ELECTROPHYSIOLOGY, AGENT-BASED MODELING AND INVERSE OPTIMAL CONTROL APPLICATIONS IN NEUROETHOLOGY

BY

MARIANNE JANSEM CATANHO

THESIS
Submitted in partial fulfillment of the requirements for the degree of Master of Science in Electrical and Computer Engineering in the Graduate College of the University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Adviser:
Adjunct Assistant Professor Todd P. Coleman
Decision-making is based on the integration of sensory information, motivational states and memory, from which assessment of outcomes and optimal action selection emerge. In foraging, animals tend to make decisions that reconcile two core goals: energy maximization and time minimization. The predatory sea-slug *Pleurobranchaea californica* has a simple, accessible nervous system and exhibits a repertory of simple behavioral elements that constitute effective foraging strategies. In this thesis, we introduce initial findings of sensory neurons identified through intracellular staining and electrophysiology, potentially linked to central sensory processing in the *Pleurobranchaea*. To further expand the principles underlying sensory integration and behavioral selection, we implemented a multi-agent based NetLogo simulation to model autonomous decision-making in the predatory sea-slug *Pleurobranchaea*. In particular, the model incorporated cost-benefit decisions in foraging by integrating sensation, internal state and learning in the virtual agent, replicating the particular behavioral selection process of the *Pleurobranchaea*. Finally, we propose two Markov decision processes to model how the animal makes decisions in its environment. Given the observed behaviors, we utilize inverse optimal control to succinctly characterize a class of utility functions the animal is maximizing. This research methodology combines principles from neurophysiology, agent-based modeling, classical conditioning, Bayesian statistics and control theory to investigate foraging decisions of the *Pleurobranchaea*, as it integrates sensation, internal state and learning mechanisms.
“Where ignorance lurks, so too do the frontiers of discovery and imagination.”
— Neil deGrasse Tyson
ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my adviser Dr. Todd P. Coleman, and my co-adviser, Dr. Rhanor Gillette. Their overwhelming support, continued optimism, motivation and immense knowledge inspired me to pursue this work and fulfill the requirements of this degree. I sincerely appreciate the collaboration and support of my fellow labmates and colleagues: Jeff Brown, Ankur Kulkarni, Vanessa Noboa and Keiko Hirayama.

I also wish to thank my mother, Teresa, who never stopped fighting and whose faith in me taught me to have faith in myself. Last, but not least, I would like to express my unending gratitude to my friend, Paul, who taught me humility and generosity, gratitude and compassion; and showed me how to live a life worth not only remembering, but celebrating.

“Remain curious,
Study hard,
Live wisely,
Love sincerely.
Prosper.”

— Paul K. Shumaker
# TABLE OF CONTENTS

**CHAPTER 1 INTRODUCTION** .......................................................... 1

**CHAPTER 2 ANATOMY, BEHAVIOR AND SENSORY SYSTEM STUDIES OF THE *PLEUROBRANCHAEA CALIFORNICA*** 4

2.1 Introduction ................................................................................. 4

2.2 Life History, Anatomy and Physiology of *Pleurobranchaea californica* ................................................................. 5

2.2.1 Anatomy ........................................................................... 6

2.2.2 Behavioral Repertoire and Physiology .................................... 6

2.2.3 Neuronal Substrates Involved in Behavior Selection ................. 7

2.2.4 Sensory Processing .............................................................. 10

2.3 Determining Critical Neurons Involved in Sensory Processing .... 12

2.3.1 Intracellular Backfilling ........................................................ 12

2.3.2 Extracellular and Intracellular Recordings ................................ 13

2.4 Results .................................................................................. 13

2.4.1 Intracellular Backfilling ........................................................ 13

2.4.2 Extracellular and Intracellular Recordings ................................ 14

2.5 Discussion .............................................................................. 18

**CHAPTER 3 COMPUTATIONAL MODEL OF BEHAVIOR, PREDATION AND ADDICTION** ......................................................... 20

3.1 Introduction .............................................................................. 20

3.2 Cost-Benefit Neural Network Model ............................................ 21

3.2.1 Basics of Agent-Based Modeling ......................................... 23

3.2.2 Rescorla-Wagner ................................................................. 25

3.3 The NetLogo Model ................................................................. 25

3.3.1 Learning Mechanism ........................................................... 26

3.4 Modeling Addiction ................................................................. 31

3.4.1 Addiction Learning .............................................................. 33

3.5 Discussion .............................................................................. 35

**CHAPTER 4 PRINCIPLES OF OPTIMIZATION THEORY AND CONTROL THEORY TO INFERENCE THE COST FUNCTION OF FORAGING ATTEMPTS** ........................................................................ 37

4.1 Introduction .............................................................................. 37
4.2 Background .................................................. 38
  4.2.1 Probability Theory ...................................... 38
  4.2.2 Markov Chains ........................................... 40
  4.2.3 Markov Decision Processes .............................. 41
  4.2.4 Inverse Optimal Control ................................. 42
4.3 Problem Statement ............................................ 44
  4.3.1 Model I .................................................. 44
  4.3.2 Model II ................................................. 48
4.4 Simulations and Results ..................................... 50
  4.4.1 Model I .................................................. 51
  4.4.2 Model II ................................................. 53

CHAPTER 5 CONCLUSION ........................................... 56

REFERENCES ..................................................... 58
In foraging, animals tend to make decisions that reconcile two core goals: energy maximization and time minimization. These decisions are well characterized in terms of optimal foraging theory [1]. In energy maximization, animals are motivated to gain as much energy as possible during a certain window of time. Conversely, in time minimization, animals attempt to use the least amount of time possible to acquire a specific quantity of energy. In a general sense, we can infer that the nervous system is trying to maximize the utility of the foraging behavior. How are the decision rules that underlie foraging behavior implemented at the level of the brain and at the level of neural circuits?

For decades, neuroscientists have been using simple animal model systems to analyze the neuronal mechanisms underlying decision-making in animals [2, 3, 4]. In particular, certain molluscs are useful preparations for investigating this question because of their simple behavior and highly accessible nervous systems. In molluscs, as in mammals, decision-making is based on the integration of sensory information, motivational states and memory, from which risk assessment and optimal action selection emerge [5]. The foraging and feeding behavior of molluscs is known to be one of the most suitable systems for the analysis of motor patterns, interactions of central pattern generating interneurons, and the role of sensory inputs in the initiation and maintenance of the behavior [2]. More commonly, decision-making studies approach behavior selection within a mathematical framework to gather behavioral data and infer utility [1, 6] or within a neuroscientific approach to characterize neuronal interactions and behavior [7, 8]. The predatory sea-slug Pleurobranchaea californica has a simple, accessible nervous system and exhibits a repertory of simple behavioral elements that constitute effective foraging strategies. As a consequence, this particular animal model provides a unique opportunity to analyze the coupling between sensory inputs and
behavioral selection within the dynamics of the decision-making processes at the level of the neuronal circuit.

The use of mathematical models in basic behavioral research to predict and control behavior has increased over the years [6]. A sound mathematical model provides a framework for understanding what might otherwise appear to be unconnected behavioral phenomena. Still, mathematical models are often presented without complex mathematical details. One of the few approaches to quantitatively characterize how mutually interacting neural circuits give rise to decision-making was the work of Kilmer and McCulloch [9]. However, their model relied on arbitrary decision rules and did not explicitly elucidate the neurochemical interactions. Within the context of Pleurobranchaea and the neuronal interactions and behavior, the neuronal circuitry that mediates approach and avoidance turning [10], as well as the control of the turn motor network by the goal-directed feeding network which manifests appetitive state [7], are known in detail. However, the neural circuitry responsible for the integration of the sensory information processing leading to the triggering of specific behaviors has yet to be described [7, 8, 11].

The intersection between behavioral neuroscience and mathematics defines a unique space to study the physiology of the nervous system and invites detailed, creative applications in both areas. Simultaneous spike-train recording techniques have been available for many decades, but the quantitative approaches to analyze data have led to resulting interpretations that can be considerably distorted or misleading [10]. In the past decade, more sophisticated approaches have been established to understand how neurophysiological signals represent brain function. In particular, point process likelihood methods applied to simultaneously recorded neurons have been shown to provide a sound framework for data analysis within the context of sensory input [12], learning and memory [13], and motor output [14].

For my thesis project, I have three aims:

1. To identify, record, and characterize functional relationships between sensory and motor neurons whose activity collectively elicits behavioral selection;

2. To elaborate a computational agent-based model to understand and observe these interactions as a mechanism based on the utility function from their foraging intake and learning;
3. To employ statistical decision theory, information theory, and control theory to infer the utility being collectively maximized by the sensory inputs, neuronal interactions, and motor outputs.

This research methodology uniquely combines state-of-the-art principles from neurophysiology, Bayesian statistics and machine learning, as well as information and control theory, to understand how sensory, feeding and locomotor networks are coupled in the nervous system of a model animal. We use these insights from biology and statistical decision theory to develop novel computational and probabilistic models of neuronal decision systems.
CHAPTER 2

ANATOMY, BEHAVIOR AND SENSORY SYSTEM STUDIES OF THE
PLEUROBRANCHAEA CALIFORNICA

2.1 Introduction

The foraging and feeding behavior of molluscs is known to be one of the most suitable systems for the analysis of motor patterns, interactions of central pattern generating interneurons, and the role of sensory inputs in the initiation and maintenance of the behavior [2, 3]. As with most gastropod molluscs, Pleurobranchaea californica’s nervous system is organized into discrete ganglia containing large nerve cell bodies. Much fictive behavior can be elicited in isolated nervous systems, semi-intact preparations, and restrained whole animals, which show similar motor output patterns as occur in the intact animal [3]. For Pleurobranchaea, these include feeding, escape swimming, crawling locomotion, and orienting and avoidance turning [5]. Critical interneurons of the underlying networks and several aspects of the Pleurobranchaea’s nervous system have been identified and studied in some detail [7, 15, 16]. However to date, critical neurons and connectivity of the sensory network have not been fully studied.

Decision-making processes are believed to be the result of integration of sensory information, motivational states and memory [17]. This fact presents a rich question: How is the nervous system in this simple animal maximizing the utility of the foraging behavior? To answer this question, an investigation of the sensory inputs and the neurons whose activity collectively elicits behavioral selection was performed. We applied direct physiological analyses and other conventional neurophysiological investigations to explore the pathways through which sensory information is transmitted within the nervous system.

\footnote{Fictive behavior refers to the generation of nervous activity that would ordinarily produce behavior; except that the nervous system has been disconnected from the effectors, i.e. the generation of a motor pattern by a nervous system without the production of any movements or muscular contractions [4].}
system of the animal. Electrophysiological experiments were performed to identify and collect data from sensory interneurons in the sensory pathways of *Pleurobranchaea*, in order to identify how sensory information is being received and integrated in terms of the critical neuronal pathways that promote behavior.

### 2.2 Life History, Anatomy and Physiology of *Pleurobranchaea californica*

*Pleurobranchaea californica* is a predatory sea-slug - an opportunistic predator on numerous invertebrates, including conspecifics; whose abilities to track appetitive chemotactile stimuli and to recognize and avoid potentially dangerous stimuli have attracted the attention of neurophysiologists. *Pleurobranchaea* is a member of the order Notaspidea, or side-gilled sea slugs, which are marine opistobranch gastropod molluscs in the family Pleurobranchidae. In nature, it inhabits deeper waters, primarily below depths of 30 to 1200 feet on the Pacific Coast of the United States [16]. Figure 2.1 shows a picture of *Pleurobranchaeas* taken in captivity.
2.2.1 Anatomy

*Pleurobranchaea*’s body is shell-less, covered with a protective mantle skirt and contains a couple of rhinophores. Rhinophores are specialized anterior chemosensory organs commonly present in sea-slugs, located on the dorsal surface of the head, and are often chemoreceptors and rheoceptors. In the posterior part, the foot functions primarily for locomotion in the forward direction. The foot’s surface is composed of both mucus-secreting cells and cilia. In the anterior part of the animal, the prolonged mantle produces a large structure called the oral veil (OV), a cowcatcher-like structure, and extensions on its edges produce the tentacles. These structures are identified in Figure 2.2. The oral veil constitutes the major chemosensory organ in the animal, and is full of sensory papillae along the anterior edge. Chemotactile afferents from the papillae converge to peripheral ganglia in the tentacle and medial region of the oral veil complex.

2.2.2 Behavioral Repertoire and Physiology

The central nervous system (CNS) of *Pleurobranchaea* is organized circumesophageally and is composed of four distinct types of ganglia, namely cerebralpleural, pedal, buccal and visceral (Figure 2.3, visceral ganglia not
Figure 2.3: *Pleurobranchaeas* nervous system shown during gross dissection.

These ganglia are interconnected. The cerebralpleural ganglion (CG) is connected to the pedal ganglia via the lateral cerebropedal connectives, and to the buccal ganglion (BG) via a pair of cerebrobuccal connectives, and to the visceral ganglion. In terms of function, the BG regulates feeding behavior, the PG locomotion and muscle contraction. The CG is believed to perform the part of the sensory processing, and it responsible for central processing of behavior. From the oral veil, the large oral veil nerve (LOVN) and the tentacle nerve (TN) collect chemotactile input from the peripheral ganglia and transmit it to the central ganglion (CG). The receptive fields of the LOVN and TN are believed to be unilateral and overlapping. The LOVN response increases as stimuli are moved between tentacle and midline, and the opposite happens for TN.

2.2.3 Neuronal Substrates Involved in Behavior Selection

**Feeding, Approach, Avoidance and Escape Swimming**

The strategy for behavior selection in the slug’s nervous system is believed to be a reconfiguration of shared neuronal networks [3]. Central neuronal circuitry integrates the information arriving from the LOVN and TN into a precise representation of stimulus localization and intensity from which motor output is elicited and the corresponding behavioral response is selected [18]. Approach, avoidance, feeding and escape swimming are the most studied
behaviors in *Pleurobranchaea* [4, 19]. Avoidance behavior is characterized by local withdrawal of the affected body part, head withdrawal of the anterior head region and oral veil, and active avoidance (avoidance turning followed by crawling) [3]. Turning away from or towards a stimulus involves the asymmetrical contraction of longitudinal lateral body wall and foot muscles. During avoidance, contralateral longitudinal muscles are contracted, and during approach ipsilateral longitudinal muscles are contracted [3]. Approach is also characterized by an orienting turn and forward crawling motion. Feeding involves a slight orienting turn, proboscis extension and biting. Escape swimming is a period of alternating dorsal and ventral body flexion, which along with strong currents, allows for a faster distancing from prey or noxious stimuli [20].

Feeding and avoidance behaviors are exclusive of each other: induction of active feeding suppresses avoidance withdrawal from mechanical stimulus [21]. This was observed when avoidance behavior (turn and locomotion), after being stimulated by feeding stimuli, was replaced by active feeding (biting) at higher stimulus strengths [21]. Proboscis extension alone does not interrupt avoidance behavior (head withdrawal, turning and locomotion). Escape swimming (predator avoidance behavior) overrides all other behaviors, feeding included [4]. This override is caused by synaptic inhibition of the feeding and turning CPGs by the activation of the central pattern generator for the escape swim, in specific of the cell A1 [3, 4]. In nature, swimming is often triggered by touch or bite of conspecific [3, 4]. Thus, food stimulation (appetent or noxious) simultaneously excites both the feeding networks and the avoidance networks [3, 4]. Depending on the synaptic strengths of these connections, the desired behavior will be elicited. Swim network activation causes a more widely spread inhibition than the one by the feeding network. Therefore, in terms of spread of the inhibition, escape swim, feeding and avoidance/turning behavior follow a decreasing hierarchy of inhibitory effect.

Critical neurons involved in the mentioned behavior networks have been successfully identified [20]. It has been shown that *Pleurobranchaea* uses a concise circuitry to achieve fairly complex behavior, with overlaps amongst networks [5]. The neuronal structure controlling the swim escape behavior is driven by a central pattern generator (CPG) composed of seven neurons from the A cluster [11]. The A cluster of the swim CPG is a group of interneuron
cell bodies (As1-4, A1/A10, and Ivs, yet to be located) that mostly project to the pedal ganglia, as well as other interneurons that connect to the CPG, and is utilized in the feeding and motor networks [3, 20, 21]. The CPG interneurons are responsible for dorsal and ventral flexions that compose the swim pattern, firing cyclically in phase to set the swim rhythm. The structure of the swim CPG is shown in Figure 2.4A.

Activation of the known pattern-generating elements of the swim CPG and A-ci1 (follower interneuron) inhibit the feeding command neurons [3]. In the feeding network, the interneuron I1, the feeding command neurons, PCp and PSE, and the interneurons I2s, are directly involved in the behavioral output of that network. Paracerebral neurons (PCNs) receive excitatory stimuli when food is presented to hungry animals, and they show prolonged inhibition in satiated animals. The feeding network is mostly composed of the PCNs that drive part of the motor network, and the interneurons group 1 (I1) and interneuron group 2 (I2). Stimulatory input to the I2s will, in turn, drive the I1s to stimulate the PCNs, producing retraction or protraction phases of feeding, depending upon the nature of the stimulus [3]. The structure of the feeding CPG and its connection to the swim CPG are shown in Figure 2.4B.

For avoidance and orienting behaviors, the network is simpler, as seen in
Avoidance and orienting turns are commanded by the pair of premotor neurons (A4 cells) from the CG [22], responsible for encoding both turn direction and amplitude. The avoidance and orienting network overlaps the swim network through the serotonergic interneuron group (As1-4) [20, 22]. The As1-4 neurons sustain A4 excitation and determine directionality of A4’s activity. This reflects the importance of the As1-4 neurons, for driving excitation in both the avoidance turning and the escape swimming CPGs [20]. Lower levels of noxious stimulation drive avoidance behavior (A4 and As1-4 coupling), whereas higher levels of excitation and stronger activation of As14 determines the activation of the swim network.

2.2.4 Sensory Processing

An interesting problem in *Pleurobranchaea californica* rises from the connection between the apparent coupling between the sensory and motor networks; and the non-trivial processing believed to happen in the periphery and in the CG [23]. Although some effort has been made to identify the role of sensory inputs in the activation of behavioral networks and behavior selection, little is known about the sensory network itself. Diverse studies have clarified the computations for directional behavior in arthropods and fish [24, 25, 26], and a few have similarly examined elements of foraging behavior in molluscs [27]. Moreover, despite studies done to elucidate the behavioral elements of orienting and avoidance, to our knowledge, no studies have actually been done to elucidate the neural bases for sensory processing.

Sensory processing and its outputs play an important role in the strategy for behavior selection in the slug’s nervous system. Behavioral experiments reveal that orientation and amplitude of turn response in this animal are governed by a simple linear relationship dependent upon satiation state, stimulus location and chemical stimulus concentration [23]. Foraging gastropods like *Pleurobranchaea* are well-known for their abilities to orient to and follow odor trails of food sources or potential predators [5]. Gillette and Jing proposed a connectionist model of the effect of appetitive or nociceptive sensory stimuli in triggering the feeding and avoidance networks (Figure 2.5) [11]. This early circuitry hypothesis takes into account the effect of the sensory stimuli after it has been processed by the CG. The mechanisms involved in the processing
Figure 2.5: Proposed organization of sensory and motor pathways of feeding and avoidance networks. Gustation and Nociception represent sensory inputs.

of sensory data prior to its arrival at the CG have yet to be elucidated.

Yafremava and al. showed the effects of appetitive stimuli in orienting and non-appetitive stimuli in avoidance turns [23]. The results demonstrated that magnitude of turn angle and direction in response to chemotactile stimuli applied to the OV are computations based on stimulus’ site, strength and modality. Also, electrophysiological experiments on the mollusk’s sensory system determined that the quality of the chemosensory stimulus is encoded in the frequency of firing of the sensory nerves (LOVN and TN) [23]. Sensory inputs from the OV are initially integrated in an interneuronal layer where amplitude and laterality of chemotactile input is encoded. This information is then sent to the CG for further processing that will subconsequently activate the appropriate network and generate the behavioral response [16, 23].

Further studies of the chemotactile sensory map of the oral veil suggested the existence of a central integrating circuit in the CG [18] and, in part, a peripheral processing circuit in the sensory ganglia of the oral veil, tentacle and rhinophore [18, 28]. Electrophysiological studies confirmed this hypothesis and suggested the presence of inhibitory interactions between the sensory elements in the oral veil [18]. Lateral inhibition is an accepted mechanism of
stimulus feature extraction and sharpening in visual [29], auditory [13, 14] and olfactory systems [30, 31] of vertebrate and invertebrate organisms. In particular for the visual system, there is evidence of peripheral signal processing happening at the retina. This processing is essentially linear, and it compresses and shapes information before transmitting it to the optic nerve. Lateral inhibition also plays a crucial role, as it sharpens the contrast between neighboring retinal regions.

In foraging gastropods, like the *Pleurobranchaea*, the neuronal circuitry that mediates approach and avoidance turning has been worked out in some detail, as well as the control of the turn motor network by the goal-directed feeding network, which manifests appetitive state. This is shown in Figure 2.5. The putative sensory processing shows receptive fields in the periphery of the oral veil, and are said to be unilateral and overlapping [18]. Nervous pathways, mainly the LOVN and TN, connect the peripheral processing to the central processing in the CG. The neural circuitry responsible for the integration of the sensory information processing leading to the triggering of specific behaviors has yet to be unveiled [7, 8].

### 2.3 Determining Critical Neurons Involved in Sensory Processing

We devised physiology and electrophysiological experiments to get a clearer picture of the sensory processing in the *Pleurobranchaea*. Cobalt backfilling technique was used to identify critical neurons potentially associated with the sensory network. Initial electrophysiological studies were performed, but more extensive experiments are underway in order to determine the connections amongst sensory and behavioral networks.

#### 2.3.1 Intracellular Backfilling

To identify the sensory neurons of interest, backfill experiments with cobalt infusions were performed. Cobalt backfilling is extensively used as a technique to delineate the morphology of invertebrate nerve cells [32]. A ganglion of interest was placed inside of a well made with petroleum jelly and filled with *Pleurobranchaea* specific saline at 14°C (420mM NaCl, 10mM KCl, 139
25mM MgCl₂, 25mM MgSO₄, 10mM CaCl₂, and 10mM MOPS buffer, pH 7.5, [22]). The nerve selected to be backfilled was then placed in a well tangent to the ganglion’s one, with the walls reinforced to prevent leakage. Excessive moisture was removed from the nerve’s connective and the nerve well filled 0.5M cobalt dissolved in distilled water. The preparation dish was flooded with 20-25ml of saline and placed in a refrigerator (8°C) for 24h. The nerve was then cut at the CG’s proximal end, and the ganglion transferred to a new preparation dish for three consecutive saline washes. The preparation dish is was precipitated with a few drops of ammonium sulfide. The ganglion was fixed for another 12h in a solution of acetic acid-alcohol (1:3). After fixation, the tissue was dehydrated in 30 min of 100%, 75%, 50% and 25% alcohol, distilled water. After, the ganglion was cleared with methyl salicylate. The degree of backfilling is strongly correlated with electrical activity, which correlates with cobalt’s ability to enter the cut end of the nerve. These procedures allowed staining of axon processes 1-2 cm from filling site.

2.3.2 Extracellular and Intracellular Recordings

Dissections were made with animals anesthetized with 1M MgCl. Isolated cerebralpleural ganglion were pinned to Sylgard, under saline. Intracellular recordings were done with standard glass micropipettes, filled with 1M KCl [22] and pulled to obtain resistances of 10-15MΩ. Data was recorded via AD Instrument’s data acquisition module LabChart 8/30 and software LabView. Stimulation was achieved in the isolated CNS by delivering mono polar shocks of varying duration and intensity through the glass electrodes to the TN or LOVN. Monosynaptic connection was tested by the postsynaptic potential’s ability to follow presynaptic spikes one-for-one.

2.4 Results

2.4.1 Intracellular Backfilling

Sensory interneurons of the Pleurobranchaea are believed to be located in the cerebralpleural ganglion. The actual connections and terminations of these cells are not known, but structural evidence suggests their axons and/or
terminals originate in the oral veil’s periphery [18]. Cobalt applied extracellularly to the LOVN and TN near the CG filled identified sensory neurons potentially receiving direct information from the interneuronal peripheral layer of the oral veil and making possible connections to motor networks (Figure 2.6). Backfills also demonstrated a crossover communication between the two sides of the CG (Figure 2.7). Most cell bodies were located in the pleural lobe of the CG. Despite the promising results, some care should be taken since it has been shown that high concentrations of cobalt can cross neuronal membranes and suggest non-specific staining [32].

2.4.2 Extracellular and Intracellular Recordings

From the mapping of cells identified during intracellular backfills, we attempted performing intracellular recordings using standard glass electrodes. Once cells were impaled and healthy electrical activity established, stimulation of the LOVN or TN was performed to verify multiplicity of the synaptic connections. Alternatively, cells were hyperpolarized or depolarized, and in-
Figure 2.7: Two stained cells in the central region of the CG showed axonal connection.

Figure 2.8: Injury discharge of impaled cell and increased activity in the LOVN.
Figure 2.9: Slow excitation of the impaled cell after stimulation was delivered to the LOVN, suggesting a polysynaptic connection.

Figure 2.10: Fixed-frequency pulses delivered to the TN producing synaptic potentials in the impaled cell. Possible monosynaptic connection.
Figure 2.11: Fixed-frequency pulses delivered to the TN producing synaptic potentials and action potentials in the impaled cell.

Figure 2.12: Fixed-frequency pulses delivered to the LOVN producing long and slow synaptic potentials and action potentials in the impaled cell.
crease or decrease in the nerves’ activity was noted.

Here, we list a few results from intracellular experiment showing correlation between cells impaled and sensory inputs. Figure 2.8 shows an increase in the activity present in the LOVN when the cell is impaled.

Figures 2.10, 2.11, and 2.13 show post synaptic potentials (PSPs) and/or spikes being generated as a result of the stimulatory pulses sent through the TN. In some cases, the correlation is one-to-one. Figures 2.9 and 2.12 show similar results, where activation due to stimulation of the LOVN produces PSPs and/or spikes in the impaled cell.

2.5 Discussion

Even though good progress has been made in the identification of major neurons involved in the reproduction of specific behaviors, the neuronal processes involved in the sensory processing, coupling and switching of behaviors warrant further investigation. In specific, the synaptic mechanism involved in the sensory processing of information arriving from the oral veil requires a thorough physiological analysis. We performed intracellular backfills to iden-
tify critical neurons in the oral veil and in the CG involved with chemotactile processing. Furthermore, electrophysiological recordings from lateral body wall nerves, associated with avoidance and orienting turns, during stimulation of the LOVN and TN were performed, to shed light into the coupling between the sensory and turning networks.

For the cobalt backfills, the nerves of interest (LOVN or TN) were exposed to the cobalt solution and the CG was later precipitated and developed. Through those backfill experiments, we were able to identify 20 neurons of interest. They are shown in Figure 2.7. The results from backfill experiments showed variability in the neurons filled across 10-15 ganglia. This variability could arise from different axonal diffusion of chemical as electrical activity in the system subsides after isolation.

Intracellular recordings, although potentially indicative of monosynaptic and polysynaptic connections between the interneurons of the LOVN/TN and sensory neurons of the CG, are still insufficient to show definitive connections. Our findings from intracellular backfilling show a distinct group of neurons in the CG with direct axonal projections to the periphery of the OV. It is possible that other neurons involved in the sensory processing are preset in other ganglia, such as the buccal and pedal. Electrophysiological results confirm the hypothesis of possible one-to-one connections between sensory interneurons in the periphery and cerebralpleural neurons. Although some monosynaptic connections were found, a higher number of multi synaptic connections further supports the hypothesis of sensory processing happening at the level of the periphery in the Pleurobranchaea [18]. Furthermore, electrophysiological studies to determine influence of stimulus intensity and frequency in the expression of fictive avoidance or approaching behaviors are underway, and will be continued in collaboration with Dr. Rhanor Gillette. With this experiment, we hope to visualize the time course of sensory activity on the ventral aspect of Pleurobranchaea’s cerebralpleural ganglion, where we think the major central sensory interneurons are located. Future work can include a hemi-preparation consisting of the head (oral veil, tentacles, rhinophores) and CNS, so that mechanically or chemically stimulation of the sensory organs can be perfumed to observe the resulting activity with voltage-sensitive dyes.
CHAPTER 3

COMPUTATIONAL MODEL OF
BEHAVIOR, PREDATION AND
ADDICTION

3.1 Introduction

From the behavioral observations and electrophysiology, we know that sensory inputs and behavioral outputs are directly linked. The simple circuitry presented in Chapter 2, under Figure 2.5, reflects the coupling between sensory inputs, brain function and behavior. The questions become: Does this circuitry tell the whole story? If this circuitry is robust and describes the mechanisms underlying behavioral selection, can we generate those complex behaviors in simuli? What other behaviors can be extrapolated from that same circuitry? To solve this problem, we use our understanding of Pleurobranchaea’s behavior and the agent-based modeling.

Animals organize their behavior to efficiently extract resources from the environment, to avoid danger and to take calculated risks. The end result, with luck, is to enhance reproductive fitness. Optimal foraging, the ability to make wise decisions in resource discovery and exploitation, is widely observed among animal species [33]. It is clear that animals make cost-benefit decisions on the basis of their appetitive states; that is, their propensity to exercise any of a repertory of goal-directed behaviors. Appetitive state is best defined as the moment-to-moment integration of sensation, internal state (the various drives) and memory. The neuronal nature of appetitive state, its manifestation and control of decision have been approached at the level of detailed neuronal circuit in the predatory sea slug Pleurobranchaea californica.

Previously, behavioral observations suggested a simple and general neuronal model for the integration of sensation and internal (nutritional) state (Figure 2.5) [5]. Food stimuli induced feeding behavior (proboscis extension and biting) at sensory thresholds that varied directly with satiation state.
Very hungry animals not only attacked quite low concentrations of a food stimulus, but also would vigorously attack mildly noxious stimuli (acidic solutions and mechanical stimuli). Such aggressive behavior in the hungry animal might help it deal with active prey, by increasing the intensity of its predatory attack. In less hungry animals with higher feeding thresholds, sub-threshold food stimuli actually induced active avoidance behavior. This would be a reasonable behavior, which could lead the predator safely out of the way when another predator, like a larger cannibal *Pleurobranchaea*, arrives.

These simple observations were taken to indicate that a simple cost-benefit computation regulates behavioral switching in the animal’s foraging behavior, where food stimuli above or below the incentive threshold level for feeding would respectively induce feeding or avoidance [5]. This decision mechanism could weigh the animal’s need for nutrient against potential risk from other predators or prey defenses and the cost of energy outlay in an attack on prey. This decision was captured in the simple, hedonically structured neural network model of Figure 2.5.

### 3.2 Cost-Benefit Neural Network Model

The simple model in Figure 2.5 presented decision-making without a learning mechanism. However, it is known that *Pleurobranchaea* learns the association of prey odor signature with the consequence of an attempted attack [8, 34]. We propose an enhanced biological model, featuring sensory integration and the minimal connections necessary to account for combined electrophysiological and behavioral data, as presented in Figure 3.1.

In this reformulated model, as also happens in real sensory systems, integration is performed in functionally parallel pathways. Two sensory integrating networks labeled *Appetence* and *Pain* are connected to either the feeding or the avoidance network, in that order. Sensory inputs encode chemosensory signature, a patterned input specific to a particular chemical stimulus; a basic nutritive food signal (amino acids like trimethylglycine for *Pleurobranchaea*, sugars for bees, etc.); and a combination pain and mechanoreceptor path that can encodes noxious stimuli. All sensory inputs access both *Appetence* and *Pain* simultaneously, and the pain path also has direct access to
Figure 3.1: Behavioral selection circuitry as hypothesized for the *Pleurobranchaea californica*. Sensory signatures (context, resource signal and nociception) are sent to the sensory integrator circuits (periphery and/or CG) and distributed to the behavior specific networks for expression.

avoidance circuits, as is also the case in many invertebrates and mammals.

Modulatory feedback pathways coming from the feeding and avoidance networks are connected to the sensory integrator networks that provide for reinforcement learning, or Hebbian learning, to occur at sensory integrating loci. This ensures that a hedonic value is assigned to specific chemosensory signatures depending on whether they were associated with active feeding or avoidance behavior. In the model of Figure 3.1, the chemosensory connections mediating odor signature (blue circles) are those susceptible to Hebbian learning.

The general model of Figure 3.1 is both minimal and testable, and it can be used to explain how animals can make the judgments that underlie optimal foraging by encoding information on their internal state and integrating that with sensation and experience. Placed in a simulated natural environment, this model could be used to test a variety of simple hypotheses.

A key feature of the model is the excitation state of the feeding network, which is the neuronal network analog of behavioral arousal – both are affected by satiation and sensation. In this case, satiation regulates the set-point of the network in terms of spontaneous activity and responsiveness to
sensation. Thus, hungrier animals require a weaker stimulus to reach active feeding threshold. In this model, the feeding network acts as a “goal-directed neuronal network”. Its central role in decision rests on two defining characteristics: 1) it is a main site for the final integration of sensation, internal state and experience; and 2) when active, it promotes behavioral elements consistent with its goal (in this case, nutrition) and suppresses those that are not. The model also incorporates the feature of learned association of the signature of a stimulus (virtual odor here) with consequence, which increases the preference for the respective appetent or aversive qualities of the stimulus signature. Readiness to feed is directly related to appetitive state, the integrated product of sensation, hunger and memory. Appetitive state is thus a function of perception of prey, a time-dependent function of both virtual nutritive content and quantity of the previous meal, and a function of the remembered consequences of previous encounters with the prey odor.

We built this model to include prey preference/avoidance learning as it plausibly acts in Pleurobranchaea, and simulated it in NetLogo, a multi-agent programmable modeling environment. Successful function in the simulation mimicking major aspects of predator/prey relationships supports the credibility of the model architecture as a core around which value-based decision-making may be based in both real and artificial nervous systems.

3.2.1 Basics of Agent-Based Modeling

An agent-based model (ABM) simulates systems comprised of autonomous, interacting agents with an encapsulated set of behaviors. ABMs can include challenging features to represent as rates or levels, such as step-by-step processes and conditional decisions [35]. Moreover, ABMs provide a process-oriented alternative to descriptive mathematical models and can simulate emergent, self-organized, nonlinear and adaptive phenomena [36]. More importantly, ABMs can address problems that system-level models cannot for being better suited to domains where the main aspect is individual or the observable, rather than the equations [35]. ABMs can be used to represent complex dynamics of ecological and social systems, group pattern formation, cooperation, prediction, manipulation and improvement collective behavior.

ABMs have three main components: agents, environment and events. This
is graphically depicted in Figure 3.2. Agents are a simplified representation of real-world entities, solving problems through sensing and reacting to stimuli from themselves, other agents, and the environment within which they operate. They are characterized by a set of properties and behaviors which promotes support for modeling decision-making. The set of behaviors under which they operate is determined by parameters, state variables and learning rules. Agents can be reflexive, goal-based, utility-based agents, and internal-state agents with increasing order of sophistication with respect to their decision-making processes. Goal- and utility-based agents have learning mechanisms that allow for decision rules to be modified in response to system dynamics. These learning mechanisms can be represented using machine learning algorithms (neuronal networks, reinforcement learning, etc.)[36]. Output behavior is also determined by the learning rules.

Environment and events are the two other fundamental components of ABMs. Environments constrain the flow of energy, information and agents across space while events play a role in representing the dynamics of real-worlds in the model. Each interaction of an agent-based model can be viewed as equivalent to a sampling event in a real-world system. Stimuli received are matched to the condition side of decision rules, triggering relevant behavioral response on the action side. This stimulus-response exchange drives an
agents’ interaction with its environment and other agents as they attempt to solve problems.

3.2.2 Rescorla-Wagner

Rescorla-Wagner (R-W) is an established reinforcement learning model presented by Robert Rescorla and Allan Wagner in 1972 [37, 38], and has been used through the years in animal learning research. It assumes that not all stimuli present during learning will control behavior by taking into account the history of the stimuli and their saliences. Based on Pavlovian conditioning, if a particular conditioned stimulus (CS) has been associated with an unconditioned stimulus (UCS) and a second CS is presented, it hypothesizes that little will be learned about the second CS. That is to say that previous conditioning will inhibit further associations to happen [38].

The RW model describes the step-by-step changes in the acquisition of associations between stimuli and motivating events in classical conditioning paradigms. Over trials, each stimulus is repeatedly paired with an event resulting in the gain of associative strength between the pair, defined as:

\[
\Delta V(t) = \alpha \times \beta \left[ \lambda - V(t - 1) \right]
\]  

(3.1)

and

\[
V(t + 1) = V(t) + \Delta V(t)
\]

(3.2)

where \( V \) is the instantaneous associative strength of the stimulus, \( \lambda \) is the maximum associative strength possible between the pair, \( \alpha \) and \( \beta \) are constants representing the salience of conditioned and unconditioned stimuli, respectively [37]. The \((\lambda-V)\) error term determines the rate of learning. This error term becomes negative when the reinforcer is not present, reducing the associative strength value to represent extinction.

3.3 The NetLogo Model

We implemented a model of the autonomous decision-making process in the predatory sea-slug *Pleurobranchaea* utilizing the modeling environment NetLogo. The interactive, animated simulation, Cyberslug, features a vir-
virtual predator with learning abilities based on the Rescorla-Wagner model of conditioning. The resulting virtual entity learns by experience to discriminate odor signatures of prey on the basis of nutrient value and noxious defense ability, copying Pleurobranchaea’s ability to learn odor discriminations [34] and to avoid dangerous prey. In particular, the model incorporated approach and avoidance decisions based on the animal’s appetitive state (readiness-to-feed). Appetitive state, manifested in feeding thresholds to food stimuli, was calculated as a moment-to-moment integration of sensation, satiation state and memory. Satiated animals will actively avoid food, while very hungry animals will attack even mildly noxious stimuli. Odor strength and appetitive state regulated magnitude of turn angle during these behaviors. Simulation and operating instructions are freely available at www.life.uiuc.edu/slugcity/cyberslug.html.

In nature, and as replicated in our model, slugs initially orient towards primary appetents produced by prey, like betaine (trimethylglycine). They learn to associate specific prey odors with consequences of predation attempts. *Pleurobranchaea* readily learns prey preferences and aversions on the bases of odor, and thus learns to avoid the noxious Spanish Shawl nudibranch, *Flabellina iodine*, while continuing to attack the related *Hermissenda*. However, very hungry animals will attack and consume the noxious *Flabellina*. Another prey in the simulation is the Batesian mimic Faux-*Flabellina*, which mimics the odor of *Flabellina* but lacks noxious defense. If *Pleurobranchaea* decides to eat the *Flabellina* mimic, it may be more inclined to eat others of the same odor in the future. Odor avoidance learning raises feeding thresholds and magnitude of turn angle during these behaviors. The slug virtual entity integrates learning variables, sensation and hunger state into appetitive state, or appetence. Following, we explore these three concepts as they were formulated in the model.

### 3.3.1 Learning Mechanism

For a learning mechanism, we used the Rescorla-Wagner model of associative learning to represent expectations and experience in learning, extinction and retrieval. The basic assumption of the model is to have the *Pleurobranchaea* learn what to eat through association of specific prey odors with consequences
of predation attempts (pleasure/pain).

Animals orient towards/away based on scent and the learned association values of odors. Learning happens during each feeding event and is described by the changes in the associative strength between stimulus and the subsequent energetic gain or reward. The animal is presented with appetitive stimuli (*Hermissenda*, *h*, green orbs) and nociceptive — or painful — stimuli (*Flabellina*, *f*, red orbs) and the Faux-*Flabellina* (*ff*, green orbs with red centers), used for the extinction component. Consumption of the Bayesian mimic Faux-*Flabellina* (*ff*, green and red orbs) promotes extinction of learned *Flabellina* avoidance. Extinction is represented as an independently learned memory, and results in the loss of the learned response. With time and re-exposure to *Flabellina*, the associative aversive values are recovered.

The Rescorla-Wagner model attempts to predict how the associative value of the conditioned stimulus (odor of prey) changes over trials during conditioning and extinction. In the NetLogo environment, a trial consists in the presentation of prey odors during the agent’s random walk across the environment. This is followed by the decision to approach and consume it, or completely avoid it. The model obeys a set of equations, which will update the associative values (*V*$_f$, *V*$_h$ and *V*$_ff$ for *Flabellina*, *Hermissenda* and Faux-*Flabellina*, respectively) upon consumption of the prey, to finally represent the strength of the learned response and to compute magnitude of turn angle for approach and avoidance.

At each time step, the slug is presented with three options for actions (*a*$_t$), contingent upon its satiation state, sensory inputs and learned memory given by the associative learning values:

\[
a_t = \begin{cases} 
0 & \text{do nothing/wander} \\
1 & \text{approach or avoid prey } i, \text{ for } i = f, h, ff
\end{cases}
\] (3.3)

Depending upon the action selected, the associative learning variable are updated as follows:

\[
\Delta V_t(a_t) = \begin{cases} 
\alpha(a_t)\beta(a_t)\left[\lambda - \sum_{m=f,h,ff} V_{t-1}(m)\right] & \text{for } a_t = f, h, ff \\
0 & \text{else}
\end{cases}
\] (3.4)

and

\[
V_t(a_t) = V_{t-1}(a_{t-1}) + \Delta V_t(a_t)
\] (3.5)
where $\Delta V$ is the change in the association strength; $\alpha$ is salience and $\beta$ is the association value of the stimuli; $\lambda$ is the maximum learned associative value and $V$ is the current associative strength.

In the simulation, the associative learning values are used to determine intensity of response to a particular odor associated with a specific prey. Thus, a higher associative value for an odor will indicate a stronger preference for the associated prey, and an approach sequence will be generated. Conversely, a lower associative value indicates rejection, and an avoidance turn will be executed.

The virtual predator exhibits behavior appropriate to naturally observed foraging strategies, as it learns to preferentially attack and eat high-nutrient, low-defense prey. However, feeding incentive changes with satiation: when quite hungry it attacks low-nutrient, high defense prey, and when satiated it actively avoids all prey. That is, it adaptively integrates sensation, internal state and experience. Actions in the virtual ecosystem have realistic consequences resembling features of realistic predatory-prey relations: introduction of a nutritive, low-defense Batesian mimic of a highly noxious prey item demonstrates protection of the mimic as expected. Moreover, the mimic’s presence increases attempted predation on the noxious prey as the predator finds that not all are noxious.

While simple, the simulation reproduces major characteristics of optimal foraging and supports the notion that the model of Figure 3.1 resembles a general decision module around which animal behavior is organized. As in real ecosystems, prey densities markedly influence predator choice. For instance, at low overall prey densities the predator may completely extinguish both populations of noxious and benign prey within the confined area. The model also incorporates other observed behavioral characteristics of the *Pleurobranchaea*, such as conspecific predation and mating.

**Satiation and Sensation**

In our model, satiation ($s$) is modeled as a variable, which coupled with the animal’s energetic state, decreases exponentially through time as the animals energy decreases. It increases with each feeding opportunity, proportionally to the energetic reward of the captured prey. Threshold levels are used to determine hunger states in the agents. Conspecifics and prey produce specific
odors and betaine, which once diffused into the environment, are sensed by the slugs through the use of specific Flabellina, Hermissenda and conspecific sensor modules. Sensory values are termed as odor strength $os_i$ (where $i = f, h, b$) for each of the prey's odor and for the betaine.

Satiation follows a simple equation, as shown below. Energy ($E$) is a variable incremented after feeding and decaying with time lapse.

$$s = (1 + ke^{(-E)})^{-2}$$

(3.6)

where $k$ is a normalization variable.

**Determining appetitive state**

Appetence and pain are filtered into the slug's feeding-network neurons, where it will decide how to proceed towards a scent based on previous encounters. The hungrier a slug, the more likely it is to bite a toxic animal. If the appetence of a slug is greater than its pain values, it will orient towards a scent, and vice-versa. In addition, the stronger a scent, the greater the angle at which the animal orients.

Following these observations from biology, our model integrates sensation to determine movement orientation (towards or away) and its magnitude (turn angle). Proximity to a prey will increase the animal’s sensation of the chemical betaine; thus, appetence and sensation are directly proportional. Also, given that Hermissenda is considered to be an appetitive stimulus; the proximity of a Hermissenda prey will increase the agent’s appetence proportionally to the associative learning value for that prey type. Thus, we can represent the model’s formulation of appetence ($app$) to be as follows:

$$app = \gamma(os_b - os_f) + V_h(os_h)$$

(3.7)

where $V_h$ is the current associative learning strength for Hermissenda; $os_b$, $os_f$ and $os_h$ represent average odor strength for the betaine, Flabellina and Hermissenda, respectively; $\gamma$ is a normalization variable.

Odor strength is computed as the diffusion of the odor value assigned to each prey at beginning of simulation and the relative distance to the predator. Similarly to appetence, pain is proportional to the animal’s sensation of
betaine odor diffusion and to the increase of *Flabellina*'s odor strength proximate to the slug. The model's representation of pain \((p)\) can be summarized as:
\[
p = \gamma (os_b - os_h) + V_f(os_f)
\]
where \(V_f\) is the current associative learning strength for *Flabelina*; \(os_b, os_f\) and \(os_h\) represent odor strength for the betaine, *Flabelina* and *Hermissenda*, respectively; \(\gamma\) is a normalization variables.

**Movement**

Movement of the agent throughout the simulation environment is determined by a combination of its satiation level, learning variables and sensory inputs. These variables will determined angle of turn of the agent’s next step. At a given time, if the agent is satiated, it will perform a random walk until it reaches feeding threshold, at which point it will engage in an approach and orientation sequence.

During approach, the angle of turn \((\theta)\) is computed as follows:
\[
\theta = \epsilon (V_h - V_f)(app)(os_h + os_f + of_b)
\]
where \(\epsilon\) is a normalizing constant and the other variables are as previously defined. Note that, when \(V_h > V_f\), the agent will have a positive angle of turn, characterizing an approach sequence, and versa for an avoidance turn.

**Conspecific Behavior: mating and reproduction**

Agents will engage in a mating and reproductive sequence when they reach a mature state (dependent upon agent’s size) and encounter a smaller conspecific. After mating, the larger agent will lay virtual eggs and a random number of those eggs will hatch into larvae. The number of hatched eggs is randomly determined as to model chances of survival in nature. The larvae will subsequently develop and grow into adult agents.
Other user features

In the simulation screen (Figure 3.3), users have access to tools designed to create an stimulating and interactive environment. Via click-interaction, users can: (i) place animat eggs and increase population; (ii) increase size of a specific slug; (iii) eliminate a slug; (iv) add odor traces of the two types of prey (Hermissenda or Flabellina) or an odor trace of a conspecific; (v) reset prey placement; and (vi) remove all prey.

3.4 Modeling Addiction

The integration of the behavioral agent-based model described above created many opportunities to study what other behaviors could be extrapolated from that same simple circuitry as described in Figure 3.1. Addiction is often believed to be a hijacking of the brain’s sensory and reward circuitry [39]. Humans, as well as other organisms, engage in behaviors that have rewarding properties. Pleasurable feelings provide positive reinforcement so that the behavior is most likely to be repeated.

Addiction is thus a state in which an organism engages in a compulsive behavior, which is reinforced or rewarded, even when faced with negative consequences. There are many theories surrounding addiction learning [40, 41, 42]; the most popular within certain groups of neuroscience identifies addictive
drugs as capable of directly acting on reward pathways, usurping the nervous system’s normal reward mechanisms. In a general sense, addiction can be thought of as a neurobiological process undertaken within the reward processes and pleasure phenomena. Natural rewarding behaviors are necessary for survival and appetitive motivation, usually prevailing over other biological behaviors like eating and reproduction [41].

With most animals studied under addiction, it has been shown that release of a specific neurotransmitter is likely key to the addiction processes. In mammals, changes in the release of the neurotransmitter dopamine – involved in many brain processes, including motivation, punishment and reward – masks the actions of drugs of abuse as rewarding [41]. Normal reward pathways are activated by the consumption of drugs as well as natural rewards [40]. In invertebrates, this neurotransmitter is believed to be serotonin [17]. Moreover, all addictive drugs are said to activate the reward system by raising the levels of neurotransmitter release and disrupting normally perfectly quantifiable relationships between reward and behavior [40, 41].

Here we model addiction using a reinforcement-learning paradigm. We build our addiction model based on the previous construct of prey tracking and avoidance. Addiction is simulated as a high reward prey, consumption of which leads to a change in the activation states of the appetitive and pain networks. We model addictive behavior as a result of the usurpation of a normal foraging behavior by a super-enhanced stimuli. Activation of goal directed networks leads to activation of reward pathways – the same used during foraging, promoting intense activation of the reward pathways. This determines a pattern similar to search images, creating unusually strong memory traces, and a strong association with the paired stimuli, that will guide future appetitive behavior. We hypothesize that high activation of the reward pathways suppresses the pain networks. As activation wears off, pain-related pathways rebound causing the irritability of the drug absence.

Following those concepts, our model simulates the addiction process in the Pleurobranchaea by introducing a high reward substance (prey*). Due to the reward associated with the drug, animals quickly develop a preference and actively seek the drug particles in the environment. The change in associative strength between the drug and its reward is calculated using the same mechanism as used for predation associative learning (Rescorla-Wagner). Consumption of the drug over time strengthens this association
similar to the homeostatic-plasticity mechanisms observed in the nervous system. It also causes a temporary inhibition of the pain network, which with increase in tolerance and diminished drive to actively seek the drug, will rebound and manifest the withdrawal behavior aspects.

Repeated use leads to destructive behavior despite very negative consequences, and involves an overwhelming compulsion, based in the alteration of brain circuits that normally regulate judgement. Tolerance occurs when the brain reacts to repeated drug exposure by adapting its own chemistry to offset the effect of the drug — it adjusts itself to tolerate the drug (counteract inhibition). Withdrawal is a different process, and it happens when the drug leaves the brain quickly, causing an imbalance in the brain’s functioning. After the brain has adapted itself to continued drug exposure, if the user stops using the drug, the brain will be counteracting an effect that is no longer being caused. That is withdrawal. Withdrawal is often symptomized by discomfort, increased pain and decreased reward association.

3.4.1 Addiction Learning

The drug learning mechanism follows the same method, the Rescorla-Wagner model, used for reinforcement learning in the predation paradigm. The basic assumption of the model is to have the *Pleurobranchaea* learn preferences through the association of specific prey odors with its rewards. Addiction is modeled through the presentation of a high-reward prey represented by an addictive stimulus (*prey*,$^*$, *d*, yellow orbs).

Similarly to the pure-learning process, The model obeys a set of equations updating the associative values $V_f$, $V_h$, $V_d$ and $V_{ff}$ for *Flabellina*, *Hermissonenda*, *prey*,$^*$ and Faux-*Flabellina*, respectively, upon consumption to finally represent the strength of the learned response and to compute magnitude of turn angle for approach and avoidance. At each time step, the slug is presented with three options for actions ($a_t$), contingent upon its satiation state, sensory inputs and learned memory given by the associative learning values:

$$a_t = \begin{cases} 0 & \text{do nothing/wander} \\ 1 & \text{approach or avoid prey } i, \text{ for } i = f, h, d, ff \end{cases}$$

\[\text{(3.10)}\]
Depending upon the action selected, the associative learning variable are updated as follows:

\[
\Delta V_t(a_t) = \begin{cases} 
\alpha(a_t)\beta(a_t) \left[ \lambda - \sum_{m=f,h,ff} V_{t-1}(m) \right] & \text{for } a_t = f, h, d, ff \\
0 & \text{else}
\end{cases} \tag{3.11}
\]

and

\[
V_t(a_t) = V_{t-1}(a_{t-1}) + \Delta V_t(a_t) \tag{3.12}
\]

where \( \Delta V \) is the change in the association strength; \( \alpha \) is salience and \( \beta \) is the association value of the stimuli; \( \lambda \) is the maximum learned associative value and \( V \) is the current associative strength.

Upon consumption of the \( \text{prey}^* \), the pain network is suppressed and a reinforcement mechanism consolidates the reward stimulus context or odor of \( \text{prey}^* \). Tolerance to \( \text{prey}^* \) accumulates over time as both a decrease in reward and an increase in the intrinsic excitability of the pain pathway are manifested due to homeostatic plasticity. The inhibitory input to the pain network is lifted after a period with the fading of effects of consuming the high-reward \( \text{prey}^* \), sending it into an excitatory rebound.

Suppression of the pain network by the inhibitory and excitatory inputs \( (S) \) is calculated as a function of tolerance and activation of the addiction and appetitive pathways, and is given by:

\[
S(t) = \begin{cases} 
(\sigma d)\tau(t) & \text{for } t < k \\
S(k) + t & \text{for } t \geq k
\end{cases} \tag{3.13}
\]

where \( \sigma \) is a normalizing constant, \( d \) is user-defined drug intensity, \( k \) is a user-determined timing constant for the addiction mechanism and \( \tau \) is tolerance, given by:

\[
\tau(t) = \begin{cases} 
0 & \text{for } t = 0 \\
\tau(t-1) + l(1 - e^{-t/k}) & \text{for } t > 0
\end{cases} \tag{3.14}
\]

where \( l \) is normalization factor for the number of \( \text{prey}^* \) consumed. After a \( 3k \) time-delay following the last consumption of \( \text{prey}^* \), tolerance will decay to zero, representing desensitization.

Addiction \( (\text{add}) \), a measure of the agent’s responsiveness to the drug and sensory input changes due to the drug’s odor signature and the agent’s prox-
imity to it, is calculated as:

\[ \text{add} = \epsilon V_d(os_d) + \frac{\epsilon}{2}(V_h - V_f)(os_h - os_f) \]  

(3.15)

where \( V_d \) is the current associative learning strength for the drug, \( os_d \) represents the odor signature’s strength for the drug, and \( \epsilon \) is a normalization parameter. The other variables are as defined previously.

Pain is reformulated for addiction simulations, as follows:

\[ p = \gamma(os_b - os_f) + V_f(os_f) + S \]  

(3.16)

where \( V_f \) is the current associative learning strength for \( Flabelina; os_b, \) and \( os_f \) represent odor strength for the betaine and \( Flabelina, \) respectively; \( \gamma \) is a normalization variable.

3.5 Discussion

The goal-directed network at the core of the cost-benefit model for the optimal forager is the central point for summation of the inputs from sensation, internal state and experience. The key feature manifesting appetitive state is the excitation of the feeding network. Its outputs can specify behavioral choice through selective activation or inhibition of distinct motor networks, or through biasing networks towards particular metastable states for expression of specific motor patterns. Modulatory outputs of the feeding and avoidance motor networks also specify which stimuli are subject to Hebbian/anti-Hebbian learning in sensory integrator networks, from which emerges reinforcement learning. Thus, decision to attack or avoid a stimulus source rests on ability to associate prey odor signature with consequence through a simple learning mechanism that selectively affects signature amplitude in sensory integrator networks, and on the regulation of appetitive state by satiation mechanisms. The excitation state of the goal-directed feeding network is the final target whose output determines decision.

The computational model represented in Figure 3.1 abstracts the properties of the optimal forager in a simplified simulation to test the similarity of cost-benefit decisions to those expected of real foraging animals. The model closely reproduces foraging decisions of \( Pleurobranchaea \) as it incor-
porates approach/avoidance decision based on appetitive state, calculated as a moment-to-moment integration of sensation, satiation state and memory. Reinforcement learning of the appetitive and noxious qualities of a prey item was modeled with a relation optimally describing learning, including extinction, processes in animals.

The virtual predator makes decisions based on its varying hunger state, and on recognition of odor signatures and their associations with previous experience. The cyberslug learned to avoid noxious prey when not particularly hungry, and to favor nutritious prey with weak defenses. When satiated, it avoided all prey. However, at the extreme reaches of hunger it would attack and consume noxious prey, similar to behavior of many active predators, including *Pleurobranchaea*.

Successful introduction of a Batesian mimic into the environment, with predictable results, suggests that the model may be further expanded to test more complex ecological and economic hypotheses, for instance, with addition of aggression and socially-based decision. We made the *Pleurobranchaea* agent an enthusiastic cannibal like the real animal and placed it in an environment where it had to learn to differentiate safe and dangerous prey by experience, similar to that in which it is found.

The model closely reproduces foraging decisions of *Pleurobranchaea*, as it integrates sensation, internal state and associative learning, incorporating approach/avoidance decisions based on appetitive state. Virtual addiction is replicated in the *Pleurobranchaea* by introducing a high reward prey (*prey*). Consumption quickly leads to a reward-based preference. Consumption of the *prey* also causes a temporary inhibition of the pain network. With increase in tolerance and diminished *prey* reward through homeostatic plasticity mechanisms, the pain network excitation state will rebound to manifest aspects of withdrawal behavior.
CHAPTER 4

PRINCIPLES OF OPTIMIZATION
THEORY AND CONTROL THEORY TO
INFERENCE THE COST FUNCTION OF
FORAGING ATTEMPTS

4.1 Introduction

Most animals’ behaviors are defined by their critical need to acquire food. This drive, coupled with natural selection, functions as an optimizing agent and has played a central role in defining the neural mechanisms that control decision making [43]. Gillette and Jing postulate that foraging decisions are an evaluation of costs and benefits of feeding attempts, leading to an optimized foraging behavior [5].

Decision making is, then, defined as the integration of stimulus percepts, internal states, and memory of experience [18, 5]. Expected nutritional gains and energy expenditure of a feeding attempt, due to risks from noxious prey defense and other predators while feeding, are calculated against the nutritional needs — or hunger level — of the organism. The underlying mechanisms of feeding versus avoidance decisions in many animals is not well understood [5]. However, for our study animal, the Pleurobranchaea californica, those behaviors are well studied [8, 16, 18, 21, 22]. In specific, the influence of nutrient need on the expression of feeding and avoidance behaviors has been studied to some detail [5].

During behavioral experiments, it was shown that feeding stimuli can activate avoidance behavior at relatively low thresholds in hungry and satiated animals [5]. Gillette et al. proposed a mechanism under which Pleurobranchaea can predict the cost-benefit values of a feeding attempt based upon the qualitative characteristic of a stimulus and the animal’s own state of satiation [5]. For satiated or partially-full animals, stimulation of food chemosensory areas by weak appetent stimuli leads to avoidance expression [5]. Such behavior can be a result of the animal’s outweighing decision of a low-level nutrient prey versus the risk of predation during feeding attempt. For very
hungry animals, it was shown that mildly noxious stimuli can induce active feeding. Also, chive feeding inhibits avoidance for those animals [5].

Those observations draw a clear picture of the relationship between quality of chemosensory inputs and internal state in the activation of neural substrates in the Pleurobranchaea [5]. A natural question rises from these observations: Is the behavior of the Pleurobranchaea optimal, and if so, how can we quantify it?

To answer this question, we utilize results from optimization theory and Markov processes to elucidate how the utility functions of the animal — its cost and reward functions — change during foraging attempts. Although this model is simplified, it has been shown that such an approach can offer important insights in the underlying decision processes, reveal new patterns and generate new hypotheses [43]. We are ignoring details of individual neurons and networks, and focusing on the collective properties that produce the behavioral outputs. We model foraging and decision-making processes of the Pleurobranchaea as a controlled Markov Process based on behavioral observations, and apply a novel result in optimization theory to calculate the process’ cost function. Section 4.2, Background, will review probability theory, Markov processes and inverse optimal control (IOC), since it serves as basis for developing the model. Sections 4.3 and 4.4 will discuss the two different models proposed.

4.2 Background

4.2.1 Probability Theory

Sample Space

Given some experiment, the set $S$ of all possible outcomes is called the samplespace. An element of $S$ is a sample point. An event $A$, a subset of the sample space $S$, is a set of outcomes.

Probability Space

A probability space, or probability triple $(\Omega, \mathcal{F}, P)$, which consists of:
1. Sample space $\Omega$, the non-empty set of all possible outcomes $\omega$.

2. $\sigma$-algebra $\mathcal{F}$ is the collection $\mathcal{F}$ of subsets of $\Omega$, or a collection of events, with the following properties:
   
   i. $\mathcal{F}$ contains the empty set, $\emptyset \subseteq \mathcal{F}$.
   
   ii. $\mathcal{F}$ is closed under complement, that is if $A$ is an event, then its complement $\overline{A} = \{\omega \in \Omega \mid \omega \notin A\}$ is also an event.
   
   iii. $\mathcal{F}$ is closed under countable unions and countable intersections, that is: If $A_i \in \mathcal{F}$ for $i = 1, 2, 3, \ldots$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ and $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

3. A function $P : \mathcal{F} \to [0, 1]$, assigning to each event $A$ the probability of the event $A$, $P(A) \in \mathbb{R}$, and following the properties:
   
   i. For any $A \in \mathcal{F}$, there exists $P(A) \geq 0$.
   
   ii. $P(\Omega) = 1$.
   
   iii. $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, for every $\{A_k, k \geq 1\}$ disjoint.

Random Variables and Random Processes

Let $(\Omega, \mathcal{F}, P)$ be a probability space as defined above. A random variable $X$ is a function $X : \Omega \to \mathbb{R}$, such that

$$\{\omega : X(\omega) \leq r\} \in \mathcal{F} \quad \forall r \in \mathbb{R} \tag{4.1}$$

The cumulative distribution function (CDF) $F : \mathbb{R} \in \mathbb{R}$ of a random variable $X$ is defined as

$$F(s) = P(\{\omega \in \Omega \mid x(\omega) \leq s\}) \tag{4.2}$$

The distribution function $F : \mathbb{R}^n \to \mathbb{R}$ of a random vector $X = \{X_1, X_2, \ldots, X_n\}$ is defined to be

$$F(s_1, s_2, \ldots, s_n) = P(\{\omega \in \Omega \mid X_1(\omega) \leq s_1, X_2(\omega) \leq s_2, \ldots, X_n(\omega) \leq s_n\}) \tag{4.3}$$
Given the probability triple \((\Omega, \mathcal{F}, P)\), a random process \(X(t)\) is a time-indexed collection of random variables on \(\Omega\).

**Markov property**

A random process \(X = (X_t : t \in T)\) is said to have the **Markov property** if the following holds

\[
P(X_n = x_n \mid X_{n-1} = x_{n-1} \ldots X_0 = x_0) = P(X_n = x_n \mid X_{n-1} = x_{n-1})
\]

4.2.2 Markov Chains

A collection of random variables \(\{X_i\}\) for \(i = 1, 2, \ldots, n\) forms a Markov chain if, given that each time the system is in state \(i\), there is some fixed probability \(P_{ij}\) such that the system will next be in state \(j\). That is

\[
P_{ij} = P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} \tag{4.4}
\]

The state transition matrix \(P \in [0, 1]^{n \times n}\), with \(\sum_j P_{ij} = 1\) and \(P\) a stochastic matrix, whose \(ij\)th entry \(P_{ij}\), contains the probability of transitioning from state \(i\) to state \(j\).

The initial state probability distribution \(\pi_0 \in [0, 1]^n\), with \(\sum_i \pi_{0i} = 1\), is a probability distribution function over the set of states, whose \(i\)th entry \(\pi_{0i}\), gives the probability of that the \(i\)th state is the initial state. Now, let \(\Pi\) be the set of all probability distribution functions \(\pi_0\). Then for any \(k \geq 0\),

\[
\pi_k := \pi_0 P^k \in \Pi \text{ is the state probability distribution after } k \text{ steps of state transitions.}
\]

The stationary distribution of \(P\) exists if \(\lim_{k \to \infty} \pi_0 P^k\) exists. The invariant distribution of \(P\) if given by \(\pi^* \in \Pi\) if \(\pi^* P = \pi^*\).

A controlled Markov chain is a Markov chain whose state transition probability matrix is a function of its control inputs. That is, a random process \(X\) is a controlled Markov chain with respect to \(U\) if:

\[
P_{X_i|X_{i-1},U_i}(x_i|x_{i-1}, u_i) = P_{X_i|X_{i-1},U_i}(x_i|x_{i-1}, u_i) \tag{4.5}
\]

where \(a^n = (a_n, \ldots, a_0)\).
4.2.3 Markov Decision Processes

Consider the scenario proposed in Figure 4.1: an environment, with set states, and an agent moving between those states. The agent, therefore, has a certain set of actions from which to choose from, and based on the current state, it elides which state to go next. Given the agent’s action, there is a new state information and some reward from the environment. We wish to minimize the cost incurring from the agent’s action and state change.

Assume that such an environment obeys the Markov property and that there is a finite number of states. A Markov decision process (MDP) is used to model the dynamics of such an environment. MDPs are discrete-time stochastic processes defined by the 4-tuple $M = \{X, U, P, R\}$, where:

- $X$ is the finite set of states $\{x_1, \ldots, x_n\}$;
- $U$ is the finite set of actions for each state;
- $P$ is the state transition function specified as probability distribution;
- $R : X \times U \to \mathbb{R}$ is the reward function. $R(x, u)$ is the reward associated with taking action $u$ in state $s$.

We can then define a policy $\pi : X \to U$ to be the mapping between states and actions.
4.2.4 Inverse Optimal Control

Cost Function

Cost represents the effect of the outcome on the agent’s fitness, but since the environment is itself unpredictable, the agent must formulate a model of its behavior that results in a minimum average cost [44]. Assume that the agent an action space $U$, and each $u \in U$ is called an action. The environment has a state space $X$, and each $x \in X$ is called a state.

A function $C : X \times U \rightarrow \mathbb{R}$ is called cost function. The cost function $C$ depends on $u$ and $x$, and for $U$ and $X$ finite, it is convenient to specify $C$ as the $|X| \times |U|$ cost matrix. The average expected cost is given by:

$$J(\pi, x_0) = \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \sum_{k=0}^{N-1} c(x_k, u_k(x_k)) \quad (4.6)$$

Inverse Optimal Control Problem

Let the state space $X$ and action space $U$ be finite, such that $X = \{1, \ldots, n\}$ and $U = \{1, \ldots, m\}$. Given a policy $\pi$ and the state transition probability $p_{ij}$ (that is, the probability of going from state $i$ to state $j$), and let $P \in \mathbb{R}^{mn \times n}$ be a matrix such that

$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_m \end{bmatrix} \quad \text{where} \quad P_u = \begin{bmatrix} p_{11}(u) & \cdots & p_{1n}(u) \\ \vdots & \ddots & \vdots \\ p_{n1}(u) & \cdots & p_{nn}(u) \end{bmatrix} ; \quad \forall u \in U \quad (4.7)$$

We want to find the cost function $C$, such that this policy $\pi$ is optimized. Let $Q$ be the feasible region of $D$, a linear program in the variables $\{q(x, u)\}_{x,u}$:

$$D \quad \begin{array}{ll}
\text{minimize} & \sum_{x \in X} \sum_{u \in U} q(x, u)c(x, u) \\
\text{subject to} & \sum_{u \in U} q(y, u) = \sum_{x \in X} \sum_{U \in U} q(x, u)p_{xy}(u), y \in X \\
& \sum_{x \in X} \sum_{u \in U} q(x, u) = 1, \\
& q(x, u) \geq 0, \quad \forall x, u.
\end{array} \quad (4.8)$$
where \( q : x \times u \to [0, 1] \) be the \textit{occupational probability}, such that:

\[
q(x,u) = P(X = x_i, U = u_i) = P(X = x_i | U = u_i)P(X = x_i) \tag{4.9}
\]

for \( u \in U \) and \( x \in X \).

Given a \( q = q^* \) optimal for \( D \), a polyhedral set \( C = \{ z | z^T b_i \leq \alpha_i, i = 1, \ldots, m \} \) the class of objectives \( c \) if optimal if and only if \(-c\) lies in the normal cone \( \mathcal{N}(q^*; C) \):

\[
\mathcal{N}(z, C) = \left\{ v \mid v = \sum_{i=1}^{m} y_i b_i, \text{ with } y_i \geq 0 \text{ for } i \in I(z), y_i = 0 \text{ for } i \notin I(z) \right\} \tag{4.10}
\]

where \( I(z) = \{ i | z^T b_i = \alpha_i \} \), and let \( m \) be an integer.

The cost function \( C \) solves the inverse optimal control problem for \( q \in Q \) if and only if there exists \( \alpha \in \mathbb{R}^m, \beta \in \mathbb{R} \) and \( 0 \leq \gamma \in \mathbb{R}^{mn} \), such that \( \gamma \perp q \) and

\[ q = [q(1, 1), \ldots, q(n, 1), \ldots, q(1, m), \ldots, q(n, n)]^T \] [44]. Let \( I_n \in \mathbb{R}^{n \times n} \) be the identity matrix of size \( n \) and \( A \in \mathbb{R}^{mn \times n} \) be the matrix with \( m \) repetitions of \( I_n \); and \( e \in \mathbb{R}^{nm} \) be the column vector with all coordinates one. Similarly, \( c = [c(1, 1), \ldots, c(n, 1), \ldots, c(1, m), \ldots, c(n, m)]^T \).

\( D \) can then be rewritten as

\[
\begin{align*}
\text{D minimize} & \quad c^T q \\
\text{subject to} & \quad q^T (A - P) = 0 \\
& \quad q^T e = 1, \\
& \quad q \geq 0
\end{align*} \tag{4.11}
\]

Finally, the normal cone \( \mathcal{N}(q, Q) \) is given by

\[
\mathcal{N}(q, Q) = \{(A - P)\alpha + \beta e - \gamma |, \alpha \in \mathbb{R}^n, \beta \in \mathbb{R}, \gamma \in \mathbb{R}^{mn}, \text{ with } \gamma_i \geq 0 \text{ if } q_i = 0, \gamma = 0 \text{ if } q_i > 0 \}\tag{4.12}
\]

With this result, Kulkarni proposes the following theorem, which stems from the work of Borkar [45]:

\textbf{Theorem 1} \( c \) solves the inverse optimal control problem for a \( q \in Q \) if and only if there exists \( \alpha \in \mathbb{R}^n, \beta \in \mathbb{R}, \) and \( \gamma_i \geq 0 \) if \( q_i = 0, \gamma = 0 \) if \( q_i > 0 \) such
that \( c = (A - P)\alpha + \beta e + \gamma \). This result is shown by Kulkarni [44].

4.3 Problem Statement

Given an optimal forager, like the *Pleurobranchaea californica*, develop a model of the foraging and decision-making processes of the animal as a controlled Markov process based on behavioral observations; calculate the process cost function to infer how utility functions of the animal change during foraging.

To address the problem above, two models are proposed, with different realizations. Although simplified, these models should infer the utility being collectively maximized by integration of the the sensory inputs, integrative and intermediating processes and displayed through the behavioral outputs.

4.3.1 Model I

Our first model focuses on the tracking aspect of the foraging behavior of the *Pleurbranchaea*. In general, the environment consists of \( n \) discrete positions for visualization; one can think of it as a circle with \( n \) possible positions, graphically represented in Figure 4.2. Prey and predator are randomly placed in the environment, and predator will search for the prey by observing the contents of the locus directly to the right and left of its own position. It will then act according to a policy stipulated in the beginning of the simulation. Upon arriving at the same locus as the prey, the predator should proceed to consume the prey or move to an adjacent locus.

In the model, there are three decision makers (DM):

1. Predator: assigned to the primary DM, and it is decision-making properties relate to those of *Pleurobranchaea*.

2. Prey: assigned to the secondary DM, that together with the

3. Environment is used as an interfacing agent to the primary DM. Environment can be considered a synthetic representation of the *Pleurobranchaea*’s experimental setup, and is built for the purposes of modeling uncertainty in the decision-making process.
Suppose that the predator makes a decision based on a set of actions it can choose from: eat.

Consider then the $n$-dimensional environment ($n \in \mathbb{R}$), in which the predator is placed at a position ($pp_i$) with a certain level of satiety ($sa_i$), that will determine its drive to attack or not a nearby prey. For simplicity, we allow for only two levels of satiation: hungry or satiated. The prey is randomly placed at a position ($pr_i$) in the environment.

Throughout the simulation, depending upon its own placement, satiation level and the prey’s positioning, the predator will make foraging decisions within the set of possible actions: to eat, approach or ignore the prey. For example, if the animal is hungry and the prey is located to its left, the animal could choose to move left. However, if the animal is satiated instead, it could choose to move right and avoid the prey. Those assumptions, although simple, conform to the behavioral observations taken by Gillette and Jing, who noted that satiated animals often expressed avoidance behavior in the presence of food stimuli [5]. The prey is allowed to randomly move in single increments through the environment space at each iteration. Figure 4.2 shows the proposed environment with predator marked as $\circ$ and prey marked as $\times$.

Based on the scenario described above, the state space is given by the cartesian product of three distinct spaces such that $X = S_A \times S_P \times S_R \rightarrow X = \{x_i : x_i = (sa_i, pp_i, pr_i); sa_i \in S_A, pp_i \in S_P, pr_i \in S_R\}$, and where:

- $S_A = \{sa_i : hungry, satiated\} = \{0, 1\}$ - satiation state
- $S_P = \{pp_i : 1, 2, 3, \ldots, n\}$ - predator position
- $S_R = \{pr_i : 1, 2, 3, \ldots, n\}$ - prey position
Similarly, the action space is given by $U = \{u_i : \text{eat}, \text{move left}, \text{move right} = 1, 2, 3\}$. We established a simple policy $\pi(sa_i, pp_i, pr_i) = u_i, u_i \in U$, derived from behavioral observations, to determine action selection in the model:

- $\pi(sa_i = 1, pp_i < pr_i, pr_i) = 2$
  For satiated animals ($sa_i = 1$), when the predator is to the left of the prey, or $pp_i < pr_i$, it will move left ($u_i = 2, pp_{i+1} = pp_i - 1$).

- $\pi(sa_i = 1, pp_i > pr_i, pr_i) = 3$
  For satiated animals, when the predator is to the right of the prey ($pp_i > pr_i$), predator will move right ($u_i = 3, pp_{i+1} = pp_i + 1$). For the two previous cases, the predator will therefore avoid the prey, theoretically reducing energy costs associated of foraging.

- $\pi(sa_i = 0, pp_i < pr_i, pr_i) = 3$
  For hungry animals ($sa_i = 0$), when the predator is to the left of the prey, or $pp_i < pr_i$, predator will move right ($u_i = 3, pp_{i+1} = pp_i + 1$).

- $\pi(sa_i = 0, pp_i > pr_i, pr_i) = 2$
  For hungry animals ($sa_i = 0$), when to the right of the prey ($pp_i > pr_i$), predator will move left ($u_i = 2, pp_{i+1} = pp_i - 1$). For the two previous cases, the predator will approach the prey in both situations, actively engaging in foraging.

- $\pi(sa_i = \{0, 1\}, pp_i = pr_i, pr_i) = 1$
  For both satiated and hungry animals, if the predator and prey are in the same loci ($pp_i = pr_i$), predator will ‘eat’ ($u_i = 1$) and the prey is reassigned to a random loci with equal probability.

**Transition Matrix**

The transition probabilities are defined, for each action selection, with $p_{ij,k} = p(x_i, x_j, u_i = k) = P(x_j|x_i, u_i = k)$ being the conditional probability of the in state $x_j \in X$ at time $j = i + 1$, given that the previous state was $x_i \in X$, and action $u_i = k \in U$ was taken. The transition probability matrix, $P$, is then

$$
P = \begin{bmatrix}
P_1 & P_2 & P_3
\end{bmatrix}
$$

(4.13)
where $P_K$ is the transition probability matrix given that action $u_k$ was taken, that is $P_k = \{ p_{ij,k} = P(x_j|x_i, u_k) : x_i, x_j \in X \text{ and } u_k \in U \}$.

For $P_1$, the state transition matrix when action $u_1$ is taken, we assume that after eating the predator will always be satiated, that is, for $sa_i = \{0, 1\}$, $sa_j = 1$ always, with $j = i + 1$. Thus, for $p_{ij,1} \in P_1$, we have that:

\[
p_{ij,1} = \begin{cases} \frac{1}{n} & \text{for } \bullet = \{ sa_i = (0, 1), pp_i = q, pr_i = q \}, u_i = 1 \\ 0 & \text{otherwise} \end{cases}
\]

and

\[
p_{ij,1} = P[\{ sa_j = 0, pp_j = \mu, pr_j = \nu \} | \bullet = \{ sa_i = (0, 1), pp_i = q, pr_i = q \}, u_i = 1] = 0
\]

for $\mu, \nu, q < n \in \mathbb{R}$ where $n$ is the dimension of the environment.

For $P_2$, the state transition matrix when action $u_2$ – move left – is taken to be:

\[
p_{ij,2} = \begin{cases} \frac{1}{n-1} & \text{for } \bullet = \{ sa_i = 0, pp_i = k, pr_i = \mu \}, u_i = 2 \\ 0 & \text{otherwise} \end{cases}
\]

and

\[
p_{ij,2} = P[\{ sa_j = 1, pp_j = k - 1, pr_j = \nu \} | \bullet = \{ sa_i = (0, 1), pp_i = k, pr_i = \mu \}, u_i = 2]
\]

\[
= \begin{cases} \frac{1}{2(n-1)} & \text{for } \bullet = \{ sa_i = (0, 1), pp_i = k, pr_i = \mu \}, u_i = 2 \\ 0 & \text{otherwise} \end{cases}
\]

for $\mu, \nu, k < n \in \mathbb{R}$ and $\nu = \mu \pm 1$, where $n$ is the dimension of the environment.

For $P_3$, the state transition matrix when action $u_3$ – move right – is taken to
be:

\[
p_{ij,2} = P \left\{ \{ sa_j = 0, pp_j = k + 1, pr_j = \nu \} \mid \bullet \right\}
\]

\[
= \begin{cases} 
\frac{1}{n-1} & \text{for } \bullet = \{ sa_i = 0, pp_i = k, pr_i = \mu \}, u_i = 3 \\
0 & \text{otherwise}
\end{cases}
\] (4.17)

and

\[
p_{ij,2} = P \left\{ \{ sa_j = 1, pp_j = k + 1, pr_j = \nu \} \mid \bullet \right\}
\]

\[
= \begin{cases} 
\frac{1}{2(n-1)} & \text{for } \bullet = \{ sa_i = (0,1), pp_i = k, pr_i = \mu \}, u_i = 3 \\
0 & \text{otherwise}
\end{cases}
\] (4.18)

for \( \mu, \nu, k < n \in \mathbb{R} \) and \( \nu = \mu \pm 1 \), where \( n \) is the dimension of the environment.

Therefore, given the fixed policy \( \pi \) and transition matrix \( P \), for a given state \( x_i \), the model independently choses from the entries in the appropriate row vector of \( P \), determining the next state \( x_j \).

The occupational probability, as described in (4.9), and the cost function \( c \) are calculated. The cost function \( c \), as described in Theorem 1, is:

\[
c = (A - P)\alpha + \beta e + \gamma
\] (4.19)

for \( \alpha \in \mathbb{R}^n, \beta \in \mathbb{R}, \) and \( \gamma_i \geq 0 \) if \( q_i = 0, \gamma = 0 \) if \( q_i > 0 \) [44].

4.3.2 Model II

In our second model we follow Raos probabilistic features of an attack cycle [46] to study the success of a predators search for a clustered – or in our case, manually fed prey. Predation is often described as composed by the following states: search, pursuit, handle and eat [46]. Although these states are often assumed to be sequential, predation can also be considered to be a non-sequential process. Here, we stipulated some freedom within the model based on behavioral observations [5]. Rao and Kshirsagar noted that physiological characteristics of predators allow for non-sequential predation, a logical change from a sequential paradigm [46].

To fit the behavior repertoire of our study predator, we have reformulated the state model, and renamed the state pursuit as approach – consisting of the typical orienting turns executed by the *Pleurobranchaea*. In terms of
behavioral and physiological observations, search consists of a random walk with no specific target, eat consists of proboscis extension and ingestion, and avoidance, a stereotypical behavior sufficient to redirect the animals away from the stimulus source [47, 18].

Suppose, at the beginning of the experiment – or simulation run — the predator starts searching for prey. Its hunger level, or satiation, at the time is high. Hunger level continues to rise until prey is caught, and its lowered upon feeding. When predator is in the search state, it will approach the prey depending upon the success of the search. The probability that it eats the prey depends on the approach success. From this, it is easy to see the pattern forming in the predation cycle. However, because of disruptions and changes in the environment, the cycle can be considered non-sequential [46]. This is shown schematically in Figure 4.3.

Following the description above, we can label the states as: $s_1$ - search, $s_2$ - approach, $s_3$ - eat, $s_4$ - digest state. The state space $S$ is therefore $S = \{s_i, i = \{1, 2, 3, 4\}\}$. The action selection is then set as go-to-state actions (i.e. go-to-approach, go-to-eat, etc), and the action space is given by $U = \{u_1, u_2, u_3, u_4\}$, such that $u_1 = \text{“go-to-search”}$, $u_2 = \text{“go-to-approach”}$, $u_3 = \text{“go-to-eat”}$, $u_4 = \text{“go-to-digest”}$.

Similar to Model I, the transition probability matrix $P$ is defined by the probability column vectors of each action selection, $P_k$, where $P_k = \{p_{ij,k}(s_i, s_j, u_i = k) : p_{ij,k} = P(s_j|s_i, u_i = k)\}$, where $p_{ij,k}$ gives the probability of going from state $i$ to state $j$ after taking action $k$.

From biological observations [5, 7, 18], some of the probabilities in the
transition matrix will be null. That happens, for example, when the animal is in the search state: it cannot go to the digestion state next since that would not follow the natural order of a predation attack. These assumptions help clear out the transition probability and we arrive at Figure 4.4. The transition probability matrix is then simplified to:

\[
P = \begin{bmatrix}
p_{11} & p_{12} & 0 & 0 \\
p_{21} & p_{22} & p_{23} & 0 \\
p_{31} & 0 & p_{33} & p_{34} \\
p_{41} & 0 & 0 & p_{44}
\end{bmatrix} ; \quad \forall u \in U
\]

(4.20)

The policy \( \pi(s_i) = u_i, u_i \in U \) is simple: \( \pi(s_1) = u_2, \pi(s_2) = u_3, \pi(s_3) = u_4 \) and \( \pi(s_4) = u_1 \). Similar to Model I, given a fixed policy \( \pi \), a transition matrix \( P \) and a state \( s_i \), the model independently choses from the entries of the row vectors of \( P \), determining the next state \( s_j \). The occupational probability \( q \) (4.9) and the cost function \( c \) (4.19) can then be calculated.

4.4 Simulations and Results

The solution proposed for the inverse optimal control problem in this chapter was applied to the two models. Results were obtained that show accuracy in the estimated behavior and generated cost function.
4.4.1 Model I

The model was simulated with a simple method, using an environment with dimension $n = 5$, without loss of generality (Figure 4.5). At each time step, depending on its current state, policy and the transition probabilities, the agent will move towards or away from the prey.

For a simulation run with initial state $x = \{0, 5, 4\}$, $\alpha \in \{0, 1\}^n$, $\beta = 0.5$, and $\gamma_i = 0.5$ if $q_i = 0$, $\gamma = 0$ if $q_i > 0$, we obtain the cost matrix shown in Figure 4.6, and an excerpt of this matrix is shown in Figure 4.7. We observed that, although the model correctly simulates and the computation of cost follows the theorem proposed, due to the polyhedral nature of the cost function, we were not able to determine a set of parameters $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$, and $\gamma$ as described above that produced the optimal cost function. Future work should focus on this optimization problem.

For a better visualization of the cost function described in Equation (4.19), we will focus on visualizing the cost set for a specific state ($x = \{1, 1, 1\}$), without loss of generality, as the parameters $\alpha$ and $\beta$ vary. The simula-
Figure 4.6: Cost matrix. Simulation results from Model I, with initial state $x = \{0, 5, 4\}$, $\alpha \in \{0, 1\}^n$, $\beta = 0.5$, and $\gamma_i = 0.5$ if $q_i = 0$, $\gamma = 0$ if $q_i > 0$.

<table>
<thead>
<tr>
<th>states ($x_i$)</th>
<th>$\pi$ ($x_i$)</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_0$</td>
<td>$p_{p_1}$</td>
<td>$p_{r_1}$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 4.7: Excerpts of the cost matrix. Due to the multidimensional component of the cost function defined in (4.19), the dimension of the cost matrix shown is not the optimal one. Simulation results from Model I, with initial state $x = \{0, 5, 4\}$, $\alpha \in \{0, 1\}^n$, $\beta = 0.5$, and $\gamma_i = 0.5$ if $q_i = 0$, $\gamma = 0$ if $q_i > 0$. 
Figure 4.8: Simulation results from Model I, plot of $c(1,1)$ with $\beta = \{-1.0, -0.9, -0.8, \ldots, 0.8, 0.9, 1.0\}$, $\alpha = \{-1.0, -0.9, -0.8, \ldots, 0.8, 0.9, 1.0\}$ and $\gamma_i = 0.7$ if $q_i = 0$, $\gamma = 0$ if $q_i > 0$, for all the trials.

The simulation was executed for 81 trials, where the parameters $\alpha$ and $\beta$ were discretized, such that $\beta = \{-0.5, -0.4, \ldots, 0.4, 0.5\}$ and $\alpha(1)$, the $\alpha$ value associated with state $x = \{1,1,1\}$, was varied in similar mode to $\beta$, $\alpha = \{-0.5, -0.4, \ldots, 0.4, 0.5\}$. Throughout the simulation, $\gamma$ was kept such that $\gamma_i = 0.7$ if $q_i = 0$, $\gamma = 0$ if $q_i > 0$, for all the trials. Each trial was composed of 20 time iterations, that is, 20 changes in state per trial. We observed the change in the plane formed by the triad $(c,\alpha,\beta)$. This is a simplification from the polyhedral cone, since plotting it in is entirety is unfeasible. The resulting plot is shown in Figure 4.8.

4.4.2 Model II

The simulation for this model is simpler than the previous one, since there is only one possible initial state and the policy is also direct.

At each time step, depending on its current state, policy and the transition probabilities, the agent will select an action and move to the corresponding state, following the model depicted in Figure 4.4. After each simulation trial was over, the cost function was calculated, as determined in Equation (4.19). The cost function parameters, $\alpha$, $\beta$, and $\gamma$ were randomly selected in each trial, with $\alpha \in \{0,1\}^4$, $\beta \in \{0,1\}$ and $\gamma_i \in \{0,1\}$ if $q_i = 0$, and $\gamma = 0$ if
Figure 4.9: Simulation results from Model II, plot of $c$ for $\pi = \{2, 3, 4, 1\}$, with $\alpha \in \{0, 1\}^4$, $\beta \in \{0, 1\}$ and $\gamma_i \in \{0, 1\}$ if $q_i = 0$, and $\gamma = 0$ if $q_i > 0$ for all the trials.

Figure 4.10: Simulation results from Model II, plot of $c$ for $\pi' = \{4, 2, 1, 3\}$, with $\alpha \in \{0, 1\}^4$, $\beta \in \{0, 1\}$ and $\gamma_i \in \{0, 1\}$ if $q_i = 0$, and $\gamma = 0$ if $q_i > 0$ for all the trials.
Several combinations of the parameters were computed, throughout a total of 100 trials with 100 iterations each. The plots below were for a simulation with optimal policy $\pi(s_1) = u_2$, $\pi(s_2) = u_3$, $\pi(s_3) = u_4$ and $\pi(s_4) = u_1$, or $\pi = \{2, 3, 4, 1\}$ for Figure 4.9; and with arbitrary policy $\pi' = \{4, 2, 1, 3\}$ for Figure 4.10.

In all trials, the cost matrix for the optimal policy $\pi$ yields results comparable to those of Figure 4.9 in which $c(i, i)$ for $i = \{1, \ldots, 4\}$ and the set $c^* = \{c(1, 2), c(2, 3), c(3, 4), c(4, 1)\}$ presented the lowest values in the matrix. Note that the set $c^*$ corresponds to the costs of the transitions for the optimal policy $\pi$. The cost matrix shown in Figure 4.10, from one simulation trial where $\pi' = \{4, 2, 1, 3\}$, has no distinguishable pattern. The same trend was observed across the 100 trials.

Although simple, this result could indicate optimality of the policy $\pi$, as assumed. Using this simple paradigm, we can explore different policy combinations and understand their implications for the animal and/or for the inverse optimal control problem solved. If this model is further developed to include satiation, learning and other components of the basic decision-making process of the *Pleurobranchaea*, it could prove to be a novel and powerful tool in behavioral simulations.
Our investigation focused on a physiological study, an agent-based model and a stochastic model of sensory processing and decision-making in our animal of study. Although each area focused on different aspects and with different goals, this study’s unified objective was to further our understanding of the neuronal processes regulating sensory information and behavioral choice in the nervous system of the *Pleurobranchaea californica*.

Intracellular staining indicated possible critical neurons associated with chemosensation and with connections in the *Pleurobranchaea*’s main sensory organ, the oral veil. The difference in the axonal diffusion of cobalt used during the staining could be responsible for the variability in the neurons identified across the ganglia filled. Although we assume that the sensory neurons inervating the oral veil are located in the cerebropleural ganglion, it is possible that sensory neurons are also present in different ganglia, such as the pedal and bucal ganglia. Current work to further this investigation includes electrophysiological studies to determine influence of stimulus intensity and frequency in the expression of fictive avoidance or approaching.

Decision-making in the *Pleurobranchaea* is the output of the integration of sensory input, memory of experience and satiation. The decision to attack or avoid a stimulus source is linked to the association of a prey odor signature with a consequence through a simple learning mechanism. The agent-based computational model presented utilizes conceptual properties of the optimal forager to demonstrate the similarity of cost-benefit decisions to those expected of real foraging animals. The model closely reproduces foraging decisions of *Pleurobranchaea* as it incorporates approach/avoidance decision based on appetitive state, calculated as a moment-to-moment integration of sensation, satiation state and memory. In the model, the *Pleurobranchaea* agent behaves as an enthusiastic cannibal, and learns to differentiate safe and dangerous prey by via sensory processing and experience, similar to the
actual animal.

Furthermore, our model of the foraging and decision-making process of the Pleurobranchaea, utilizing insights from Markov processes and inverse optimal control theory, can be utilized to infer the cost function of the animal in two different simulation environments. However, there is obvious future work to extend these models and to better represent the cost function generated. Future work will investigate easily testable conditions for optimization of the cost function’s parameters. Finally, it would be beneficial to test the usefulness of the cost function solution under simulations utilizing transition matrixes extracted from behavioral experiment data. For example, in the case of addicted animals versus sober animals, the cost function could lead to novel hypothesis and methods in addiction research.
REFERENCES


