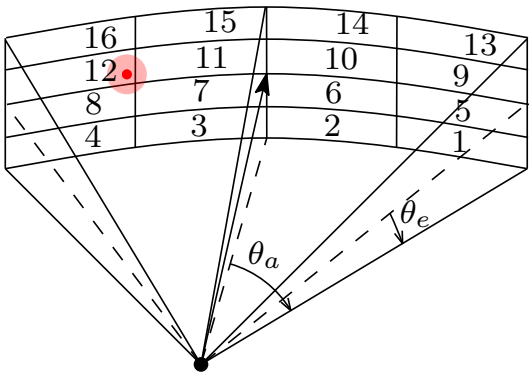


Figure 2.4 The sensor model. The black dot is the position of the agent; the solid arrow is the agent's flying direction. The red dot is the position of the incoming animal. Due to bearing error and elevation error, our agent may perceive the incoming animal at any position within the pink area.

As long as the incoming animal comes into the agent's sensor range (denoted as D_R) and into the visible space, it appears in a certain observation cell. For example, in Figure 3.3, the red dot represents the incoming bee, and it lies in 12. However, due to bearing error and elevation error, there will also be small probabilities that that agent observes the incoming animal to lie in cells 8, 7 and 11, respectively, and this brings uncertainties to the observation results. The bearing error is described by a normal distribution with zero mean and σ_b degree standard deviation; similarly, the elevation error is described by a normal distribution with zero mean and a standard deviation σ_e.

Other factors that contribute to observation uncertainties are false negative and false positive errors. False positive error is the probability of perceiving the incoming animal when it is out of range; false negative error is the probability of not perceiving the incoming animal when it is in range. Our sensor model can be described by the parameters in Table 2.1.

Table 2.1 Sensor Parameters

	Parameter
Range limit	D_R
Azimuth limit	θ_a
Elevation limit	θ_e
Bearing error standard deviation	σ_b
Elevation error standard deviation	σ_e
False positive probability	p_{fp}
False negative probability	p_{fn}

2.3.3 Reward Model

We assume that the outgoing flying animal is a rational agent that minimizes its risk of mid-air collision with the incoming flying animal, while avoiding collision with static objects in the environment. Furthermore, we assume the flying animal tries to use as few manoeuvres as possible to avoid collision. To model such behaviour in our POMDP agent, we use the following additive reward function $R(s, act) = R_C(s) + R_W(s) + R_M(s, act)$, where $R_C(s)$ is the penalty imposed if at state $s \in \mathcal{S}$, the outgoing and incoming animals collide, $R_W(s)$ is the penalty imposed if at state $s \in \mathcal{S}$, the outgoing animal collides with one or more static objects in the environment, and $R_M(s, act)$ is the cost for the outgoing animal to perform action $act \in \mathcal{A}$ from state $s \in \mathcal{S}$.

2.4 The Simulator

Given the POMDP problem as modelled in Section 2.3, the motion strategy for the outgoing animal is generated by solving the POMDP problem. Any POMDP solver can be used. In this work, we use Monte Carlo Value Iteration (MCVI) [3], which has been shown to perform well on POMDP-based TCAS [2], the model we have adopted for modelling motion strategies of flying animals in avoiding mid-air collision. Our simulator simulates the behaviour of the flying animal that uses this motion strategy to avoid mid-air collision in various head-on encounter scenarios. The head-on encounter

scenarios can be generated based on data or information about flight plans of the flying animal under study.

Biologists can then use the simulator to generate and observe how the agent avoids mid-air collision in various environments and encounter scenarios, to obtain an intuition on how the flying animals might avoid mid-air collisions.

One may argue that our simulator assumes that the flying animal acts rationally, in the sense that it tries to maximize a certain objective function, while the real animal may not act rationally. Indeed, this is true. However, our preliminary tests on honeybees data indicate that the underlying motion strategy of honeybees in avoiding mid-air collision may not be far from that of a rational agent (Section 2.5).

One may also argue that the objective function we set may not be the same as the objective function of the flying animal. Again, this is correct. However, if we acquire additional information that leads us to believe that the reward function needs to be modified, we can easily do so by revising the reward function in the model, regenerating the motion strategy, and revising the simulator to implement the new motion strategy.

2.5 Case Study on Honeybees

This case study is based on 100 pair-wised honeybee encounters in a 3-dimensional tunnel space. The size of the tunnel space is $930mm \times 120mm \times 100mm$. The possible coordinate values for x, y, z are $-30 \leq x \leq 900$, $-60 \leq y \leq 60$, and $-50 \leq z \leq 50$. Each encounter consists of the trajectories of two bees, in the format of $(x_1, y_1, z_1, x_2, y_2, z_2)$ at each timestamp, where (x_1, y_1, z_1) is the position of the outgoing bee and (x_2, y_2, z_2) is the position of the incoming bee. The data is sampled at 25 frames per second. Figure 2.5 shows one example of an encounter scenario. In it, the line-path represents the outgoing bee's flying trajectory, while the star-path represents the incoming bee's flying trajectory.

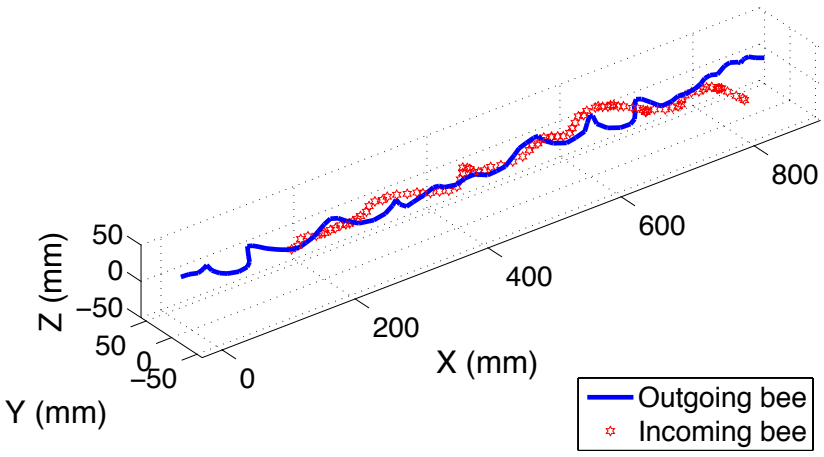


Figure 2.5 The 3D tunnel space in which the line-path represents the outgoing bee's trajectory and the star-path represents the incoming bee's trajectory.

2.5.1 Setting the Parameters

In our *in-silico* system, the outgoing bee is modelled as a POMDP agent as described in Section 2.3.

In the POMDP model, we assume the outgoing bee and the incoming bee share the same flying dynamics, i.e., they have the same horizontal velocity u, the same maximum/minimum vertical acceleration $\pm a_m$ and the same maximum/minimum turn rate $\pm\omega_m$. This is a reasonable assumption considering that both bees are of the same species, i.e., honeybees, and both behave in the same environment, i.e., the tunnel space. To get the exact values for these parameters, we perform statistical analysis on the data set. From this analysis, we can set $u = 300mm/s$, $a_m = 562.5mm/s^2$, and $w_m = 375deg/s$.

Now, we set the sensing parameters (Table 2.1). Since bees can see quite far and the length of the tunnel is less than one meter, we set the range limit to D_R to be infinite, to model the fact that the range limit of the bee's vision will not hinder its ability to see the other bee. The viewing angle of the bees remain limited. We set the azimuth limit θ_a to be 60 degrees and the elevation limit θ_e to be 60 degrees. The bearing error standard deviation σ_b and the elevation error standard elevation σ_e are

both set to be 1 degree. We assume that the false positive probability p_{fp} and the false negative probability p_{fn} are both 0.01.

We use the reward model as described in Section 2.3.3 to model the risk of near mid-air collisions and the risk of colliding with the tunnel boundaries.

We consider a state $s = (x_1, y_1, z_1, \theta_1, u_1, v_1, x_2, y_2, z_2, \theta_2, u_2, v_2) \in \mathcal{S}$ to be a collision state whenever the centre-to-centre distance between two parallel body axes is smaller than the wing span of the bee. By analysing the data and based on the biologist's observations on when collision occurs, we set this centre-to-centre distance (or wing span) to be 12mm. And define a state to be in collision when the two bees are within a cross-section distance (in YZ-plane) of 12mm and an axial distance (in X-direction) of 5mm, i.e., $\sqrt{(y_1 - y_2)^2 + (z_1 - z_2)^2} \leq 12$ and $\|x_1 - x_2\| \leq 5$. We assign collision penalty to be -10,000 as suggested in [2], i.e., $R_C(s) = -10,000$ whenever s is a collision state. In addition, to discourage unnecessary manoeuvres, we also assign a small penalty of -0.1 as suggested in [2], i.e., $R_M(s, act) = -0.1$ when act has a non-zero vertical speed or non-zero turn rate.

Bees have a tendency to fly in the centre of the tunnel. To mimic this flying tendency, we impose a penalty $R_W(s)$ when the bee is too close to the tunnel walls. Specifically, in the Y-axis, when our agent bee flies in the centre area of the tunnel ($-20 \leq Y \leq 20$), no penalty applies; beyond that, a penalty applies linearly proportional to the distance to the wall; when our agent hits walls, a maximum penalty $-10,000$ is imposed. Similarly, the gradient-based penalty mechanism is also applied in the z-axis.

2.5.2 Experimental Setup

The goal of this experiment is to measure the resemblance of the trajectories produced by the original outgoing bee and trajectories produced by POMDP for the outgoing bee. For this comparison, we use two measurements, derived from the necessary conditions for the two trajectories to be equivalent. The first measurement is the *Collision rate*, which is the percentage of colliding encounters (among the 100 encounters). The second measurement is the *Minimum Encounter Distance (MED)*, which is the

smallest Euclidean distance between the outgoing bee and the incoming bee during one whole encounter process. If the trajectories produced by POMDP are similar to the trajectories of the original outgoing bee, then both collision rate and MED should be similar too.

To generate the trajectories, we first need to solve the POMDP problem. For this purpose, we implement our POMDP problem in C++ and solve it using MCVI [3]. Since MCVI is a randomized algorithm, we generate 30 different policies to get reliable measurements. To reliably capture the effect of stochastic uncertainty on the collision avoidance strategy, for each policy, we run 100 simulations. Each simulation consists of 100 different encounter processes, where in each encounter, the simulated incoming bee follows one of the trajectories observed from a real bee. Each simulation run produces a collision rate. The average collision rate of the trajectories generated by the *in-silico* system is then the average collision rate over the 30×100 simulation runs. The average MED for a particular encounter situation is then the average MED over 30×100 simulation runs too.

All experiments are carried out on a Linux platform with a 3.6GHz Intel Xeon E5-1620 and 16GB RAM.

2.5.3 Experimental Results

Table 2.2 Collision Rates

Policy	Collision rate	Margin of error (95% Conf.)
Bee	3%	—
POMDP	3.84%	±0.13%

Applying the collision definition of our POMDP agent to the 100 bee data, we found the collision rate of this set of data is 3%. Table 2.2 shows the collision rates of the original data and the average collision rate of the POMDP-based *in-silico* system. The two collision rates are less than 1% difference, which indicates that the trajectories

produced by the POMDP-based *in-silico* system are similar to trajectories produced by the actual bees in terms of collision rate.

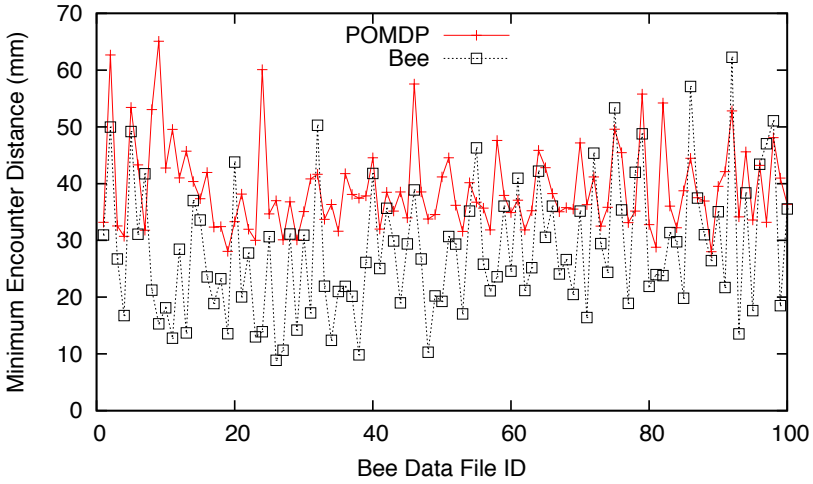


Figure 2.6 *Minimum Encounter Distance (MED)* for the bee data and the simulated encounter in our POMDP-based *in-silico* system.

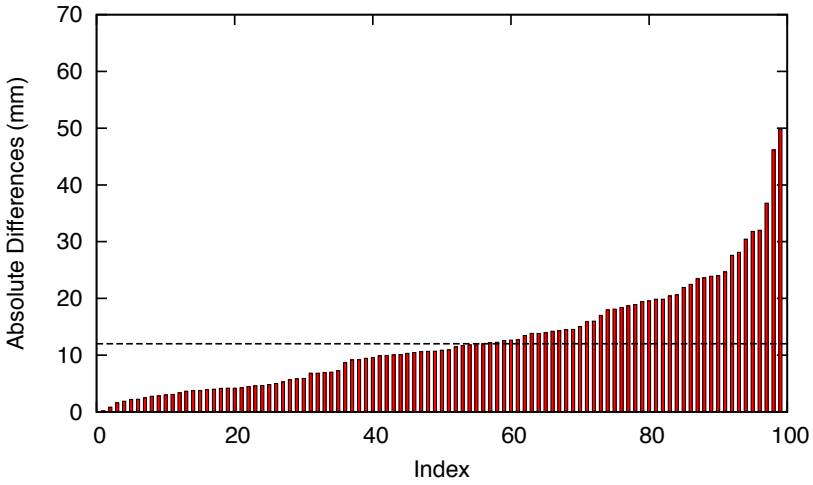


Figure 2.7 The sorted absolute difference between the *MED* measure of real bee encounters and the average *MED* of the simulated encounters generated by the *in-silico* system

Figure 2.6 shows the *Minimum Encounter Distance (MED)* for the real bee data and the simulated encounter in our POMDP-based *in-silico* system. Figure 2.7 shows a histogram depicting the increasing sorted absolute differences between the *MED* measurement of the real bee data and the average *MED* measurement of the simulated encounters in our POMDP-based *in-silico* system. This histogram shows that in 98%

of the encounters, the absolute difference between the *MED* of the real bee data and that of the simulated encounters is less than 35mm, while in 60% of the encounters, the absolute difference between the *MED*s is less than 12mm. In fact, the average absolute difference between the *MED* measure of the real bee data and the average *MED* measure of the POMDP-based *in-silico* system is 12.60, with 95% confidence interval of 1.85.This 12mm average absolute difference is roughly equivalent to the estimated average wing span of the honeybees in our data, which implies that in terms of *MED* measure, our POMDP-based *in-silico* system produces similar results to the original bee data.

2.6 Summary

In this chapter, we propose a POMDP-based *in-silico* system to help biologists study mid-air collision avoidance strategies of flying animals. Our system is distinct from flying animal dynamics and trajectory simulation, as the focus is on the strategy behind the observed motions, rather than the specific motions in space. Our *in-silico* system consists of a model and a simulator. We model the animals as decision making agents under the POMDP framework. The solution to this POMDP problem is an optimal motion strategy for the agent to avoid mid-air collision with another flying animal. Our simulator simulates the behaviour of a flying animal that uses this motion strategy in various head-on encounter scenarios. The head-on encounter scenarios are generated based on data and information about flight plans of the flying animal under study.

We tested our system on 100 honeybee encounters. We measure how close our *in-silico* system to the actual bee using two measurements —collision rate and minimum encounter distance— that are derived from the necessary conditions for the two systems to be equivalent. Preliminary results indicate that our POMDP-based *in-silico* system is a promising tool to study mid-air collision avoidance strategies of flying animals, *in-silico*. Such a tool may help biologists better understand mid air collision-avoidance

strategies of flying animals faster and with much less cost, which in turn may benefit the robotics community in developing better mid air collision-avoidance system.

Many avenues are possible for future work. First is the measurement to determine if the *in-silico* system generates similar trajectories as the real data. In this work, we have used measurements derived from the necessary conditions. A better measurement should be derived from the sufficient and necessary condition. Second is to test the system on more data and various different scenarios. Third is to expand the system to handle more complex encounter scenarios.

Chapter 3

In-silico Behavior Discovery System: A Prototype

It is now widely accepted that a variety of interaction strategies in animals achieve optimal or near optimal performance[1]. The challenge is in determining the performance criteria being optimized. A difficulty in overcoming this challenge is the need for a large body of observational data to delineate hypotheses, which can be tedious and time consuming, if not impossible. To alleviate this difficulty, we propose a system — termed "*in-silico* behavior discovery" — that will enable ethologists to simultaneously compare and assess various hypotheses with much less observational data. Key to this system is the use of Partially Observable Markov Decision Processes (POMDPs) to generate an optimal strategy under a given hypothesis. POMDPs enable the system to take into account imperfect information about the animals' dynamics and their operating environment. Given multiple hypotheses and a set of preliminary observational data, our system will compute the optimal strategy under each hypothesis, generate a set of synthesized data for each optimal strategy, and then rank the hypotheses based on the similarity between the set of synthesized data generated under each hypothesis and

[1]This chapter is copied from the following publication: H. Wang, H. Kurniawati, S. P. N. Singh, and M. V. Srinivasan. In-silico Behavior Discovery System: An Application of Planning in Ethology. In *Proc. International Conference on Automated Planning and Scheduling (ICAPS)*, 2015. (Outstanding Student Paper Award)

the provided observational data. In particular, this chapter considers the development of this approach for studying mid-air collision-avoidance strategies of honeybees. To perform a feasibility study, we test the system using 100 data sets of close encounters between two honeybees. Preliminary results are promising, indicating that the system independently identify the same hypothesis (optical flow centering) as discovered by neurobiologists/ethologists.

3.1 Introduction

What are the underlying strategy that animals take when interacting with other animals? This is a fundamental question in *ethology*. Aside from human curiosity, the answer to such a question may hold the key to significant technological advances. For instance, understanding how birds avoid collisions may help develop more efficient collision avoidance techniques for Unmanned Aerial Vehicles (UAVs), while understanding how cheetahs hunt may help develop better conservation management systems.

Although it is now widely accepted that a variety of interaction strategies in animals have been shaped to achieve optimal or near optimal performance [5, 9], determining the exact performance criteria that are being optimized remains a challenge. Existing approaches require ethologists to infer the criteria being optimized from many observations on how the animals interact. These approaches present two main difficulties. First, the inference is hard to do, because even extremely different performance criteria may generate similar observed data under certain scenarios. Second, obtaining observational data are often difficult. Some interactions rarely occur. For instance, to understand collision avoidance strategies in insects and birds, it is necessary to observe many near-collision encounters, but such events are rare because these animals are very adept at avoiding collisions. Ethologists can resort to experiments that deliberately cause close-encounter events, but such experiments are tedious, time-consuming, and may not faithfully capture the properties of the natural environment. Furthermore, such experiments may not be possible for animals that have become extinct, such as

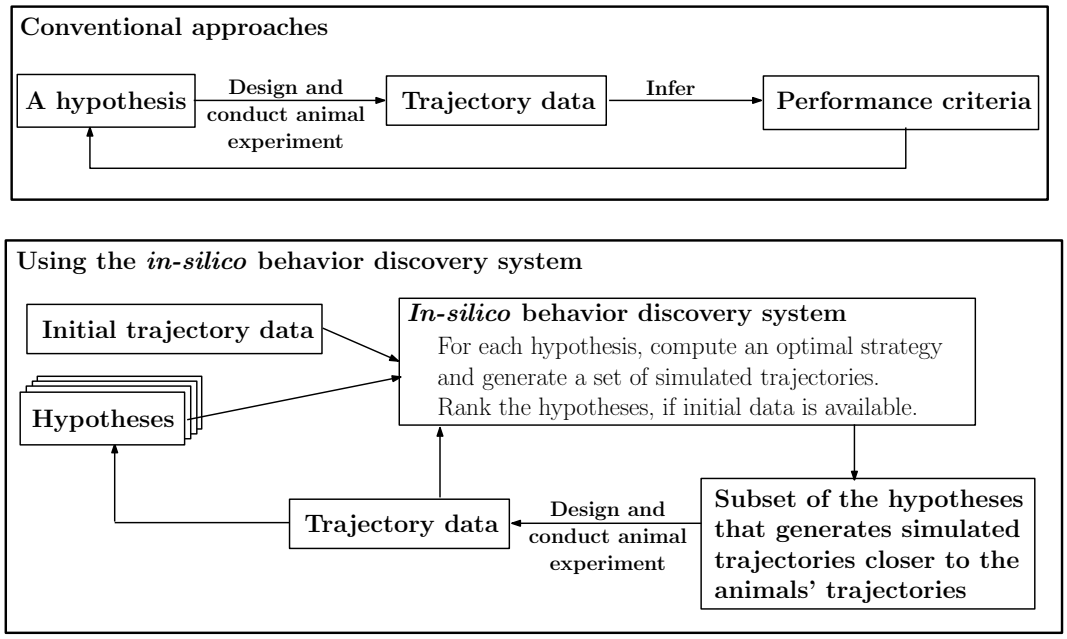


Figure 3.1 Conventional and proposed approach to study performance criteria used by honeybees to avoid collision with each other.

dinosaurs, in which case ethologists can only rely on limited historical data, such as fossil traces.

This chapter presents our preliminary work in developing a system — termed "*in-silico* behavior discovery" — to enable ethologists study animals' strategies by simultaneously comparing and assessing various performance criteria on the basis of limited observational data.

In this preliminary work, we focus on developing and verifying the feasibility of an *in-silico* behavior discovery system to study mid-air collision-avoidance strategies of honeybees. The difference between an approach using our system and conventional approaches is illustrated in Figure 3.1. Our system takes as many hypotheses as the user choose to posit, and data from preliminary experiments. Preliminary data are collision-avoidance encounter scenarios, and each encounter consists of a set of flight trajectories of all honeybees involved in one collision-avoidance scenario. Each hypothesis is a performance criterion that may govern the honeybees' collision-avoidance strategies. For each hypothesis, the system generates an optimal collision-avoidance strategy and generates simulated collision-avoidance trajectories based on that. Given the set of

simulated collision-avoidance trajectories, the system will rank the hypotheses based on the similarity between the set of simulated trajectories and the preliminary trajectory data. By being able to generate simulated data under various hypotheses, the *in-silico* behavior discovery system enables ethologists to "extract" more information from the available data and better focus their subsequent data gathering effort, thereby reducing the size of exploratory data required to find the right performance criteria that explains the collision avoidance behavior of honeybees. This iterative process is illustrated in Figure 3.1.

Obviously, a key question is how to generate the collision-avoidance strategy under a given performance criterion. In our system, we use the Partially Observable Markov Decision Processes (POMDPs) framework. One may quickly argue that it is highly unlikely an insect such as a bee runs a POMDP solver in its brain. This may be true, but the purpose of our system is *not* to mimic honeybee neurology. Rather, we use the widely accepted idea in biology — i.e., most interaction strategies in animals achieve optimal or near optimal performance — to develop a *tool that helps ethologists predict and visualize their hypotheses prior to conducting animal experiments to test the hypotheses*, thereby helping them to design a more focused and fruitful animal experiments. In fact, POMDPs allows us to relax the need to model the exact flight dynamics and perception of the honeybees. No two animals are exactly alike, even though they are of the same species. This uniqueness causes variations in various parameters critical to generating the strategies. For instance, some honeybees have better vision than others, enabling them to sense impending collisions more accurately and hence avoid collisions more often, different honeybees have different wing beat frequencies causing varying manoeuvrability, etc. Our system frames these variations as stochastic uncertainties — commonly used modelling in analysing group behavior — and takes them into account when computing the optimal collision-avoidance strategy under a given performance criterion.

We tested the feasibility of our *in-silico* behavior discovery system using a data set comprising 100 close encounter scenarios between two honeybees. The results indicate

that the system independently identify the same hypothesis (optical flow centering) as discovered by neurobiologists/ethologists.

3.2 Background and Related Work

A POMDP model is defined by a tuple $\langle S, A, O, T, Z, R, \gamma, b_0 \rangle$, where S is a set of states, A is a set of actions, and O is a set of observations. At each time step, the POMDP agent is at a state $s \in \mathcal{S}$, performs an action $act \in \mathcal{A}$, and perceives an observation $o \in \mathcal{O}$. POMDP represents the uncertainty in the effect of performing an action as a conditional probability function, called the transition function, $T = f(s' \mid s, act)$, with $f(s' \mid s, act)$ representing the probability the agent moves from state s to s' after performing action act. Uncertainty in sensing is represented as a conditional probability function $Z = g(o \mid s', act)$, where $g(o \mid s', act)$ represents the probability the agent perceives observation $o \in \mathcal{O}$ after performing action act and ends at state s'.

At each step, a POMDP agent receives a reward $R(s, act)$, if it takes action act from state s. The agent's goal is to choose a sequence of actions that will maximize its expected total reward, while the agent's initial belief is denoted as b_0. When the sequence of actions may have infinitely many steps, we specify a discount factor $\gamma \in (0, 1)$, so that the total reward is finite and the problem is well defined.

The solution of a POMDP problem is an *optimal policy* that maximizes the agent's expected total reward. A policy $\pi \colon B \to \mathcal{A}$ assigns an action act to each belief $b \in B$, and induces a value function $V(b, \pi)$ which specifies the expected total reward of executing policy π from belief b. The value function is computed as

$$V(b, \pi) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t, act_t) | b, \pi] \tag{3.1}$$

To execute a policy π, a POMDP agent executes action selection and belief update repeatedly. Suppose the agent's current belief is b. Then, it selects the action referred to by $act = \pi(b)$, performs action act and receives an observation o according to the

observation function Z. Afterwards, the agent updates b to a new belief b' given by

$$\begin{aligned} b'(s') &= \tau(b, act, o) \\ &= \eta Z(s', act, o) \int_{s \in S} T(s, act, s') ds \end{aligned} \tag{3.2}$$

where η is a normalization constant.

A more detailed review of the POMDP framework is available in [15].

Although computing the optimal policy is computationally intractable [26], results over the past decade have shown that by trading optimality with approximate optimality for speed [20, 27, 33, 35], POMDP can start becoming practical for various real world problems [2, 13, 17, 44].

To the best of our knowledge, this is the first system that applies planning under uncertainty to help ethologists reduce the number of necessary observational data. The closest to this work is Chapter 2. It uses POMDP to generate a near-optimal collision avoidance strategies of honeybees under a given hypothesis, and lets biologists observe the simulated trajectories, manually. In contrast, this work proposes a system that takes multiple hypotheses at once and provides a ranking of how likely the hypotheses generate the observational data.

3.3 The *In-silico* Behavior Discovery System

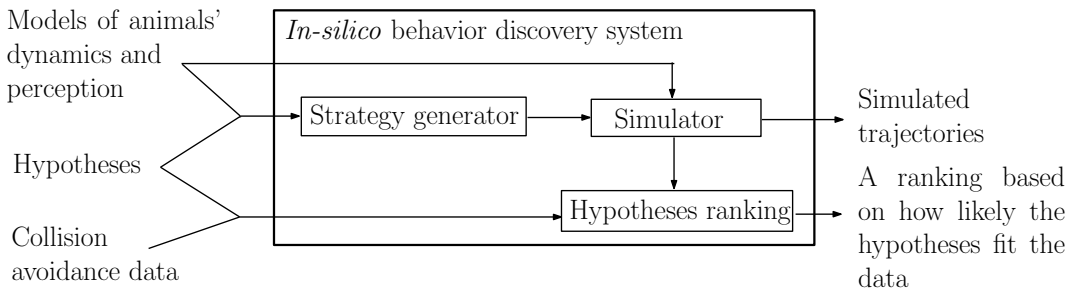


Figure 3.2 The inputs, outputs, and main components of the proposed *in-silico* behavior discovery system.

Figure 3.2 illustrates an overview of the *in-silico* behavior discovery system for studying collision avoidance strategies of honeybees. It is a software system that takes as input the animals' flight dynamics and perception models, a set of hypotheses on the performance criteria used by honeybees to avoiding mid-air collision, and a set of collision avoidance trajectories of honeybees. These trajectories are usually small in number and act as preliminary observational data. The system computes the optimal collision avoidance strategy for each hypothesis. It outputs a set of simulated trajectories under each strategy, along with a ranking on which hypotheses are more likely to explain the preliminary data. The ranking is based on the similarity between the simulated trajectories of the hypotheses and the preliminary data. The system consists of three main modules, i.e.:

- Strategy Generator, which computes the optimal strategy under each hypothesis.
- Simulator, which generates the simulated trajectories under each strategy that has been computed by the Strategy Generator module.
- Hypothesis Ranking, which identifies the hypotheses that are more likely to explain the observational data. For each hypothesis, the strategy generator and the simulator modules generate the simulated trajectories under the hypothesis. Once the sets of simulated trajectories have been generated for all hypotheses, the hypothesis ranking module will rank the hypotheses based on the similarity between the simulated trajectories and the observational data.

The details of each module are described in the following sub-sections.

3.3.1 Strategy Generator and Simulator

The strategy generator module is essentially a POMDP planner that generates an optimal collision avoidance strategy under hypothesized performance criteria used by honeybees to avoid mid-air collision in various head-on encounters. Since this chapter focuses only on head-on encounters, the number of honeybees involved in each encounter is only two. In this work, we also assume that the bees do not communicate/negotiate

when avoiding collision. This assumption is in-line with the prevailing view in the relatively open question of whether bees actually negotiate for avoiding collision. Furthermore, it simplifies our POMDP model in the sense that it suffices to model each bee independently as a single POMDP agent, rather than all bees at once as a multi-agent system.

The POMDP framework is used to model a honeybee “agent" that tries to avoid collisions with another honeybee, assuming the agent optimizes the hypothesized performance criteria. The flying dynamics and perception models become the transition and observation functions of the POMDP model, while each hypothesis is represented as a reward function of the POMDP model. POMDP enables the system to take into account variations in the honeybees' flight dynamics, for instance due to their weight and wingspan, or variations in the honeybees' perceptive capacities, and captures the agent's uncertainty about the behavior of the other bee.

One may argue that even the best POMDP planner today will not achieve the optimal solution to our problem within reasonable time. This is true, but a near optimal solution is often sufficient. Aside from results in ethology that indicate animals often use near optimal strategies too [5, 9], our system can help focus subsequent animal experiments as long as the strategy is sufficient to correctly identify which hypotheses are more likely to be correct, based on the similarity of the simulated trajectories under the hypotheses and the trajectories from real data. In many cases, we can correctly identify such hypotheses without computing the optimal collision-avoidance strategies, as we will show in our Results section.

Another critique of using POMDP is that POMDP requires Markov assumption that is unlikely to be true in bees' motion. However, POMDP is Markovian in the belief space. Since beliefs are sufficient statistics of the entire history, a POMDP agent, and hence our simulated bees, selects the best actions by considering the entire history of actions and observations. POMDP does require the transition function to be Markovian. However, this can often be satisfied by suitable design of the state and action space. In this work, we assume bees are kinematic — a commonly used

simplification in modelling complex motion — where the next position and velocity is determined by the current position, velocity, and acceleration. More details on this model are discussed in subsequent paragraphs.

Our POMDP model is an adaptation of the POMDP model [2] designed for the Traffic Alert and Collision Avoidance System (TCAS) — a collision avoidance system mandatory for all large commercial aircraft. One would argue that this model is not suitable because the flight dynamics and perception model of aircraft are totally different than those of honeybees. Indeed their dynamics and perception are different. However, the model in [2] is a highly abstracted flight dynamics and perception model of aircraft, such that if we apply the same level of abstraction to the flight dynamics (simplified to its kinematic model) and perception of honeybees (simplified to visibility sensors), we would get a similar model, albeit with different parameters. In this chapter, we adjust the parameters based on the literature and data on the flight dynamics and perception capabilities of honeybees.

For completeness, we describe the POMDP model here together with the required parameter adjustment. Although our POMDP will only control one of the honeybees involved in the close-encounter scenarios, the position, heading, and velocity of the two honeybees determine the collision avoidance strategy. Therefore, the state space $\mathcal{S}$ consists of the joint flight state spaces of the two honeybees involved. A flight state of a honeybee is specified as (x, y, z, θ, u, v), where (x, y, z) is the 3D position of the bee, θ is the bee's heading angle with respect to the positive direction of X axis, u is the bee's horizontal speed, and v is the bee's vertical speed.

The action space $\mathcal{A}$ represents the control parameters of only one of the honeybees. It is a joint product of vertical acceleration a and turn rate ω. Since most practical POMDP solvers [20, 33] today only perform well when the action space is small, we use bang-bang controller, restricting the acceleration a to be $\{-a_m, 0, a_m\}$ and the turning rate ω to be $\{-\omega_m, 0, \omega_m\}$, where a_m and ω_m are the maximum vertical acceleration and the maximum turn rate, respectively. Although the control inputs are continuous, restricting their values to extreme cases is reasonable because under the

danger of near mid-air collisions, a bee is likely to maximize its maneuvering in order to escape to a safe position as fast as possible. And control theory has shown that maximum-minimum (bang-bang) control yields time-optimal solutions under many scenarios [21]. We assume that the other bee —whose action is beyond the control of the POMDP agent— has the same possible control parameters as the POMDP agent. However, which control it uses at any given time is unknown and is modelled as a uniform distribution over the possible control parameters.

The transition function represents an extremely simplified flight dynamics. Each bee is treated as a point mass. And given a control (a, ω), the next flight state of an animal after a small time duration Δt is given by

$$\begin{aligned} x_{t+1} &= x_t + u_t \Delta t cos\theta, & \theta_{t+1} &= \theta_t + \omega \Delta t, \\ y_{t+1} &= y_t + u_t \Delta t sin\theta, & u_{t+1} &= u_t, \\ z_{t+1} &= z_t + v_t \Delta t, & v_{t+1} &= v_t + a \Delta t. \end{aligned}$$

Although a honeybee's perception is heavily based on optical flow [32, 37], to study the collision avoidance behavior, we can abstract its perception to the level of where it thinks the other bee is, i.e., the perception after all sensing data has been processed into information about its environment. Therefore, we can model the bee's observation space in terms of a sensor that has a limited field of view and a limited range.

The observation space $\mathcal{O}$ is a discretization of the sensor's field of view. The discretization is done on the elevation and azimuth angles such that it results in 16 equally spaced bins along the elevation and azimuth angles. Figure 3.3 illustrates this discretization. The observation space $\mathcal{O}$ is then these bins plus the observation `NO-DETECTION`, resulting in 17 observations in total.

As long as the incoming animal comes into the agent's sensor range (denoted as D_R) and into the visible space, it appears in a certain observation grid, with some uncertainty. The observation function models the uncertainty in bearing and elevation, as well as false positives and false negatives.

The parameters for the observation model are:

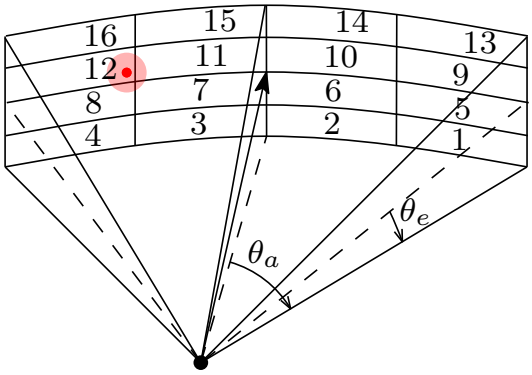


Figure 3.3 The observation model. The black dot is the position of the agent; the solid arrow is the agent's flying direction. The red dot is the position of the incoming honeybee. Due to bearing and elevation errors, our agent may perceive the other bee to be at any position within the shaded area.

- Range limit, parameterized as D_R.
- Azimuth limit, parameterized as θ_a.
- Elevation limit, parameterized as θ_e.
- Bearing error standard deviation, parameterized as σ_b.
- Elevation error standard deviation, parameterized as σ_e.
- False positive probability, parameterized as p_{fp}.
- False negative probability, parameterized as p_{fn}.

The reward function will be different for different hypotheses of the performance criteria used by the honeybees in avoiding collision. We will discuss them in the Simulation Setup section, when describing the hypotheses that we use to test the system.

A POMDP simulator is used, in the sense that it takes the POMDP model and policy as inputs, and then generates the collision avoidance trajectories of the bee under various head-on encounter scenarios. The scenarios we use in the simulator are similar to the encounter scenarios in the real trajectories. Recall that a collision-avoidance trajectory is a set of flight trajectories of all the honeybees involved in the encounter scenario. In this work, only two honeybees are involved in each scenario, as we focus on head-on encounters. Our simulator uses similar encounter scenarios as the real

data, in the sense that we only simulate the collision avoidance strategies of one of the two honeybees, while the other bee follows the flight trajectory of the real data. Therefore, each set of collision-avoidance trajectories generated by our simulator will have a one-to-one mapping with the set of real collision-avoidance trajectories that has been given to the system. For statistical significance, in general, our system generates multiple sets of simulated trajectories.

3.3.2 Hypotheses Ranking

The key in this module is the metric used to identify the similarity between a set of simulated collision-avoidance trajectories and a set of real collision-avoidance trajectories. Recall that each set of simulated collision avoidance trajectories has a one-to-one mapping with the set of real collision-avoidance trajectories. Let us denote this mapping by g. Suppose $\mathcal{A}$ is a set of simulated collision-avoidance trajectories and $\mathcal{B}$ is the set of real collision-avoidance trajectories given as input to the system. Then we define the similarity $sim(\mathcal{A}, \mathcal{B})$ between $\mathcal{A}$ and $\mathcal{B}$ as a 3-tuple $\langle \overline{F}, \overline{M}, C \rangle$, where:

- The notation $\overline{F}$ is the average distance between the flight path of the simulated trajectories and that of the real trajectories. Suppose L is the number of trajectories in $\mathcal{A}$. Then, $\overline{F(\mathcal{A}, \mathcal{B})} = \frac{1}{L} \sum_{i=1}^{L} F(A_i, B_i)$ where $F(A_i, B_i)$ denotes the **Fréchet Distance** between the curve traversed by the simulated bee in $A_i \in \mathcal{A}$ and the curve traversed by the corresponding bee in $B_i = g(A_i) \in \mathcal{B}$. In our system, each curve traversed by a bee is represented as a polygonal curve because, the trajectory generated by our simulator assumes discrete time steps (a property inherited from the POMDP framework).

  The Fréchet Distance computes the distance between two curves, taking into account their course. A commonly used intuition to explain Fréchet Distance is based on an analogy of a person walking his dog. The person walks on one curve and the dog on the other curve. The Fréchet Distance is then the shortest leash that allows the dog and its owner to walk along their respective curves,

from one end to the other, without backtracking [6, 7]. Formally,

$$F(A_i, B_i) = \min_{\substack{\alpha[0,1]\to[0,N] \\ \beta[0,1]\to[0,M]}} \left(\max_{t\in[0,1]} dist\Big(A_i(\alpha(t)), B_i(\beta(t)) \Big) \right)$$

where *dist* is the underlying distance metric in the honeybees' flight space. In our case, it is the Eucledian distance in $\mathbb{R}^3$. N and M are the number of segments in the polygonal curves A_i and B_i respectively. The function α is continuous with $\alpha(0) = 0$ and $\alpha(1) = N$ while β is continuous with $\beta(0) = 0$ and $\beta(1) = M$. These two functions are possible parameterizations of A_i and B_i.

- The notation $\overline{M}$ denotes the average absolute difference in **Minimum Encounter Distance** (*MED*). MED of a collision-avoidance trajectory computes the smallest Euclidean distance between the two honeybees, e.g., for a collision-avoidance trajectory A_i, $MED(A_i) = \min_{t=1}^{T} dist(A_i(t), A_i'(t))$ where T is the smallest last timestamp among the trajectories of the two honeybees, $A_i(t)$ and $A_i'(t)$ are the trajectories of bee-1 and bee-2 in A_i at time t respectively, and *dist* is the Euclidean distance between the two positions. MED measures how close two honeybees can be during one encounter. Small $\overline{M}$ is a necessary condition for a simulated trajectory to resemble the real trajectory, in the sense that if the simulated trajectories of the incoming and the outgoing bees are similar to the observed trajectories, then the minimum encounter distance between the simulated incoming and outgoing bees should be similar to that of the observed trajectories.
- The notation C denotes the absolute difference in the **Collision Rate**. The *Collision rate* is defined as the percentage of the collision that occur. Small C is a necessary condition for a simulated trajectory to resemble the real trajectory, in the sense that if the simulated and the real bees have similar capabilities in avoiding collisions, then assuming the trajectories and environments are similar, the collision rate of the simulated and real bees should be similar.

For statistical analysis, in general, our system generates multiple sets of simulated collision-avoidance trajectories for each hypothesis. The goodness of the hypothesis in explaining the input data is then defined as the average 3-tuple metric over all sets of simulated trajectories generated by the system. Suppose the system generates K sets of simulated trajectories, e.g., $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_K$ for hypothesis H_1. Then the goodness of H_1 in explaining the input data is a 3-tuple where the first element is $\frac{1}{K}\sum_{i=1}^{K}\overline{F(\mathcal{A}_i, \mathcal{B})}$, the second element is $\frac{1}{K}\sum_{i=1}^{K}\overline{M(\mathcal{A}_i, \mathcal{B})}$, and the third element is $\frac{1}{K}\sum_{i=1}^{K} C(\mathcal{A}_i, \mathcal{B})$ where $\mathcal{B}$ is the set of real collision-avoidance trajectories that the system received as inputs. The order in the tuple acts as prioritization. The system assigns a higher rank to the hypothesis whose goodness value has the smaller first element. If the goodnesses of two hypotheses have a similar first element, i.e., they are the same with more than 95% confident based on student t-test hypothesis testing, then the second element becomes the determining factor, and so on. Now, this ranking system may not be totally ordered, containing conflict on the the ordering. When such a conflict is found, we apply the Kemeny-Young voting method [22] to enforce a total ordering of the resulting ranking.

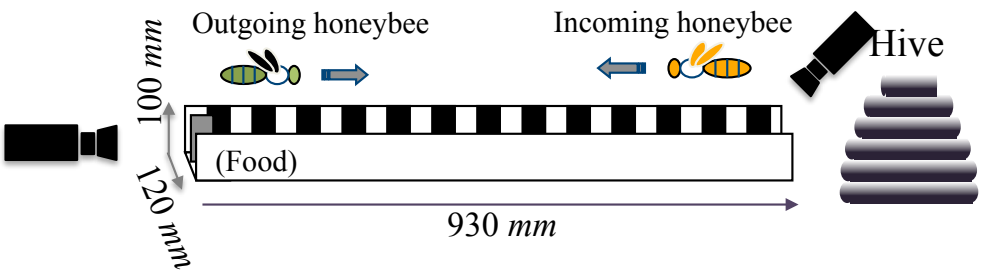
Note that although in this chapter, there are only two honeybees involved in each collision-avoidance trajectory, it is straightforward to extend the aforementioned similarity metric and ranking strategy to handle encounter scenarios where many more honeybees are involved.

3.4 System Verification

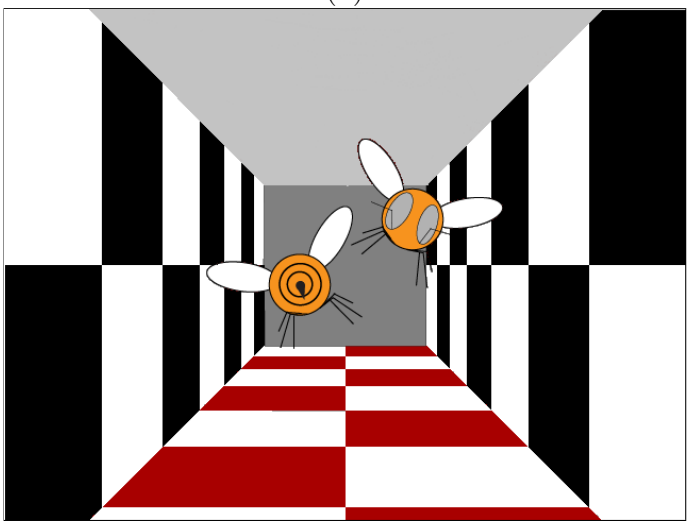
To verify the system, we will use the system to rank several hypotheses in which the performance criterion closest to the correct one is known.

3.4.1 Collision-Avoidance Trajectories of Real Honeybees

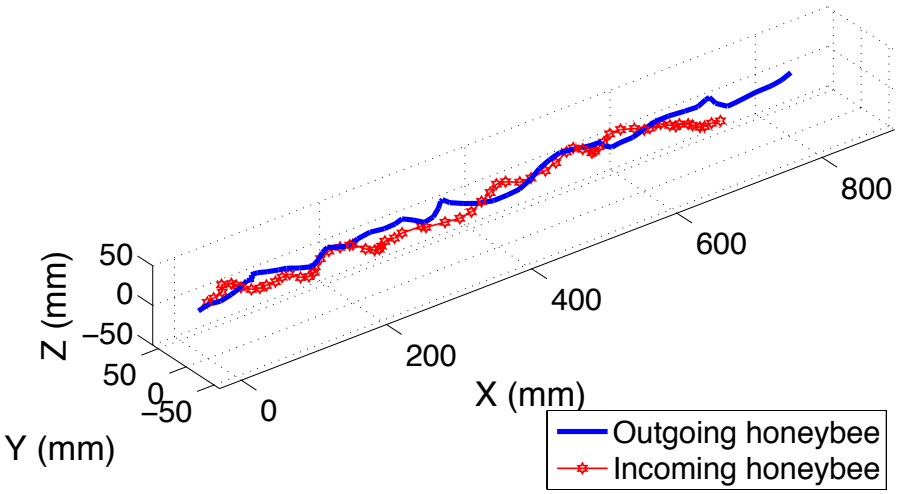
To verify the applicability of our system, we use 100 sets of collision-avoidance trajectories as preliminary data.The data are gathered from experiments conducted at



(a)



(b)



(c)

Figure 3.4 Illustration of experimental design and setup for gathering collision-avoidance trajectories of real honeybees. These trajectories are used as an input (Initial Trajectory Data) to our system. (a) The tunnel. (b) The inner part of the tunnel. (c) A collision-avoidance trajectory gathered from this experiment.

the Neuroscience of Vision and Aerial Robotics in the Queensland Brain Institute. These data are the results of experimental recording of 100 head-on encounters of two honeybees flying along a 3-dimensional tunnel. Figure 3.4 illustrates the experimental setup to gathered the data. The tunnel dimensions are $930mm \times 120mm \times 100mm$. The roof of the tunnel is transparent. The left, right, and bottom wall of the tunnel are covered with checkerboard patterns, where each square is of size $2.2cm \times 2.2cm$. The left and right patterns are colored black/white, while the bottom pattern is red/white, to aid the detection of the honeybees, which are generally dark in color. These patterns aid the honeybees' navigation through the tunnel. The tunnel is placed with its entrance near a beehive, and a sugar water feeder is placed inside the tunnel at its far end. To record a collision-avoidance trajectory, a bee is first released from the hive to the tunnel. This bee will fly towards the feeder, collect the food, and then fly back to the hive. When the bee starts to fly back to the hive, another bee is released from the hive to the tunnel and flies towards the feeder. We denote the bee flying towards the feeder as the incoming bee and the bee flying towards the hive as the outgoing bee.

The trajectories of the honeybees are recorded using two cameras —one positioned above the tunnel, looking down, and another camera positioned at the far end of the tunnel, looking axially into the tunnel. The stereo cameras capture the bees' flight at 25 frames per second. Based on the positioning and the resolution of the two stereo cameras, the estimated precision of the reconstructed 3D trajectories is approximately $2mm \times 2mm \times 2mm$.

Figure 3.4(c) shows the coordinate frame and one example of a collision-avoidance trajectories reconstructed in 3D. The possible coordinate values are $-30 \leq X \leq 900$, $-60 \leq Y \leq 60$, and $-50 \leq Z \leq 50$. Each collision-avoidance trajectory consists of the trajectories of the two honeybees, represented as a sequence of positions of the two honeybees. Each element of the sequence follows the following format $(x_1, y_1, z_1, x_2, y_2, z_2)$, where (x_1, y_1, z_1) is the position of the outgoing honeybees and (x_2, y_2, z_2) is the position of the incoming bee.

3.4.2 Hypotheses

To verify our system, we use six hypotheses as the input to our system. These hypotheses are selected in a way that we know exactly which hypotheses are closer to the correct performance criteria. Each hypothesis is represented as a reward function in the POMDP problem. It is essentially a summation of the component cost and reward. We will discuss the detailed values of all component costs and reward in the Simulation Setup section. The hypotheses are:

- H_{Basic} is the basic collision avoidance hypothesis. In this hypothesis, the reward function is the summation of collision cost and movement cost.
- $H_{BasicDest}$ is the hypothesis that the honeybees do not forget their goal of reaching the feeder or the hive, even though they have to avoid mid-air collision with another bee. In this hypothesis, we provide a high reward when the bee reaches its destination. The reward function is then the summation of the collision cost, the movement cost, and the reward for reaching the goal. This behavior is evident from the 100 collision-avoidance trajectories that were recorded. All trajectories indicate that the honeybees fly toward both ends of the tunnel, instead of wandering around within the tunnel or turning back before reaching their goals.
- H_{LR} is the hypothesis that the honeybees tend to perform horizontal centering. This is related to the optical flow matching nature of honeybee visual flight control [32, 36]. By optical flow we mean the observed visual gradient in time due to the relative motion (of the honeybee) and the objects in the scene. It has been shown that a honeybee navigates by matching the optical flow of the left and right eyes, which suggests that a honeybee has a mechanism and tendency to perform horizontal centering, but not one for vertical centering. This behavior is also visible in our 100 sets of honeybees' collision-avoidance trajectories. If we project all data points from the trajectories onto XY-plane, we find that the mean of all Y values is 1.13, with a standard deviation 12.49. A plot of this

projection is shown in Figure 3.5.

In this hypothesis, the reward function is the summation of the collision cost, the movement cost, and the penalty cost for moving close to the left or right walls.

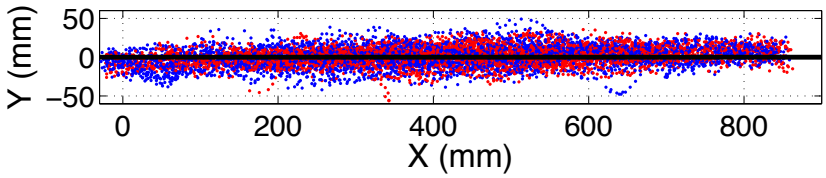


Figure 3.5 Data points of the 100 encounters are projected to XY-plane. Red points are projected data points from incoming honeybees, while blue points are projected data points from outgoing honeybees.

- H_{UD} is the hypothesis that the honeybees tend to perform vertical centering. This is actually an incorrect hypothesis we set to verify that the system can delineate bad hypotheses. The honeybees are actually biased to fly in the upper half of the tunnel because they are attracted to light. The transparent roof and solid bottom means more light is coming from the top. This behavior is evident from the 100 recorded collision-avoidance trajectories of honeybees. If we project all data points from the trajectories onto the XZ-plane, we find that 80% of the data points lies in the upper side of the tunnel. In fact, the Z values of all the data points have a mean of 12.89, a median of 15.67 and a standard deviation of 16.76, which again confirms the biased distribution toward the ceiling of the tunnel. A plot of this projection is shown in Figure 3.6.

  In this hypothesis, the reward function is the summation of collision cost, movement cost, and the penalty cost for moving close to the top or bottom walls.
- H_{LRUD} is the combination of the previous two hypotheses: H_{LR} and H_{UD}. The reward function is the summation of collision cost, movement cost, the penalty cost for moving near the left and right walls, and the penalty cost for moving near the top and bottom walls.

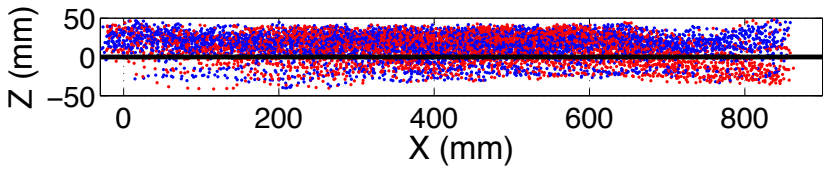


Figure 3.6 Data points of the 100 encounters, projected on to XZ-plane. Red points are projected data points from incoming honeybees, while blue points are projected data points from outgoing honeybees.

- H_{LRDest} is the combination of H_{LR} and $H_{BasicDest}$, i.e., the reward function is the summation of collision cost, movement cost, the penalty cost for moving near the left and right walls, and the reward for reaching the goal. This hypothesis is the closest to the correct performance criteria, based on the existing literature.

Table 4.1 presents a summary of these hypotheses.

Table 3.1 Hypotheses with the Corresponding Component Cost/Reward Functions

Penalties or Rewards	Hypotheses					
	H_{Basic}	$H_{BasicDest}$	H_{LRUD}	H_{LR}	H_{UD}	H_{LRDest}
Collision Cost	✓	✓	✓	✓	✓	✓
Movement Cost	✓	✓	✓	✓	✓	✓
LR-Penalty	—	—	✓	✓	—	✓
UD-Penalty	—	—	✓	—	✓	—
Destination Reward	—	✓	—	—	—	✓

3.4.3 Simulation Setup

We use POMCP [33] to generate near optimal solutions to the POMDP problem that represents each of the hypotheses. POMCP is an online POMDP solver, which means it will plan for the best action to perform at each step, execute that action, and then re-plan. The on-line computation of POMCP helps to alleviate the problem with long planning horizon problem of this application — since bees can see relatively far, the collision-avoidance manoeuvring may happen far before the close encounter scenario actually happens. In our experiments, POMCP was run with 8,192 particles.

For each hypothesis, we generate 36 sets of collision-avoidance trajectories. Each of these sets of trajectories consists of 100 collision-avoidance trajectories, resulting in a total of 3,600 simulated collision-avoidance trajectory for each hypothesis. Each trajectory corresponds to exactly one of the encounter scenarios in the initial trajectory data gathered from the experiments with real honeybees. In each of the simulated collision-avoidance trajectories, our system generates the outgoing honeybee's trajectory based on the POMDP policy and sets the incoming bee to move following the incoming bee in the corresponding real collision-avoidance trajectory. All experiments are carried out on a Linux platform with a 3.6GHz Intel Xeon E5-1620 and 16GB RAM.

Now, we need to set the parameters for the POMDP problems. To this end, we derive the parameters based on the experimental setup used to generate the initial trajectory data (described in the previous subsection) and from the statistical analysis of the data.

For the control parameters, we take the median over the velocity and acceleration of honeybees in our data and set $u = 300mm/s$, $a_m = 562.5mm/s^2$, and $\omega_m = 375deg/s$.

For the observation model, since honeybees can see far and the length of the tunnel is less than one meter, we set the range limit to D_R to be infinite, to model the fact that the range limit of the bee's vision will not hinder its ability to see the other bee. The viewing angle of the honeybees remain limited. We set the azimuth limit θ_a to be 60 degrees and the elevation limit θ_e to be 60 degrees. The bearing error standard deviation σ_b and the elevation error standard elevation σ_e are both set to be 1 degree. We assume that the false positive probability p_{fp} and the false negative probability p_{fn} are both 0.01.

Following the definition used by ethologists, a state $s = (x_1, y_1, z_1, \theta_1, u_1, v_1, x_2, y_2, z_2, \theta_2, u_2, v_2) \in \mathcal{S}$ is a collision state whenever the centre-to-centre distance between two parallel body axes is smaller than the wing span of the bee. Based on the ethologists' observations on average wingspan of a honeybee, we set this centre-to-centre distance to be 12mm. And we define a state to be in collision when the two honeybees are within

a cross-section distance (in YZ-plane) of 12mm and an axial distance (in X-direction) of 5mm, i.e., $\sqrt{(y_1 - y_2)^2 + (z_1 - z_2)^2} \leq 12$ and $\|x_1 - x_2\| \leq 5$.

As for the reward functions, we assign the following component costs and rewards as follows:

- Collision cost: $-10,000$.
- Movement cost: -10.
- LR-Penalty:
$$R_{LR}(s) = \begin{cases} 0 & \text{if } |y_1| \leq 12, \\ -20 \times \frac{|y_1|-12}{60-12} & \text{otherwise.} \end{cases}$$
- UD-Penalty:
$$R_{UD}(s) = \begin{cases} 0 & \text{if } |z_1| \leq 12, \\ -20 \times \frac{|z_1|-12}{50-12} & \text{otherwise.} \end{cases}$$
- Destination reward: $+10,000$.

The numbers are set based on the ethologists intuition on how important a particular criteria is.

One may argue that when the ethologists have little understanding on the underlying animal behavior, then setting the above values are impossible. Indeed setting the correct value is impossible. However, note that different cost and reward values can construct different hypothesis. And, one of the benefits of the system is exactly that the ethologists can simultaneously assess various hypotheses. Therefore, when the ethologists have little understanding, they can construct many hypotheses with different cost and reward values, and then use our *in-silico* behavior discovery system to identify hypotheses that are more likely to explain the input data better.

3.4.4 Results

Table 3.2 shows each component of the goodness of each hypotheses along with their 95% confidence intervals. It also shows the ranking of the hypotheses, where 1 means best.

Table 3.2 Hypotheses with corresponding rankings, where 1 indicates the most promising hypothesis. The observational bee data has a *collision rate* of 0.030 and an averaged *MED* of 30.61. Each metric value is the absolute difference of the corresponding metric values between the hypothesis and Bee. The value is in the format of mean and 95% confidence interval. The units for $\overline{F}$ and $\overline{M}$ are mm.

	Goodness of the hypotheses			
Hypotheses	$Average(\overline{F})$	$Average(\overline{M})$	$Average(C)$	Ranking
H_{Basic}	147.19 ± 0.833	33.02 ± 0.250	0.023 ± 0.0023	6
H_{UD}	149.19 ± 1.142	26.83 ± 0.266	0.021 ± 0.0036	5
H_{LRUD}	126.59 ± 1.057	12.11 ± 0.279	0.001 ± 0.0065	4
H_{LR}	121.22 ± 0.784	13.29 ± 0.260	0.006 ± 0.0042	3
$H_{BasicDest}$	115.47 ± 0.594	15.79 ± 0.257	0.002 ± 0.0049	2
H_{LRDest}	115.27 ± 0.678	11.49 ± 0.254	0.007 ± 0.0056	1

The results indicate that the *in-silico* behavior discovery system can identify the best hypothesis, i.e., the hypothesis that represents the performance criteria closest to that of a honeybee avoiding mid-air collision, which is maintaining its position to be at the center horizontally and reaching its destination (H_{LRDest}).

The results show that the ranking does indicate the known behavior of the honeybees. For instance, H_{LR} is ranked higher than H_{LRUD} and H_{LRUD} is ranked higher than H_{UD}, which means that the system can identify that horizontal centering is a criteria the honeybees try to achieve, but vertical centering is not, which conform to the widely known results as discussed in the Hypotheses subsection.

Furthermore, $H_{BasicDest}$ is ranked higher than H_{Basic}, which indicates that the system does identify that honeybees tend to remain focussed on reaching their destination even in head-on encounter scenarios, which conforms to the widely known results as discussed in the Hypotheses subsection.

3.5 Summary

This chapter presents an application of planning under uncertainty to help ethologists study the underlying performance criteria that animals try to optimize in an interaction. It is widely accepted that a variety of interaction strategies in animals achieve optimal

or near optimal performance, but determining the performance criteria being optimized remains a challenge. The main difficulty in this problem is the need to gather a large body of observational data to delineate hypotheses, which can be tedious and time consuming, if not impossible. This chapter presents a system — termed "*in-silico* behavior discovery" — that enables ethologists to simultaneously compare and assess various hypotheses with much less observational data, and therefore overcome the above challenge. Key to this system is the use of POMDPs to generate optimal strategies for various postulated hypotheses. Preliminary results indicate that, given various hypothesized performance criteria used by honeybees, our system can correctly identify and rank criteria according to how well their predictions fit the observed data. These results indicate that the system is feasible and may help ethologists in designing subsequent experiments or analysis that are much more focused, such that with a much smaller data set, they can reveal the underlying strategies in various animals' interaction. Such understanding may be beneficial to inspire the development of various technological advances.

Nature, additionally, involves a multi-objective optimization. Another strength of this approach is to help tease out the mixing of these objectives. For example, honeybee flight is regulated not only by optical flow, but also by overall illumination (i.e., phototaxis). As seen between the Central Tendency and Left/Right Central Tendency hypothesis test, the method can help clarify the weighting of the mixing (between optical flow and phototaxis). Another advantage of this approach is that it reduces the number of experiments where one has to hold other secondary conditions (e.g., temperature, food sources, etc.) stationary, thus saving time and further aiding discovery of the underlying behaviours.

Chapter 4

In-silico Behavior Discovery System: Extensive Validations

In previous chapter, we have verified that the *in-silico* behavior discovery system is able to independently discover that the *outgoing honeybee* has a horizontal centering tendency while at the same time may not have the vertical centering tendency. However, these results are still preliminary. In this chapter, we further verify the *in-silico* system with multiple sets of parameters for both the system modeling the *outgoing honeybee* as the rational agent and the system modeling the *incoming honeybee* as the rational agent. The results are promising, indicating that the *in-silico* system is able to independently identify that the horizontal centering tendency dominates the vertical centering tendency as discovered by biologists. Moreover, the system can even discover the subtle behavioral difference in the vertical centering tendencies between the incoming honeybee and the outgoing honeybee; that is, the outgoing honeybee is more easily affected by the translucent roof than the incoming honeybee. This tendency, amazingly, has already been verified by Menzel and Greggers's study being carried out 30 years ago.

4.1 Introduction

Many studies [10, 31, 37] have shown that when a honeybee flies through a narrow tunnel, it displays the *horizontal centering tendency*. As Srinivasan in [38] writes, "When a bee through a narrow passage, it positions itself such that both eyes experience approximately the same image velocity. This ensures that the two walls of the passage are at the same distance from the bee, enabling a collision-free flight through the middle of the tunnel." Other studies [4, 11, 24, 30] have shown that honeybees are phototactic, meaning that they are affected by light. Hence when a honeybee flies in a tunnel, due to the translucent roof, it tends to fly toward the roof and as a result does not display much of the *vertical centering tendency*.

This phenomenon had been independently discovered by the *in-silico* behavior discovery system in Chapter 3. However, the experimental results are still preliminary considering that:

1. The results are based on only one set of parameter setting, based on ethologists' understanding to the performance criteria.

2. The results are valid only for the *outgoing honeybee* as the *in-silico* system only models the outgoing honeybee as the rational agent.

To further verify the *in-silico* system, in this chapter, we not only use multiple sets of parameters, but also model the *incoming honeybee* as the rational agent. In particular, we validate the system by focusing on three aspects.

1. Would the system still be able to identify the dominating *horizontal centering tendency* as compared with the *vertical centering tendency* even when the ethologists do not have much prior knowledge on the performance criteria?

2. Would the system still be able to identify the dominating *horizontal centering tendency* as compared with the *vertical centering tendency* even it models the *incoming honeybee* as the rational agent?

3. Would the system be able to identify any behavioral difference between the two encountering honeybees in terms of relative weights between the *horizontal centering tendency* and the *vertical centering tendency*?

These three aspects are validated in the following three sections, respectively.

4.2 When the Agent is the Outgoing Honeybee

This section verifies the *in-silico* system with multiple sets of parameters for the outgoing honeybee. It focuses on discovering the relationship between the *horizontal centering tendency* and the *vertical centering tendency*.

In particular, we systematically adjust the weights for LR-Penalty (enforcing the horizontal centering tendency) and UD-Penalty (enforcing the vertical centering tendency), and under such circumstances, we verifty whether the *in-silico* system is still be able to independently discover the fact that the horizontal centering tendency is more important than the vertical centering tendency.

4.2.1 Generating Multiple Hypotheses

We generate multiple hypotheses in a way of systematically adjusting the weights for LR-Penalty and UD-Penalty, while holding other components as constants. In particular, we can either adjust the weights only for LR-Penalty, or only for UD-Penalty, or for both LR-Penalty and UD-Penalty. The rest of the parameters are set to be the same values as in Chapter 3.

Setting LR-Penalty Only

In this case, we assign various values to LR-Penalty, while keeping the parameter values of the rest components the same as in previous chapter. The set of possible values for LR-Penalty are: -1, -100, -500, $-5,000$, and $-10,000$.

Under each assignment of LR-Penalty, there are 4 possible hypotheses: H_{LR}, H_{LRDest}, H_{LRUD}, and $H_{LRUDDest}$. Table 4.1 shows the consisting components of the reward function for each hypothesis.

For each assignment of LR-Penalty, we use notation $LRPX/$ to represent that specific category of hypotheses, where X is the corresponding parameter value. For example, for assignment -100, we represent the 4 hypotheses in that category as: $H_{LRP100/LR}$, $H_{LRP100/LRDest}$, $H_{LRP100/LRUD}$, and $H_{LRP100/LRUDDest}$. As a result, there are 20 hypotheses in this setting; and among them, only in category $LRP1$, the LR-Penalty is less important than the UD-Penalty, as -1 penalizes less than -20; in the rest of the category, namely $LRP100$, $LRP500$, $LRP5000$, and $LRP10000$, the LR-Penalty is more important then UD-Penalty.

Setting UD-Penalty Only

In this case, we assign various values to UD-Penalty, while keeping the parameter values of the rest components the same as in previous chapter. The set of possible values for UD-Penalty are: -1, -100, -500, $-5,000$, and $-10,000$.

Under each assignment of UD-Penalty, there are 4 possible hypotheses: H_{UD}, H_{UDDest}, H_{LRUD}, and $H_{LRUDDest}$. Table 4.1 shows the consisting components of the reward function for each hypothesis.

For each assignment of UD-Penalty, we use notation $UDPX/$ to represent that specific category of hypotheses, where X is the corresponding parameter value. For example, for assignment -100, we represent the 4 hypotheses in that category as: $H_{UDP100/UD}$, $H_{UDP100/UDDest}$, $H_{UDP100/LRUD}$, and $H_{UDP100/LRUDDest}$. As a result, there are 20 hypotheses in this setting; and among them, only in category $UDP1$, the UD-Penalty is less important than the LR-Penalty, as -1 penalizes less than -20; in the rest of the category, namely $UDP100$, $UDP500$, $UDP5000$, and $UDP10000$, the UD-Penalty is more important then LR-Penalty.

Setting Both LR-Penalty and UD-Penalty

In this case, we assign various values to both LR-Penalty and UD-Penalty, while keeping the parameter values of the rest components the same as in previous chapter. The set of possible values for both LR-Penalty and UD-Penalty are: -1, -100, -500, $-5,000$, and $-10,000$.

Under each assignment of LR-Penalty and UD-Penalty, there are 2 possible hypotheses: H_{LRUD} and $H_{LRUDDest}$. Table 4.1 shows the consisting components of the reward function for each hypothesis.

For each assignment of LR-Penalty and UD-Penalty, we use notation $LRUDPX/$ to represent that specific category of hypotheses, where X is the corresponding parameter value. For example, for assignment -100, we represent the 2 hypotheses in that category as $H_{LRUDP100/LRUD}$ and $H_{LRUDP100/LRUDDest}$. As a result, there are 10 hypotheses in this setting, in which the LR-Penalty and the UD-Penalty share the same importance as they are assigned the same values.

Table 4.1 Hypotheses with the corresponding reward functions and consisting components.

Hypotheses	Penalties or Rewards				
	Collision cost	Movement cost	LR-Penalty	UD-Penalty	Destination Reward
H_{LR}	✓	✓	✓	—	—
H_{UD}	✓	✓	—	✓	—
H_{LRUD}	✓	✓	✓	✓	—
H_{LRDest}	✓	✓	✓	—	✓
H_{UDDest}	✓	✓	—	✓	✓
$H_{LRUDDest}$	✓	✓	✓	✓	✓

To this end, we have constructed 50 hypotheses in total to test, including 20 hypotheses from category LRP, 20 hypotheses from category UDP, and 10 hypotheses from category $LRUDP$.

4.2.2 Results

Table 4.2 lists the top 10 ranked hypotheses generated by the *in-silico* behavior discovery system for the above 50 hypotheses, where a smaller rank indicates a more promising hypothesis and 1 means the most promising hypothesis.

Table 4.2 Top 10 ranked hypotheses for the outgoing honeybee, where 1 in the ranking column indicates the most promising hypothesis. LR-Penalty and UD-Penalty are the parameters in hypotheses to enforce the horizontal centering tendency and the vertical centering tendency respectively. A larger parameter value gives more weights on the enforcement of the tendency.

Hypotheses	LR-Penalty	UD-Penalty	Ranking
$H_{LRP500/LRDest}$	-500	0	1
$H_{LRP5000/LRUDDest}$	$-5,000$	-20	1
$H_{LRP100/LRUDDest}$	-100	-20	1
$H_{LRP10000/LRUDDest}$	$-10,000$	-20	1
$H_{LRP500/LRUDDest}$	-500	-20	5
$H_{LRP10000/LRDest}$	$-10,000$	0	5
$H_{LRP5000/LRDest}$	$-5,000$	0	7
$H_{LRP5000/LRUD}$	$-5,000$	-20	8
$H_{LRP500/LR}$	-500	0	9
$H_{LRP100/LRDest}$	-100	0	9
$H_{LRUDP100/LRUDDest}$	-100	-100	9

Among the top ranked hypotheses, except that the hypothesis $H_{LRUDP100/LRUDDest}$ has equal values on both LR-Penalty and UD-Penalty, the rest of the hypotheses are all LRP based, i.e., they have a much larger LR-Penalty with a realatively smaller UD-Penalty. This result indicates that the horizontal centering is more important than the vertical centering. As a matter of fact, if we take a look at the complete ranking results as shown in Table A.1 in Appendix A, we find that the best UDP based hypothesis – the hypothesis in which the vertical centering tendency weights more than the horizontal centering tendency – is ranked at 19th place in the ranking spectrum, after 18 LRP or $LRUDP$ based hypotheses, suggesting that those hypotheses enforcing more vertical centering than horizontal centering are ranked relatively lower, which in turn proves that the horizontal centering tendency dominates the vertical centering tendency. Therefore, the *in-silico* behavior discovery system for the outgoing honeybee

can independently identify the correct relationship between the horizontal centering tendency and the vertical centering tendency as discoverd by many previous studies (see the detailed discussion in Section 4.1).

In fact, if we project all data points of the 100 enounter trajectories of the outgoing honeybees to the XY-plane (Figure 4.1), we found that these points are nearly symmetrically distributed around the horizontal center line, indicating that the outgoing honeybees have horizontal centering tendencies. Similarly, if we project all data points of the 100 enounter trajectories of the outgoing honeybees to the XZ-plane (Figure 4.2, we found that these points are biased distributed toward the roof, as the z value has a mean $14.87mm$, a median $17.35mm$, and a standard deviation $16.03mm$. This indicates that the outgoing honeybees have tendencies to fly toward roof; in other words, they do not display so much vertical centering tendencies.

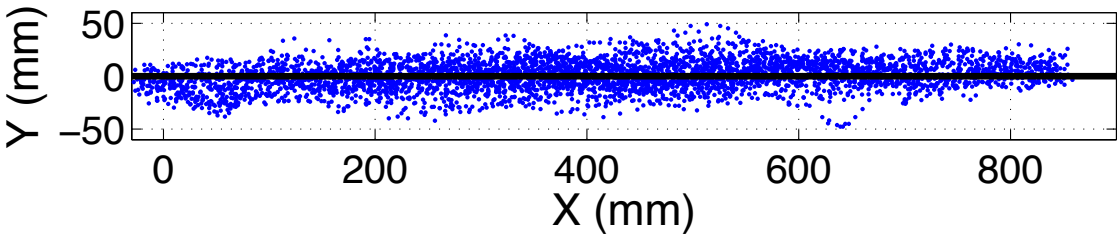


Figure 4.1 All data points of the 100 enounter trajectories of the outgoing honeybees are projected to the XY-plane. The y value has a mean $0.79mm$, a median $0.86mm$, and a standard deviation $13.45mm$.

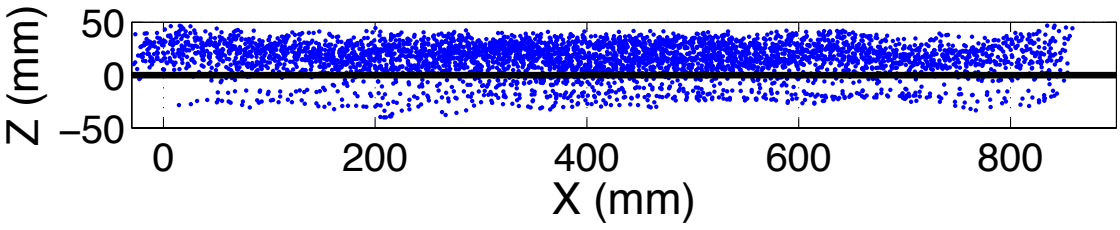


Figure 4.2 All data points of the 100 enounter trajectories of the outgoing honeybees are projected to the XZ-plane. The z value has a mean $14.87mm$, a median $17.35mm$, and a standard deviation $16.03mm$.

Interestingly, among the highest ranked hypotheses, three out of four consist of both LR-Penalty and UD-Penalty, thought the UD-Penalty is much smaller than the

LR-Penalty, indicating that the combination of horizontal centering tendency and *a less weighted* vertical centering tendency achieves the best ranking, despite the fact that the vertical centering is not that true.

4.3 When the Agent is the Incoming Honeybee

This section verifies the *in-silico* system with multiple set of parameters for the incoming honeybee. It focuses on discovering the relationship between the *horizontal centering tendency* and the *vertical centering tendency* for the incoming honeybee.

4.3.1 Generating Multiple Hypotheses

We generate multiple hypotheses for the incoming honeybee in the same way as in previous section for the outgoing honeybee and use the same notations except that we add an extra letter '2' at the end of each notation to denote that the particular hypothesis is for the incoming honeybee.

4.3.2 Results

Table 4.3 lists the top 10 ranked hypotheses generated by the *in-silico* behavior discovery system for the above 50 hypotheses, where a smaller rank indicates a more promising hypothesis and 1 means the most promising hypothesis.

Among the top 10 ranked hypotheses, nearly half them are $LRUDP$ based, meaning that the LR-Penalty and UD-Penalty share the same importance, while the other half are LRP based, in which the LR-Penalty weights much more than then UD-Penalty. It is worth noting that among these top ranked hypotheses, no one weights the UD-Penalty more than the LR-Penalty, indicating that the horizontal centering tendency for the incoming honeybee is more important than the vertical centering tendency. In fact, if we take look at the complete ranking results for the incoming honeybee in Table A.2 in Appendix A, we may find that the best ranking for UDP based hypotheses, i.e., the hypotheses in which the vertical centering tendency weights more than the

Table 4.3 Top 10 ranked hypotheses for the incoming honeybee, where 1 in the ranking column indicates the most promising hypothesis. LR-Penalty and UD-Penalty are the parameters in hypotheses to enforce the horizontal centering tendency and the vertical centering tendency respectively. A larger parameter value gives more weights on the enforcement of the tendency.

Hypotheses	LR-Penalty	UD-Penalty	Ranking
$H_{LRUDP100/LRUDDest2}$	-100	-100	1
$H_{LRUDP500/LRUDDest2}$	-500	-500	2
$H_{LRUDP5000/LRUDDest2}$	$-5,000$	$-5,000$	2
$H_{LRUDP10000/LRUDDest2}$	$-10,000$	$-10,000$	4
$H_{LRP100/LRUDDest2}$	-100	-20	5
$H_{LRP10000/LRUDDest2}$	$-10,000$	-20	5
$H_{LRUDP5000/LRUD2}$	$-5,000$	$-5,000$	7
$H_{LRP5000/LRDest2}$	$-5,000$	0	7
$H_{LRP500/LRUDDest2}$	-500	-20	9
$H_{LRP5000/LRUD2}$	$-5,000$	-20	9

horizontal centering tendency, is at the 20th place (the $H_{UDP100/LRUDDest2}$), after 19 LRP or $LRUDP$ based hypotheses in the ranking spectrum, demonstrating that those hypotheses with more vertical centering tendency than the horizontal centering tendency are not ranked at the top, and this provides more evidence that for the incoming honeybee, the horizontal centering tendency dominates the vertical centering tendency.

Therefore, the *in-silico* behavior discovery system for the incoming honeybee can also independently identify the correct relationship between the horizontal centering tendency and the vertical centering tendency as discoverd by previous studies (see the detailed discussion in Section 4.1).

Indeed, if we project all data points of the 100 enounter trajectories of the incoming honeybees to the XY-plane (Figure 4.3), we found that these points are approximately symmetrically distributed around the horizontal center line, indicating that the incoming honeybees have horizontal centering tendencies. Similarly, if we project all data points of the 100 enounter trajectories of the incoming honeybees to the XZ-plane (Figure 4.4, we found that these points are biased distributed toward the roof, as the z value has a mean $10.92mm$, a median $13.21mm$, and a standard deviation $17.22mm$. This

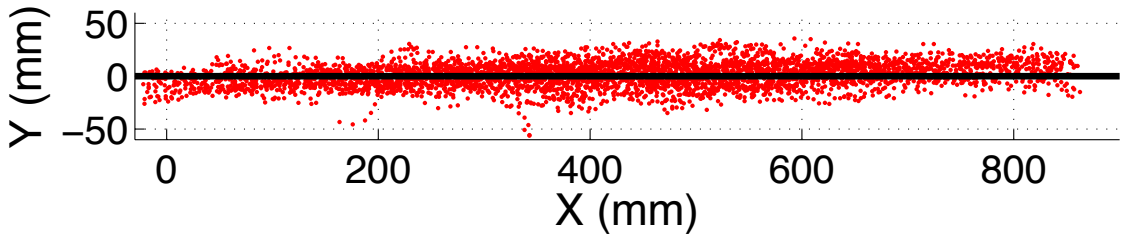


Figure 4.3 All data points of the 100 enounter trajectories of the incoming honeybees are projected to the XY-plane. The y value has a mean $1.47mm$, a median $1.66mm$, and a standard deviation $11.44mm$.

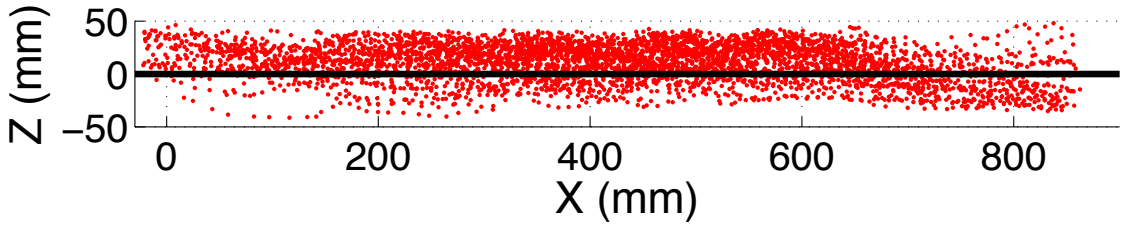


Figure 4.4 All data points of the 100 enounter trajectories of the incoming honeybees are projected to the XZ-plane. The z value has a mean $10.92mm$, a median $13.21mm$, and a standard deviation $17.22mm$.

indicates that the incoming honeybees have tendencies to fly toward roof; in other words, they do not display so much vertical centering tendencies.

4.4 Comparative Study

If we compare the top hypotheses ranking results of the outgoing honeybee with the results of the incoming honeybee, we may find that there are nearly half of the hypotheses are $LRUDP$ based in the results of the incoming honeybee, compared with only one hypothesis is $LRUDP$ based in the results of the outgoing honeybee. This indicates that the incoming honeybee puts more weights on the UD-Penalty, implying that the incoming honeybee has a stronger tendency on the vertical centering than the outgoing honeybee.

Amazingly, this result is in line with the result in [24] done 30 years ago, in which the biologists verified that when honeybees leave a dark food source and prepare to fly back

to the hive, they are positively phototactic, meaning that they are attracted to light. In our case, when the outgoing honeybees leaves the food and starts flying back, due to its positive phototaxis, it is more easily affected by the translucent roof and therefore flies closer to the roof than the incoming honeybee. Consequently, the incoming honeybee shows a stronger tendency on the vertical centering than the outgoing honeybee.

This is an interesting discovery, especially considering the fact that the difference in the behavior of both encountering honeybees is so *subtle* that it may not be noticed by normal human observers by just looking at the resulting trajectory data – it needs much more expertise plus carefully designed experiments to be disclosed.

As a matter of fact, even if we project all the 100 encounter trajectories of the incoming honeybees as well as the 100 encounter trajectories of the outgoing honeybees into the XZ-plane, in Figure 4.5 and Figure 4.6, respectively, we find that the difference is not that obvious, especially when we take the statistic significance into considerations.

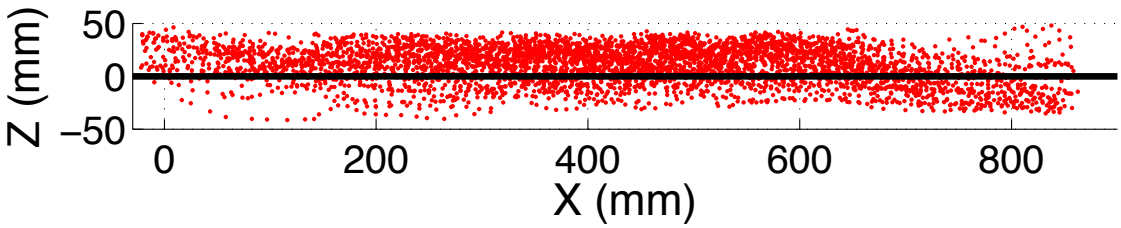


Figure 4.5 All data points of the 100 enounter trajectories of the incoming honeybees are projected to the XZ-plane. The z value has a mean $10.92mm$, a median $13.21mm$, and a standard deviation $17.22mm$.

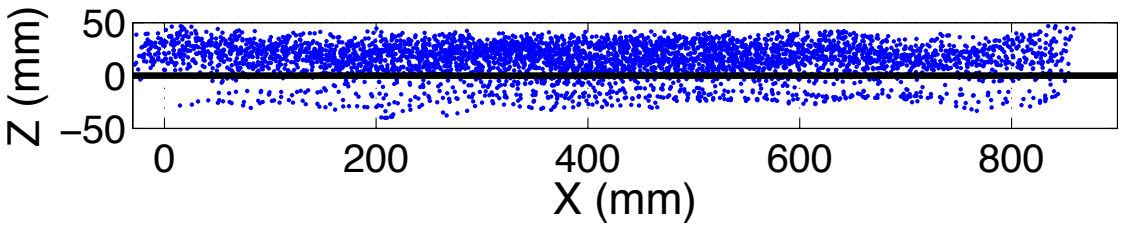


Figure 4.6 All data points of the 100 enounter trajectories of the outgoing honeybees are projected to the XZ-plane. The z value has a mean $14.87mm$, a median $17.35mm$, and a standard deviation $16.03mm$.

However, in our *in-silico* system, such subtle difference becomes evident when we focus on the comparison of the top ranked hypotheses. This is yet another evidence on the usefulness as well as validness of the *in-silico* behavior discover system in that it is capable of discovering subtle behaviors that are beyond the discernment of human observers. Such hidden-behavior-discovery capabilities could provide new insights to ethologists to enable them better understand animal behaviors.

4.5 Summary

In the first part of this chapter, we have evaluated the *in-silico* behavior discovery system with multiple sets of parameters for both the system modeling the outgoing honeybee as the rational agent and the system modeling the incoming honeybee as the rational agent. The results were promising, indicating that both system can independently identify that the horizontal centering tendency dominates the vertical centering tendency as discovered by biologists.

It is worth noting that the multiple sets of parameters were generated in a systematically manner, only based on the consisting components of performance criteria. This method has the potential to expand the capabilities of *in-silico* system in that the ethologists only need to provide a set of consisting components concerning the performance criteria as input, instead of providing all testing hypotheses, as which could be tedious and sometimes even may not be possible due to:

1. The ethologists may not have prior knowledge on the performance criteria.

2. The ethologists may have some knowledge on the performance criteria, but they are able to convert those knowledge into appropriate parameters values.

Therefore, this systematically weighting method can reduce the dependence on ethologists' understanding to the performance criteria and enable the ethologists focus on only the consisting components of the criteria, thereby aiding their discoveries on animal behavior.

In the later part of this chapter, we have evaluated the *in-silico* behavior discovery system with a comparative study on the ranking results of both systems in the earlier part. The results demonstrated that the *in-silico* system is capable of discovering the subtle behavioral difference in the vertical centering tendencies between the incoming honeybee and the outgoing honeybee, which, amazingly, had already been verified in a study done by biologists 30 years ago. This indicates that the *in-silico* behavior discovery system is capable of providing new insights into the behavior discovery process and aiding ethologists to understand the underlying behavior better.

Overall, these extended validations provide a strong evidence on the usefulness and validness of the *in-silico* behavior discovery system.

Chapter 5

Conclusions

It is now widely accepted that animal behavior optimizes certain performance criteria. The challenge is in identifying which criteria are being optimized? To conquer this challenge and alleviate the difficulties in conventional methods, We proposed an *in-silico* system – termed "*in-silico* behavior discovery" – to explore the feasibility of performing an *in silico* study on the underlying performance criteria. Experimental results were encouraging, demonstrating that the *in-silico* system was able to independently identify the same set of most promising hypotheses as discovered by ethologists. Moreover, the system was tested with an extensive set of hypotheses generated from multiple sets of parameters. Even in this case, the *in-silico* system still consistently identified the same set of hypotheses embodied the known facts as discovered by ethologists. Last but not least, the system discovered subtle behaviors that are beyond the discernment of human observers, implying that the system is capable of providing new insights to ethologists and helping them discover hidden behaviors.

The above results demonstrate that it is feasible to perform an in silico study to the underlying performance criteria toward the understanding of animal behaviors, and the *in-silico* behavior discovery system could help such studies.

Note that this *in-silico* behavior discovery system is not trying to replace ethologists and study the underlying performance criteria directly; quite the contrary, it tries to *aid* ethologists to study these performance criteria so that the discovery process

could be easier and faster. In particular, the *in-silico* behavior discovery system allows ethologists simultaneously compare and assess various hypotheses, therefore enables them better focus on the subsequent experiments and would enable them understand animal behavior with fewer animal trials.

References

[1] R. M. Alexander. *Animal Mechanics*. University of Washington Press, Seattle, 1969.

[2] H. Bai, D. Hsu, M. J. Kochenderfer, and W. S. Lee. Unmanned aircraft collision avoidance using continuous-state pomdps. *Robotics: Science and Systems VII*, page 1, 2012.

[3] H. Bai, D. Hsu, W. S. Lee, and V. A. Ngo. Monte carlo value iteration for continuous-state pomdps. In *Algorithmic Foundations of Robotics IX*, pages 175–191. Springer, 2011.

[4] Y. Ben-Shahar, H.-T. Leung, W. Pak, M. Sokolowski, and G. Robinson. cgmp-dependent changes in phototaxis: a possible role for the foraging gene in honey bee division of labor. *Journal of Experimental Biology*, 206(14):2507–2515, 2003.

[5] M. D. Breed and J. Moore. *Animal Behavior*. Academic Press, Burlington, San Diego, London, 2012.

[6] E. W. Chambers, É. Colin de Verdière, J. Erickson, S. Lazard, F. Lazarus, and S. Thite. Walking your dog in the woods in polynomial time. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 101–109. ACM, 2008.

[7] E. W. Chambers, É. Colin de Verdière, J. Erickson, S. Lazard, F. Lazarus, and S. Thite. Homotopic fréchet distance between curves or, walking your dog in the woods in polynomial time. *Computational Geometry*, 43(3):295–311, 2010.

[8] J. P. Chryssanthacopoulos and M. J. Kochenderfer. Accounting for state uncertainty in collision avoidance. *Journal of Guidance, Control, and Dynamics*, 34(4):951–960, 2011.

[9] N. B. Davies and J. R. Krebs. *An Introduction to Behavioural Ecology*. Wiley-Blackwell, Chichester, 2012.

[10] J. P. Dyhr and C. M. Higgins. The spatial frequency tuning of optic-flow-dependent behaviors in the bumblebee bombus impatiens. *The Journal of experimental biology*, 213(10):1643–1650, 2010.

[11] J. Erber, J. Hoormann, and R. Scheiner. Phototactic behaviour correlates with gustatory responsiveness in honey bees (apis mellifera l.). *Behavioural brain research*, 174(1):174–180, 2006.

[12] Groening, Julia and McLeod, Laura and Liebsch, Nikolai and Schiffner, Ingo and Srinivasan, Mandyam V. When left is right and right is wrong: Collision avoidance in honeybees. *Frontiers in Behavioral Neuroscience*, 2012.

[13] M. Horowitz and J. Burdick. Interactive Non-Prehensile Manipulation for Grasping Via POMDPs. In *ICRA*, 2013.

[14] K. Hsiao, L. Kaelbling, and T. Lozano-Perez. Grasping POMDPs. In *ICRA*, pages 4685–4692, 2007.

[15] L. Kaelbling, M. Littman, and A. Cassandra. Planning and acting in partially observable stochastic domains. *AI*, 101:99–134, 1998.

[16] M. J. Kochenderfer and J. P. Chryssanthacopoulos. Robust airborne collision avoidance through dynamic programming. *Massachusetts Institute of Technology, Lincoln Laboratory, Project Report ATC-371*, 2011.

[17] M. Koval, N. Pollard, and S. Srinivasa. Pre- and post-contact policy decomposition for planar contact manipulation under uncertainty. In *RSS*, Berkeley, USA, July 2014.

[18] H. Kurniawati, T. Bandyopadhyay, and N. Patrikalakis. Global motion planning under uncertain motion, sensing, and environment map. *Autonomous Robots: Special issue on selected papers from RSS 2011*, 30(3), 2012.

[19] H. Kurniawati, Y. Du, D. Hsu, and W. Lee. Motion planning under uncertainty for robotic tasks with long time horizons. *IJRR*, 30(3):308–323, 2011.

[20] H. Kurniawati, D. Hsu, and W. S. Lee. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In *Proc. Robotics: Science and Systems*, 2008.

[21] J.-C. Latombe. Robot motion planning. *Kluwer Academic Publishers*, 1991.

[22] J. Levin and B. Nalebuff. An introduction to vote-counting schemes. *The Journal of Economic Perspectives*, pages 3–26, 1995.

[23] K. Y. Ma, P. Chirarattananon, S. B. Fuller, and R. J. Wood. Controlled flight of a biologically inspired, insect-scale robot. *Science*, 340(6312):603–607, 2013.

[24] R. Menzel and U. Greggers. Natural phototaxis and its relationship to colour vision in honeybees. *Journal of Comparative Physiology A*, 157(3):311–321, 1985.

[25] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231–268, 2001.

[26] C. Papadimitriou and J. Tsitsiklis. The Complexity of Markov Decision Processes. *Math. of Operation Research*, 12(3):441–450, 1987.

[27] J. Pineau, G. Gordon, and S. Thrun. Point-based value iteration: An anytime algorithm for POMDPs. In *IJCAI*, pages 1025–1032, 2003.

[28] M. Raibert, K. Blankespoor, G. Nelson, R. Playter, and the BigDog Team. Bigdog, the rough-terrain quadruped robot. In *Proc. of the 17th IFAC World Congress*, 2008.

[29] I. Ros, L. Bassman, M. Badger, A. Pierson, and A. Biewener. Pigeons steer like helicopters and generate down-and upstroke lift during low speed turns. *Proceedings of the National Academy of Sciences*, 108(50):19990–19995, 2011.

[30] R. Scheiner, C. I. Abramson, R. Brodschneider, K. Crailsheim, W. M. Farina, S. Fuchs, B. Grünewald, S. Hahshold, M. Karrer, G. Koeniger, et al. Standard methods for behavioural studies of apis mellifera. *Journal of Apicultural Research*, 52(4):1–58, 2013.

[31] J. R. Serres, G. P. Masson, F. Ruffier, and N. Franceschini. A bee in the corridor: centering and wall-following. *Naturwissenschaften*, 95(12):1181–1187, 2008.

[32] A. Si, M. V. Srinivasan, and S. Zhang. Honeybee navigation: properties of the visually drivenodometer'. *Journal of Experimental Biology*, 206(8):1265–1273, 2003.

[33] D. Silver and J. Veness. Monte-Carlo planning in large POMDPs. In *NIPS*, 2010.

[34] K. A. Smith, M. Kochenderfer, W. A. Olson, and A. E. Vela. *Collision Avoidance System Optimization for Closely Spaced Parallel Operations through Surrogate Modeling*. American Institute of Aeronautics and Astronautics, 2013.

[35] T. Smith and R. Simmons. Point-based POMDP algorithms: Improved analysis and implementation. In *UAI*, July 2005.

[36] M. Srinivasan, S. Zhang, M. Lehrer, and T. Collett. Honeybee navigation en route to the goal: visual flight control and odometry. *Journal of Experimental Biology*, 199(1):237–244, 1996.

[37] M. V. Srinivasan. Honeybees as a model for the study of visually guided flight, navigation, and biologically inspired robotics. *Physiological Reviews*, 91(2):389–411, 2011.

[38] M. V. Srinivasan, R. J. Moore, S. Thurrowgood, D. Soccol, and D. Bland. *From biology to engineering: insect vision and applications to robotics*. Springer, 2012.

[39] S. Temizer, M. J. Kochenderfer, L. P. Kaelbling, T. Lozano-Pérez, and J. K. Kuchar. Collision avoidance for unmanned aircraft using markov decision processes. In *Proc. AIAA Guidance, Navigation, and Control Conference*, 2010.

[40] S. Thurrowgood, R. J. Moore, D. Soccol, M. Knight, and M. V. Srinivasan. A biologically inspired, vision-based guidance system for automatic landing of a fixed-wing aircraft. *Journal of Field Robotics*, 31(4):699–727, 2014.

[41] M. S. Triantafyllou and G. S. Triantafyllou. An efficient swimming machine. *Scientific American*, 272(3):64–70, 1995.

[42] H. Wang, H. Kurniawati, S. P. N. Singh, and M. V. Srinivasan. Animal Locomotion In-Silico: A POMDP-Based Tool to Study Mid-Air Collision Avoidance Strategies in Flying Animals. In *Proc. Australasian Conference on Robotics and Automation*, 2013.

[43] H. Wang, H. Kurniawati, S. P. N. Singh, and M. V. Srinivasan. In-silico Behavior Discovery System: An Application of Planning in Ethology. In *Proc. International Conference on Automated Planning and Scheduling (ICAPS)*, 2015.

[44] J. D. Williams and S. Young. Scaling POMDPs for spoken dialog management. *IEEE Trans. on Audio, Speech, and Language Processing*, July 2007.

Appendix A

Complete Ranking Results

A.1 Results When The Agent Is The Outgoing Honeybee

Table A.1 shows the complete ranking results given by the *in-silico* behavior discovery system for the outgoing honeybee with 50 hypotheses, where a smaller rank indicates a more promising hypothesis and 1 means the most promising hypothesis.

Table A.1 Hypotheses for the outgoing honeybee and the corresponding rankings, where 1 indicates the most promising hypothesis. The value is in the format of mean and 95% confidence interval. The units for $\overline{F}$ and $\overline{M}$ are mm.

	Goodness of the hypotheses			
Hypotheses	$Average(\overline{F})$	$Average(\overline{M})$	$Average(C)$	Ranking
$H_{LRP500/LRDest}$	114.328 ± 0.6754	15.666 ± 0.1839	0.023 ± 0.0063	1
$H_{LRP5000/LRUDDest}$	115.219 ± 0.7754	15.451 ± 0.1909	0.018 ± 0.0047	1
$H_{LRP100/LRUDDest}$	114.446 ± 0.6945	15.200 ± 0.1811	0.018 ± 0.0040	1
$H_{LRP10000/LRUDDest}$	115.667 ± 0.7743	15.277 ± 0.1862	0.023 ± 0.0058	1
$H_{LRP500/LRUDDest}$	116.271 ± 0.8572	15.280 ± 0.1825	0.017 ± 0.0036	5
$H_{LRP10000/LRDest}$	116.178 ± 0.7968	15.568 ± 0.1902	0.022 ± 0.0051	5

Continued on next page

Table A.1 – continued from previous page

Hypotheses	Goodness of the hypotheses			
	$Average(\overline{F})$	$Average(\overline{M})$	$Average(C)$	Ranking
$H_{LRP5000/LRDest}$	116.375 ± 0.7991	15.543 ± 0.1909	0.020 ± 0.0043	7
$H_{LRP5000/LRUD}$	116.612 ± 0.8018	15.976 ± 0.1966	0.018 ± 0.0055	8
$H_{LRP500/LR}$	117.179 ± 0.8404	16.248 ± 0.1938	0.019 ± 0.0042	9
$H_{LRP100/LRDest}$	114.632 ± 0.6922	16.003 ± 0.1861	0.019 ± 0.0040	9
$H_{LRUDP100/LRUDDest}$	117.374 ± 0.8493	15.843 ± 0.1902	0.015 ± 0.0044	9
$H_{LRP500/LRUD}$	118.819 ± 0.9698	15.853 ± 0.1945	0.017 ± 0.0058	12
$H_{LRP10000/LRUD}$	117.939 ± 0.8899	16.202 ± 0.1932	0.021 ± 0.0045	12
$H_{LRP10000/LR}$	118.138 ± 0.8708	15.974 ± 0.1946	0.020 ± 0.0059	12
$H_{LRUDP500/LRUDDest}$	114.602 ± 0.6868	16.535 ± 0.1906	0.016 ± 0.0032	15
$H_{LRP5000/LR}$	118.359 ± 0.9081	16.146 ± 0.1984	0.021 ± 0.0052	15
$H_{LRUDP5000/LRUDDest}$	114.602 ± 0.6868	16.535 ± 0.1906	0.016 ± 0.0032	15
$H_{LRUDP10000/LRUDDest}$	118.944 ± 0.9326	16.155 ± 0.1930	0.016 ± 0.0042	18
$H_{LRUDP500/LRUD}$	115.745 ± 0.6725	16.408 ± 0.1934	0.016 ± 0.0035	19
$H_{LRUDP5000/LRUD}$	115.745 ± 0.6725	16.408 ± 0.1934	0.016 ± 0.0035	19
$H_{UDP1/LRUD}$	115.745 ± 0.6725	16.408 ± 0.1934	0.016 ± 0.0035	19
$H_{UDP5000/LRUD}$	115.745 ± 0.6725	16.408 ± 0.1934	0.016 ± 0.0035	19
$H_{UDP500/LRUD}$	115.745 ± 0.6725	16.408 ± 0.1934	0.016 ± 0.0035	19
$H_{UDP10000/LRUD}$	115.745 ± 0.6725	16.408 ± 0.1934	0.016 ± 0.0035	19
$H_{UDP100/LRUD}$	115.745 ± 0.6725	16.408 ± 0.1934	0.016 ± 0.0035	19
$H_{LRP100/LR}$	117.449 ± 0.7740	16.769 ± 0.1901	0.018 ± 0.0050	26
$H_{LRP100/LRUD}$	119.285 ± 0.9876	15.865 ± 0.1947	0.021 ± 0.0049	27
$H_{UDP10000/LRUDDest}$	115.490 ± 0.5865	19.138 ± 0.1981	0.013 ± 0.0027	28
$H_{UDP5000/LRUDDest}$	115.490 ± 0.5865	19.138 ± 0.1981	0.013 ± 0.0027	28
$H_{LRP1/LRUDDest}$	115.490 ± 0.5865	19.138 ± 0.1981	0.013 ± 0.0027	28
$H_{UDP1/LRUDDest}$	115.490 ± 0.5865	19.138 ± 0.1981	0.013 ± 0.0027	28

Continued on next page

Table A.1 – continued from previous page

Hypotheses	Goodness of the hypotheses			
	$Average(\overline{F})$	$Average(\overline{M})$	$Average(C)$	Ranking
$H_{UDP100/LRUDDest}$	115.490 ± 0.5865	19.138 ± 0.1981	0.013 ± 0.0027	28
$H_{UDP500/LRUDDest}$	115.490 ± 0.5865	19.138 ± 0.1981	0.013 ± 0.0027	28
$H_{UDP5000/UDDest}$	115.667 ± 0.5762	19.078 ± 0.2025	0.012 ± 0.0030	34
$H_{UDP1/UD}$	116.244 ± 0.6300	19.206 ± 0.1998	0.011 ± 0.0022	34
$H_{UDP10000/UDDest}$	115.667 ± 0.5762	19.078 ± 0.2025	0.012 ± 0.0030	34
$H_{UDP10000/UD}$	116.244 ± 0.6300	19.206 ± 0.1998	0.011 ± 0.0022	34
$H_{UDP5000/UD}$	116.244 ± 0.6300	19.206 ± 0.1998	0.011 ± 0.0022	34
$H_{UDP500/UD}$	116.244 ± 0.6300	19.206 ± 0.1998	0.011 ± 0.0022	34
$H_{UDP1/UDDest}$	115.667 ± 0.5762	19.078 ± 0.2025	0.012 ± 0.0030	34
$H_{UDP100/UDDest}$	115.667 ± 0.5762	19.078 ± 0.2025	0.012 ± 0.0030	34
$H_{UDP100/UD}$	116.244 ± 0.6300	19.206 ± 0.1998	0.011 ± 0.0022	34
$H_{UDP500/UDDest}$	115.667 ± 0.5762	19.078 ± 0.2025	0.012 ± 0.0030	34
$H_{LRP1/LRDest}$	116.616 ± 0.5900	20.250 ± 0.2030	0.015 ± 0.0031	44
$H_{LRUDP1/LRUDDest}$	116.203 ± 0.6076	20.405 ± 0.2019	0.012 ± 0.0027	44
$H_{LRUDP100/LRUD}$	125.032 ± 1.0901	16.988 ± 0.2022	0.013 ± 0.0033	46
$H_{LRUDP10000/LRUD}$	127.473 ± 1.1686	17.111 ± 0.1997	0.017 ± 0.0050	47
$H_{LRUDP1/LRUD}$	142.112 ± 0.8389	30.212 ± 0.2294	0.020 ± 0.0024	48
$H_{LRP1/LR}$	140.970 ± 0.7714	31.611 ± 0.2206	0.022 ± 0.0025	49
$H_{LRP1/LRUD}$	146.867 ± 1.1475	27.813 ± 0.2250	0.019 ± 0.0031	50

A.2 Results When The Agent Is The Incoming Honeybee

Table A.2 shows the complete ranking results given by the *in-silico* behavior discovery system for the incoming honeybee with 50 hypotheses, where a smaller rank indicates a more promising hypothesis and 1 means the most promising hypothesis.

Table A.2 Hypotheses for the incoming honeybee and the corresponding rankings, where 1 indicates the most promising hypothesis. The value is in the format of mean and 95% confidence interval. The units for $\overline{F}$ and $\overline{M}$ are mm.

	Goodness of the hypotheses			
Hypotheses	$Average(\overline{F})$	$Average(\overline{M})$	$Average(C)$	Ranking
$H_{LRUDP100/LRUDDest2}$	94.534 ± 0.6291	14.783 ± 0.1892	0.010 ± 0.0027	1
$H_{LRUDP500/LRUDDest2}$	96.025 ± 0.7156	14.874 ± 0.1912	0.012 ± 0.0033	2
$H_{LRUDP5000/LRUDDest2}$	96.637 ± 0.8015	15.131 ± 0.1912	0.011 ± 0.0038	2
$H_{LRUDP10000/LRUDDest2}$	96.351 ± 0.7554	15.291 ± 0.1938	0.014 ± 0.0040	4
$H_{LRP100/LRUDDest2}$	95.750 ± 0.7995	15.237 ± 0.1864	0.011 ± 0.0026	5
$H_{LRP10000/LRUDDest2}$	96.043 ± 0.7957	15.223 ± 0.1902	0.011 ± 0.0028	5
$H_{LRUDP5000/LRUD2}$	97.668 ± 0.8285	15.555 ± 0.1986	0.012 ± 0.0032	7
$H_{LRP5000/LRDest2}$	95.769 ± 0.7952	15.308 ± 0.1894	0.015 ± 0.0033	7
$H_{LRP500/LRUDDest2}$	93.718 ± 0.5593	15.395 ± 0.1894	0.013 ± 0.0029	9
$H_{LRP5000/LRUD2}$	95.999 ± 0.8089	15.580 ± 0.1919	0.011 ± 0.0028	9
$H_{LRP10000/LRUD2}$	96.623 ± 0.7976	15.356 ± 0.1928	0.012 ± 0.0028	11
$H_{LRP5000/LR2}$	96.381 ± 0.8329	15.749 ± 0.1977	0.011 ± 0.0030	12
$H_{LRP500/LRUD2}$	95.836 ± 0.8311	15.605 ± 0.1961	0.010 ± 0.0026	12
$H_{LRP500/LRDest2}$	94.713 ± 0.5735	15.364 ± 0.1937	0.010 ± 0.0028	14
$H_{LRP5000/LRUDDest2}$	95.534 ± 0.7195	15.406 ± 0.1911	0.012 ± 0.0025	14
$H_{LRP10000/LRDest2}$	94.591 ± 0.6113	15.968 ± 0.1993	0.011 ± 0.0025	14
$H_{LRUDP500/LRUD2}$	99.410 ± 0.9267	15.617 ± 0.1960	0.011 ± 0.0031	14

Continued on next page

Table A.2 – continued from previous page

Hypotheses	Goodness of the hypotheses			
	$Average(\overline{F})$	$Average(\overline{M})$	$Average(C)$	Ranking
$H_{LRP10000/LR2}$	96.547 ± 0.8311	15.705 ± 0.2006	0.012 ± 0.0029	14
$H_{LRP500/LR2}$	96.470 ± 0.7374	16.395 ± 0.2012	0.009 ± 0.0025	19
$H_{LRP100/LRDest2}$	94.769 ± 0.5785	15.990 ± 0.1917	0.009 ± 0.0024	20
$H_{UDP100/LRUDDest2}$	96.311 ± 0.7606	16.780 ± 0.1977	0.012 ± 0.0030	20
$H_{LRP100/LRUD2}$	100.457 ± 1.0238	15.759 ± 0.2021	0.011 ± 0.0034	20
$H_{UDP100/UDDest2}$	96.589 ± 0.7066	18.243 ± 0.1924	0.016 ± 0.0029	23
$H_{UDP1/LRUDDest2}$	95.937 ± 0.6686	17.566 ± 0.1965	0.012 ± 0.0024	23
$H_{LRP100/LR2}$	101.250 ± 0.9937	16.919 ± 0.2039	0.010 ± 0.0024	23
$H_{LRP1/LRUDDest2}$	96.086 ± 0.7212	17.631 ± 0.1983	0.012 ± 0.0031	23
$H_{UDP1/UDDest2}$	95.577 ± 0.6286	19.333 ± 0.2105	0.013 ± 0.0030	27
$H_{LRP1/LRDest2}$	95.118 ± 0.5340	19.169 ± 0.2009	0.013 ± 0.0033	27
$H_{LRUDP10000/LRUD2}$	99.656 ± 0.9978	15.634 ± 0.1982	0.016 ± 0.0038	27
$H_{LRUDP100/LRUD2}$	103.574 ± 1.0855	15.599 ± 0.1958	0.009 ± 0.0031	30
$H_{LRUDP1/LRUDDest2}$	95.682 ± 0.5485	19.606 ± 0.2046	0.017 ± 0.0033	30
$H_{UDP500/LRUDDest2}$	99.330 ± 0.7510	19.268 ± 0.2012	0.011 ± 0.0030	32
$H_{UDP500/UDDest2}$	97.952 ± 0.6883	20.340 ± 0.2067	0.018 ± 0.0028	33
$H_{UDP5000/UDDest2}$	99.791 ± 0.7461	21.324 ± 0.2037	0.018 ± 0.0027	34
$H_{UDP10000/UDDest2}$	100.843 ± 0.8182	21.451 ± 0.2098	0.019 ± 0.0027	35
$H_{UDP5000/LRUDDest2}$	102.564 ± 0.8631	21.396 ± 0.2084	0.018 ± 0.0033	36
$H_{UDP10000/LRUDDest2}$	102.122 ± 0.8880	21.119 ± 0.2041	0.018 ± 0.0033	36
$H_{UDP500/LRUD2}$	108.501 ± 1.1039	21.570 ± 0.2121	0.015 ± 0.0028	38
$H_{UDP100/LRUD2}$	111.663 ± 1.2282	19.756 ± 0.2165	0.014 ± 0.0030	39
$H_{UDP5000/LRUD2}$	106.413 ± 0.9897	22.139 ± 0.2102	0.018 ± 0.0024	39
$H_{UDP10000/LRUD2}$	107.322 ± 1.0546	22.234 ± 0.2074	0.015 ± 0.0026	41
$H_{UDP5000/UD2}$	108.099 ± 1.0734	22.150 ± 0.2082	0.021 ± 0.0023	42

Continued on next page

Table A.2 – continued from previous page

Hypotheses	Goodness of the hypotheses			Ranking
	$Average(\overline{F})$	$Average(\overline{M})$	$Average(C)$	
$H_{UDP10000/UD2}$	106.963 ± 0.9550	22.488 ± 0.2084	0.019 ± 0.0027	42
$H_{UDP1/LRUD2}$	116.942 ± 1.3772	20.190 ± 0.2180	0.014 ± 0.0031	44
$H_{UDP500/UD2}$	111.206 ± 1.1091	22.921 ± 0.2080	0.019 ± 0.0031	45
$H_{UDP100/UD2}$	123.396 ± 1.4457	23.814 ± 0.2171	0.018 ± 0.0029	46
$H_{LRP1/LRUD2}$	143.856 ± 1.8434	24.343 ± 0.2245	0.022 ± 0.0029	47
$H_{LRUDP1/LRUD2}$	157.199 ± 1.8114	28.772 ± 0.2391	0.025 ± 0.0025	48
$H_{LRP1/LR2}$	155.518 ± 1.8001	30.124 ± 0.2580	0.024 ± 0.0026	49
$H_{UDP1/UD2}$	163.300 ± 1.8331	29.629 ± 0.2431	0.025 ± 0.0022	50