SENSORY ACQUISITION FOR EMERGENT BODY REPRESENTATIONS IN NEURO-ROBOTIC SYSTEMS

by

Jasmine Anaíís Berry

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(COMPUTER SCIENCE)

December 2020

Dedication

This dissertation is dedicated to those who have contributed to both my personal and professional growth through the years as a Doctoral scholar. Your unwavering support and encouragement over the years have made this journey worthwhile.

To Dwayne, Gwendolen, Jaymes, Yvette, Li'Ma
and especially to God.

Acknowledgements

I would like to first acknowledge my friends and family for their continual support throughout my time in graduate school. The Ph.D. is yours just as much as it is mine. I want to thank Dr. Francisco Valero-Cuevas (BBDL), my PI, for his advisement towards developing the research methods for my dissertation. Our many discussions helped pave the way for the material grounding of self-modeling and my professional dedication towards becoming a great scientist and engineer. Much appreciation is due to Dr. Alice Parker (BioRC Group), my research advisor and mentor, for her guidance, technical insight, and encouragement throughout my Ph.D. journey. She is credited for sparking my interest in the topics of Consciousness and Self-Awareness. I am grateful to Dr. Michael Arbib (USC Brain Project), my advisor, for introducing me to the concept of Brain Theory and its principles within. I also acknowledge the contribution of my examiners and qualifications committee, Dr. Bhaskar Krishnamachari, Dr. Wei-Min Shen, and Dr. Paul Rosenbloom, for their valuable feedback towards the betterment of the thesis. This Ph.D. was made possible through the combination of several disciplines, including Computer Science, Biomedical Engineering, Neuroscience, and Psychology. The dissertation would not have been possible without the support and contributions from various funding agencies who deserve recognition. To Dr. Timothy Pinkston, an eternity of gratitude for having always having an open office for me. You were one of the initial reasons I joined USC and embarked on this great opportunity. To Dr. Aftab Ahmad, my undergraduate professor and mentor, I am indebted to you for your guidance and shared intellect since NSU. I am grateful to all international collaborators, Ph.D. and Master's students for their knowledge and shared expertise. Thank you all, and I hope to give back throughout my career in ways to reflect my gratitude.

Table of Contents

List Of Tables

List Of Figures

Abstract

An ongoing engineering challenge is achieving agility, information processing, and flexibility in robotic systems. Building neuromorphic robots called NeuRoBots (i.e., robots that imitate the mechanisms of neural sensorimotor processing in animals) is one approach to accomplishing this goal. NeuRoBots offer several advantages over traditional robots and also serve as test beds for understanding the sensorimotor dynamics of mammalian neuromuscular physiology. The notion of how the anatomical brain builds a sense of self and how neuro-robotic agents can utilize body schemas (or representations) to build a sense of self have not been particularly successful due to varied and often contradictory accounts. In this dissertation I present a critical step in forming self-identified body schemas, based on physiological simulation of proprioceptive afferent signals, to determine the plausibility of whether biological signals can be used to inform the operation of a state machine. First, I demonstrate that a given movement gives rise to a distinct sensory manifold embedded in the 12-D space of muscle spindle information that is largely independent of the choice of muscle coordination strategy. Given that muscle lengths and velocities are fully determined by joint kinematics, such manifolds provide a rich set of information to use in its control. Secondly, I show that high-dimensional multi-muscle proprioceptive ensembles can usefully discriminate limb states and be utilized as a suitable classifier for inter-trajectory comparisons—but only after minimal pre-processing. Lastly, I present the concept of Sensory-Motor Gestalt, which provides a unifying framework for constructing body states into useful behaviors to understand the foundations of sense of self in hybrid robots and synthetic biological agents.

Chapter 1

Introduction

1.1 Motivation

Our motivation for this dissertation is to assist in the efforts towards building robotic systems that can acquire unique movement capabilities online that lead to classifying possible actions amid encounters of changing external circumstances. Such systems should have the capability to learn how to perform movements given their unique set of anatomical and physiological constraints. Achieving this goal requires us first to survey the main challenges within the field of computing that our intermediate goals will address.

As conventional computing reaches practical limitations for performance [172], additional computing methods are sought after for the next generation of autonomous devices and systems. A longtime technological goal of artificial intelligence (AI) and robotics is to create computational systems with functions similar to biological brains with the prospect of machines behaving and thinking like humans or mammals. Neuromorphic computing is increasingly becoming one plausible approach for accomplishing this goal as it introduces fundamental architectures that can potentially perform similarly to anatomical neurons. This essentially attempts to emulate in silicon what the biological cortex does in vivo. One application of neuromorphic computing is found in biorobotics and bio-inspired machines for enhanced information processing and hosting software models of neural processes that typically address problems of memory, perception, motor control, and multisensory integration [11]. A significant problem for both traditional robots and biorobots is the complexity of building general-purpose autonomy for executing various cross-domain tasks while learning new skills without catastrophic forgetting [173]. These systems do not entirely exist yet, as the field is still heavily challenged with single-task robotics. However, a closer examination of how our brains and bodies are designed can provide the means to make incremental strides in this direction.

The brain's primary mode of operation is to assure its host's survival, especially amid adverse and unpredictable conditions. In many ways, so should our autonomous machines behave. Biorobotics ought to be constructed with the basic physiological needs in mind, too. Llinás [111] and Carter [24] suggest our brains must execute the several objectives to properly maintain a sufficient level of autonomy:

1. Generate internal sensory signals that indicate our bodies' primal needs (i.e., food, rest, safety, and security).

2. Form a map of the world to direct us to locations to satisfy our needs.

3. Produce the appropriate actions and energy to move us there.

4. Alert us of opportunities and threats (both present and foreseeable).

5. Lastly, tailor our actions based on the requirements and constraints of our current state, goals, and environment.

At the core of these faculties is the concept of the physical self, body representations, and bodily awareness, which all together form the basis of this thesis. One can also extrapolate from this list the basis for how an organism uses its body to generate actions (according to specifics of the system's needs, mechanical makeup, and physical constraints) and what underlying physiological mechanisms are responsible for the selection of these actions. Without a functional body representation, there will be disruptions in the sense of agency, therefore negatively affecting actions towards goal-directed behaviors. Actions, choices, and decisions that an agent should be mostly identity-congruent as it has been observed in humans [136].

For brain-based systems like NeuRoBots, we seek to determine the scope of parameters necessary to comprise an artificial self? Consequently, we can answer address the implications for future biorobotic systems? We are particularly interested in how a NeuRobot can improve its "capabilities by being able to automatically synthesize, extend, or adapt to a model of its body" [82] from action-oriented body representations. These are the primary factors motivating this research. By examining sensory-motor contingencies, we can explore how bodies and their models are cognitively encoded and decoded to produce meaningful behaviors that are tailored to the system's physical constraints. The development of new hardware materials and manufacturing schemes gives way to an advanced generation of robots that seek to become increasingly power-efficient, multifunctional, adaptable, and autonomous in ways similar to biological organisms. The present work aims to provide a cognitive architecture to enable these robots to optimize their actions for decision making, efficient locomotion, and planning through the use of self-modeling techniques and body representations. Through computational methods of sensory acquisition in simulated tendon-driven limbs, we are then motivated to build a system that determines unique movement capabilities online, leading to classifying possible actions amid encounters of changing circumstances. How neuromorphic systems and tendon-driven robotics are capable of performing movements given its anatomical and physiological constraints will be further explored.

1.2 Problem Statement

Generally, in practice, robotic agents are pre-programmed to perform a pre-assigned set of specific tasks in a controlled environment. When systems are programmed to learn in such a way, they typically do so through imitation [42] [100], exhaustive iterations of execution, and simulation of the motor-to-sensory maps. In contrast, vertebrate animals usually learn by limited trial-and-error interactions with the physical world [5]. The biological approach supports learning new tasks that overlay existing capabilities, essentially demonstrating that novel behaviors can evolve and emerge through trials. To achieve this learning for our robotic agent (i.e., NeuRoBot), I propose for the acquisition and redundant integration of sensory information to be used as the driver of motor map development instead of the consequence of motor behavior; a methodological process we will refer to as *Motor Learning by Active Sensing.*

The cortical-motor-physical-sensory feedback loop, which is made possible by the NeuRoBot, should be capable of supporting independent exploration in the physical world. Including a model of self that emerges from the formalized construction and classification of sensory afferents can address several issues. One is ensuring biorobotics continue to closely align their mechanisms with biological systems, namely mapping sensorimotor representations of the body for action. Another is improving closed-loop control with sensory feedback that can be predicted in advance before it is perceived — thus being useful for systems with feedback delays. Our implementation should serve as a fundamental layer to self-modeling systems while helping launch robots capable of continuous, autonomous, and cumulative learning. For NeuRoBots, this level of functionality is ideal for deployment in environments that are not completely observable. Our assumptions are based on the following Fundamental Premises:

- **Fundamental Biological Premise**: The nervous system constantly assesses and enforces its current experienced-based estimate of body model against incoming sensory input and feasible motor actions [166].

- **Fundamental Robotic Premise #1**: The brain's body-model constantly assesses and enforces its current "hypothesis of body representation" against incoming sensory input and feasible motor actions.

- **Fundamental Robotic Premise #2**: A system's physical actions and interactions with the environment provide sufficient information to build a minimal representation of a robot's physical properties.

    - Then, the minimal body representation (implicit) of a robot is defined as a repertoire of physical actions, resulting movements and possible transitions between the actions.

    - Here, a minimal body representation may alternatively be referred to as the minimal sense of physical self.

1.3 Hypothesis

We hypothesize muscle spindle and Golgi tendon organs (GTO) signals that are available to the mammalian nervous system are also useful enough for classifying different movements with sufficient discriminability. Preliminary experimental observations have demonstrated how the evolution of feasible sensory sets (FSS) can provide variables of interest to be extracted for collection and statistical analysis. We then ultimately seek to show the validity in utilizing FSS for goal-directed motor mapping of neural-driven limbs. Anatomical bodies, such as the proposed neuromorphic and neuromechanical system [130, 85], can be modeled as a time-invariant control system that channels sensory inputs into actionable states for task classification. The afferent data collected from our analyses form topological manifolds. Therefore, by definition, we can expect the entirety of the model's state space to be maintained as a non-linear representation in the form of topological manifold maps that satisfy countability and separability conditions such that in N-dimensional, each point $p \in X$ in the topological space M has a local neighborhood that is homeomorphic to Euclidean space $\mathbb{R}^N$ for some $n \geq 0$ [76].

The combination of linear and non-linear manifold mapping methods can delineate the feasible transitions between adjoining states. Aggregating the states into a unified whole will then form the minimal computational frame a body representation (BR) that emerges as a byproduct of experiences may be specified as a self or identity that is constructed from the agent's subjective experiences throughout the runtime. When mathematically expressed, the emergent BR entails knowledge of *how* an agent can move within its anatomical constraints and how it will classify the internal dependencies (e.g., muscle lengths, muscle velocity moment arms, etc.) controlling each task action.

Consequently, this state classification method is a prerequisite for determining how individual body representations (i.e., body schemas) for neuromuscular-driven robotic systems can be self-generated from physiological sensory signals that result from a range of immature to skilled motor actions, prompting learning of useful actions and feasible behaviors. The resulting model of the neural-driven system will be produced by demonstrating that a system can learn a dynamic model of its neuromuscular behavior, self, and identity from afferent constraints (proprioception and somatosensory stimulation). Applying intrinsic motivation within the robotic agent's exploration of the state space may be accomplished with randomly selected movements coded in the form, or neural firing rates could be implemented for undirected curiosity or exploration of the state space [4].

1.4 Research Approach

Our end-goal is to equip a neuro-robot (i.e., NeuRoBot, a robot with an artificial nervous system) that forms a repertoire of physiologically-inspired sensory and motor couplings to explore and exploit physical actions and transitions among them. This dissertation's foundation is grounded in the fundamental premise that the brain's body-model continuously assesses and enforces its current "hypothesis of a body representation" against incoming sensory input and feasible motor actions. We further hypothesize that computationally modeling the biological central nervous system's bottom-up method for extracting sensory afferents to form unique motor maps can build functional action-oriented body representations for our NeuRoBot system. Revealing this coupling can show how sensory synergy is intertwined with motor combinations, thus enabling improvements to the biorobots' design and control.

To demonstrate this, I will use an alternative approach to self-modeling that incorporates supervised and unsupervised learning approaches. We acknowledge that a machine learning (ML) approach does not fully align with the brain's methodical process of forming online body schemas. Simultaneously, a manifold description more intuitively utilizes the biological data distribution in a multidimensional space [180]. Although useful for many applications, ML techniques may be prone to inaccuracies but can still be somewhat reliable when revealing the dimensional structure of the raw data manifolds collected. Combining the mathematics of manifolds and Gestalt law principles can provide a more intuitive expression of the intrinsic structure, shape, and space complexity. The dissertation deliverable is a system that performs arbitrary, flexible tasks using the topologically refined sensory-to-motor maps. Fig. 1.1 depicts the phases of this research plan that will be implemented in order to achieve the desired goal. The phases incrementally build on top of each other for the body representation of our NeuRoBot. Upon completion, we will evaluate the implementation's success against three main criteria that encompass the four phases. The system's final result should provide a computational foundation for constructing body representations, with contributions from each of the following phases.

1. Self-recognition: Extending the Mirror Neuron System, II (MNS2) for Agency in Reaching and Hand Grasping
2. Sense: Quantifying High Dimensional Feasible Sensory Sets
3. State: Organization of Sensory Afferents to Classification of Actionable States
4. Sensory-Motor Gestalt: Exploring a perceptual continuum for constructing an artificial self via Gestalt Laws

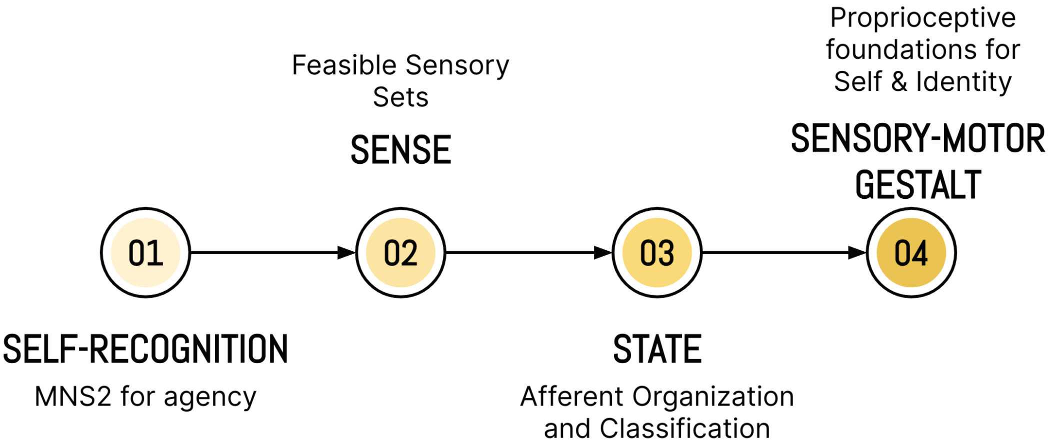


Figure 1.1: Research phase plan for building unique body representations. The research begins with a study on Self-recognition (Chapter 3) then on to Sense (Chapter 4), State (Chapter 5), and lastly Sensory-Motor Gestalt (Chapter 6).

1.4.1 Self-recognition

The first phase of our research examines the Mirror Neuron System II (MNS2). We provide a theoretical extension of the model that identifies principal neural correlates and Brain Operating Principles (BOPs) that are useful for functions of agency in autonomous systems. We use several

BOPs in two tasks: 1) a simulation of self-recognition using and 2) a hand reaching and grasping in an interactive user interface environment.

1.4.2 Sense

We transition from the identification of brain theory principles of agency in the *Self-Recognition* phase to the *Sense* phase, where we analyze the usability of physiological proprioceptive signals in the muscle spindle. The *Sense* phase of our research focuses on quantifying high-dimensional Feasible Sensory Sets (FSSs), through geometric interpretation, that can detect and categorize functional movements and tasks. It justifies the practicality of recording plausible movements for the quadruped robotic platform through a single-limb inspection of a 6-muscle, 2-link experimental human arm.

1.4.3 State

The second observable functionality of our system will be its ability to generate and identify feasible transitions among the functional movements detected in the *Sense* milestone. The *State* milestone uses simulated physiological proprioceptive signals to organize sensory afferents and classify them into actionable states. Here we determined that high-dimensional multi-muscle proprioceptive ensembles can usefully discriminate limb states—but only after minimal pre-processing. Importantly, this finding may explain the documented subcortical pre-processing of afferent signals, such as cutaneous signals processing by the cat's cuneate nucleus.

1.4.4 Sensory-Motor Gestalt

The final metric of my dissertation's success is the demonstration of sensory-to-motor-sensory maps as useful body representations of actionable states. We propose the *Sensory-Motor Gestalt* as a perceptual continuum for constructing self via Gestalt Laws. This phase is expected to assist in autonomously updating the sensory-to-motor maps of proprioceptive afferents and spindle information by forming and constraining the topology and geometry of manifold shapes.

1.5 Dissertation Outline

The remaining content of this dissertation is summarized as follows. Chapter 2 presents the necessary background information, current knowledge, and related literature for framing the right perspective of the research on constructing body representations (i.e., body schemas). This chapter will additionally acknowledge the core contributions from specialists in the field of body schemas while highlighting their competing views. Chapters 3-6 describe the experimental methodology for the four phases of work (i.e., *Self-recognition*, *Sense*, *State*, and *Sensory-Motor Gestalt*, respectively) and their relevant contributions to forming emergent body representations in neuro-robotic systems. Lastly, in Chapter 7, we interpret our results, their significance and discuss the new insights obtained.

Chapter 2

Background and Related Work

This chapter introduces our fundamental understanding of the neuroscience, physiology, and biology involved in the behavioral and cognitive phenomenon of constructing body representations. I begin with a primer on past and contemporary views on how the brain forms representations of the body. Next, I will provide insight into how the shaping of a self (i.e., body schema) may form from the mathematics of the neuromechanical perspective. Following is an understanding of how the sensory and motor interactions in muscle redundancy (i.e., the fact that we have ‘too many’ muscles). Lastly, I will analyze comparable models and applications in the field of robotics and cognition to present future improvements or trade-off for the use of our proposed model.

2.1 The Brain Represents the Body: Neuroanatomical and Artificial Shaping of the Self

Body representations are valuable as functional utilities in both biological agents and hybrid systems. The neuroanatomical basis of shaping body representations into an embodied self has been observed through empirical research, therefore providing potential computational analogs for shaping an artificial self in bio-inspired robots. Engineering flexible bio-inspired robots require an understanding of the neural interactions and computations between the brain and body. We choose to investigate the brain-body connection through the perspective of neuromechanics for vertebrates. Neuromechanics examines the nervous system's functions from within the body's mechanical constraints and anatomical structure. Within the context of this research, neuromechanics addresses the neural computational complexities that occur when the brain controls body movement. The implementation of neuromechanics presented here will cover the span of several concepts that include neuroscience, computational geometry, muscle mechanics, and anatomy [178]. Several neural computational problems arise within the brain's circuitry while it controls the body's mechanical movement. One such problem is that of spatiotemporal representations in the brain and its resulting control abilities. The brain is a self-organizing and self-repairing circuit. Its plasticity allows dynamic construction of the generated maps when external and internal changes are experienced and observed.

We argue that the nature of modeling the individual self with all of its attributes is a necessary precursor for biological agents to control their actions directly. However, some have interestingly undervalued its role and made counter-claims against it being a requirement, namely in collective systems [124]. Others have argued that a primal sense of self in animals and humans develops due to internal prospective foraging in the environment, also known as exploration and exploitation, which we will take advantage of in our implementation [81]. Next, let us consider what attempts have been made to build a model of the self [105, 191, 79] and address the self's representation within biological neural correlates [95]. We argue that the self is a by-product of the formation of sensory manifolds made available through perceptual learning. Unfortunately, the precise neural correlates for forming the self and achieving self-awareness are mostly unidentified. Gallagher's [55] initial step to this approach is constructing a primitive version of the self that ignores irrelevant features, called the "minimal" self. He assumes this model of self is most pertinent to robotic models and is reinforced by neurocognitive disorders [46] such as schizophrenia that affects the prefrontal cortex, an area thought to be critical for the formation of self. Gallese [57] provided insight on how the body in action is necessary for the building of self, which is a finding we agree with and aligns with the Exploration and Exploitation concept. The neural mechanisms that constitute the differentiation between self and other were determined by some to be the link for self-actualization [143, 182, 184]. However, several other approaches have suggested for identifying the brain regions involved: using imaging techniques (PET, MRI, and fMRI) to evaluate healthy brains and making contrasts with studies of impaired brain patients [91].

The same pattern of a lack of consensus in defining self-awareness follows in the construction of a neural basis of the self. Morin [126] challenges a common stance in neuroscience that self-awareness is based in the right hemisphere, particularly in the anterior insular cortex, which has been noted for its integration of interoceptive and exteroceptive signals in the body [34, 161, 35]. This region was observed to activate with the detection of mismatches and discrepancies between predicted signals and interoceptive signals that were perceived [68]. Morin [126] also assessed the hemispheric activity in the mirror self-recognition (MSR) and theory-of-mind (ToM) tests, amongst other self-awareness evaluation tests (e.g., self-description, autobiography). One of the conclusions drawn from this study was that both hemispheres were active during the tasks, which insinuates a distributed network of connections in self-referential activities rather than in localized brain regions; thus, debunking the right hemispheric claim. Another study involving a patient with severe brain damage to three specific regions that were once considered imperative for self-awareness development surprisingly exhibited no signs of mental degradation during self-awareness tasks. Philippi et al. [138] agree with Morin [126] in stating that we cannot pinpoint self-awareness processes to a single brain area or lobule, but instead rather distributed neural networks. Others have theorized the brainstem, posteromedial cortices, thalamus, and spindle cells in the anterior cingulate in the frontal lobe are responsible for self-awareness development [114].

The sense of agency [25] (i.e., subjective ownership and control for one's actions) is the next trait realized after establishing a model of the implicit self [86]. How agency develops is also a controversial matter. However, ownership and intent may be anatomically represented in the brain and subsequently used for the dynamic model of self. During the cognitive assessment of interoceptive signals resulting from efferent motor intentions, intentional action was observed to cause amplified activity in the anterior insular region[19]. Although there are competing theories and evidence on the matter, the prevailing theme is the distribution of neural activity across the brain is more likely for self-shaping than localized activity.

We recommend having emergent self models of the neuromuscular systems and proposing a dynamic property that emphasizes plasticity according to experiences, which is an aspect our self-model implementation will feature. Other researchers [104] have previously demonstrated robots that can build and calibrate themselves according to their subjective properties. Bongard et al.'s [15] model showed how the self could develop via movements made under the exploration of its current locomotive capabilities. "Injuries" to the four-legged robot that render a particular limb ineffective would prompt the robot to update the various models of its morphology. Consequently, compensating behavior was observed via the system's inference of its topology and parametric changes. In contrast to this implementation, which uses an actuation-sensation method to reason its own structure, we will obtain a holistic mathematical representation of the self as it evolves. Instead of forming opposing robot internal models and "generating actions to maximize disagreement between predictions of these models," our approach will primarily look at all the physically possible states of transition that occur within the manifold space.

2.2 Tendon-Driven Neuromechanics: Sensorimotor Control in Muscle Redundancy

Determining the role of sensory information in the body is an underappreciated area of study within sensorimotor research. Sensorimotor control research, both past and ongoing, has made efforts to predominately provide evidence for how the brain influences the body's actions and perceptive capabilities [72] [162]. However, the counter to these works (i.e., how the body's perception of sensory afferent shapes the brain [28] [5] [135]) is not as extensively considered until recent years as shown in Fig. 6.1. Often not taken into account are sensory states and their effects on building the brain's body composite model necessary for involuntary and voluntary behaviors. Such behaviors serve as a form of self-expression to evaluate the efficacy of one's use of functional behaviors and practical actions. Arguments have been made both for and against the view of whether sensory information's presence and quality is a necessary condition for implicit self-awareness [98]. Does the existence of a self-modeled body schema or self-awareness hinge upon the availability of sensory afferents? Most have answered this question in the affirmative and provided empirical data to support their claim. The importance of established manifolds is found in the coherence of sensory signals for kinetic energy optimization of arm movements during object manipulation tasks [47]. Platek et al. [140] hypothesized self-related information that emanates from the various sensory domains (e.g., visual, auditory, and olfactory) affects self-face recognition in such a way that enhances the priming of a model of the self, and also models of familiar faces and strangers. Incorporating different sensory domains leads to the discussion of multisensory integration [169] and how such can be a sensory set representation of an action.

2.3 Proprioception and Its Role in Bodily State Estimation

There is insufficient research on sensory paradigms in sensorimotor neuroscience compared to the numerous works solely on motor activity. However, the two should not be separated so widely. Motor planning is indeed happening in sensory space and should be taken into account. Motor control is often viewed as a direct outcome of neural activity descending from the brain's motor areas: brain stem, basal ganglia, cerebellum, and the primary motor cortex (M1). Nevertheless, we need to take it a step further. Much less attention is given to how an agent would react from an inverse of this activity. In other words, how would motor control look instead as the consequence of sensory input? We start by examining this issue by inspecting proprioception, a sense usually associated with body awareness. Proprioception can affect our learning, focus(attention), and behavior. Such a sensory system of receptors located in our muscles, ligaments, and joints is designed to articulate where our body is in space without visual stimuli. It is evident in many studies that the mind (brain) shapes the body. Conversely, how the body shapes the mind (and in turn, affects our behavior) is a dynamic that is left without sufficient understanding. These conceptual streams coincide with debates on the vitality of cohesive perception and action for effective sensorimotor control (Iberall and Arbib 1990, Mechsner et al. 2001). In this thesis, we seek to showcase how far precisely one can get with their motor control abilities by primarily targeting the senses of vision, somatosensory signals, and proprioception, as they all have the most significant bearing on output motor control. Additionally, our sensory states will initially span proprioception of limb position, joint torques, spindle signals, skin sensation, Golgi tendon organs, and kinematic frames of reference. If successful with these modalities, we can move onward to examine other senses, such as tactile feedback and auditory signaling.

2.4 Body Representations for Self-Awareness in Animals and Machines

Before explaining what self-awareness means for machine intelligence, let us ascertain how it is defined. The importance of a body representation and its features should not be overlooked when designing the architecture for autonomous individual and collective systems. It is a critical feature that will become progressively vital as technologies continue to advance in the coming years. However, the questions “What is a body representation?” and “What do body representations offer autonomous systems?” evokes answers that are neither well-defined nor understood, and are often subjectively characterized by the disciplines that define them. The perspective we chose to inspect body representations from is from the context of self-awareness. For example, Morin [125] from the field of psychology, defines self-awareness “as the capacity to become the object of one’s attention.” Nagel and Searle, philosophers of the mind, identified three features necessary for the formation of self-awareness in an agent (or self): subjectivity, unity, and intentionality [127, 53]. Subjectivity denotes the awareness of the self as a private and distinct experience of sensations. Unity in self-awareness conveys the unified singular experience that an individual or agent may have instead of separate sensory modalities. Lastly, intentionality directs consecutive moments that occur within our self-aware state to a designated goal. Damasio [37], Koch [96], and Crick [33] also agree that these features must be attended to for full comprehension of self-awareness. Of the three features, we have chosen to evaluate subjectivity, which relates to building an appropriate introspective model of the self.

Within the scope of computer science, self-awareness is fundamentally viewed as a combination of the ability to possess information about one’s internal state (private self-awareness), possessing knowledge about one’s external environment for insight on how it is perceived by others (public self-awareness), and maintaining information about future actions and decisions it could potentially make [110]. In biology, self-awareness is self-directed behavior guided by external environmental factors [10]. Some cognitive scientists portray self-awareness as the embodiment of a sense of agency and a sense of ownership [55]. McGeer [118] claims that the target of attention within any experience is the true meaning of self-awareness, but also further classifies the difference between an agent simply experiencing something and then actually having an alerted meta-awareness of such experience.

Most approaches to achieving self-awareness should be an interdisciplinary effort due to its inherently subjective nature and bias. A computational view and definition of self-awareness allow for scalability [110] and highly complex integration of nodes in a network, with the choice of even implementing self-awareness directly or as an emerging property [60]. Amongst the varied definitions, there are also many categorical types of self-awareness. Moreover, those types are discrete levels to gauge how much an agent is self-aware. Prominent researchers who have made level-type distinctions include Rochat [151], Neisser [129], Piaget [139], and Lewis et al. (2015). Rochat [151] was motivated to observe children's behavior in what has been deemed the conventional self-awareness test, the mirror self-recognition test. He questioned how the self develops over time and at what stage of development does one view themselves as a separate entity in relation to the world. He concluded that there was a range of five levels needed to describe the maturation a child experiences to reach the self-aware state:

- Level 0 – Confusion
- Level 1 – Differentiation
- Level 2 – Situation
- Level 3 – Identification
- Level 4 – Permanence

- Level 5 – Self-consciousness or “meta” self-awareness

Rochat’s approach reveals that we need to be cautious of having a dualistic view of an agent either possessing self-awareness or not, with no intermediary stages.

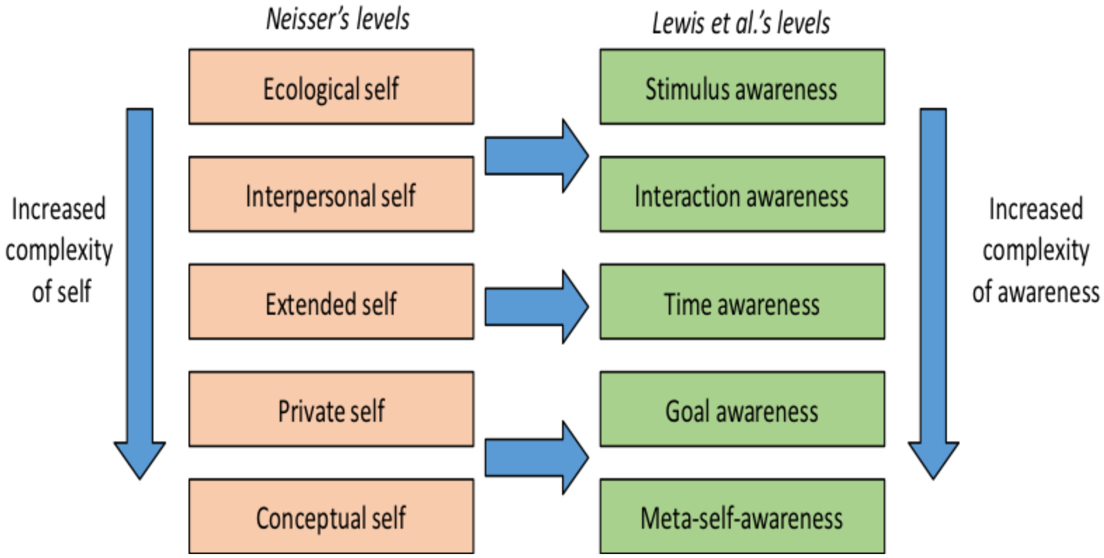


Figure 2.1: Neisser’s [129] appraisal of the various levels of self-related knowledge one must attain to reach self-awareness (left). Lewis [109] juxtaposes this with their own framework (right) of a computational perspective beginning with stimuli awareness and concluding with meta-self-awareness.

Neisser [129] proposed five different types of selves that we gradually become knowledgeable in early development to get to the self-aware state. The selves span from the ecological self in which the self is perceived with respect to the physical environment to the conceptual self, where one forms a concept of self in a social-like structure. Lewis et al. [109] took on the challenge to convert these levels, formerly based in psychology, into engineering for architecture design of computational systems Fig. 2.1. Starting with stimulus awareness, the agent is capable of using incoming stimuli to respond to events. Interaction awareness and Time awareness prompt the agent to form interactions with other systems in the environment and procure knowledge related to past and potential future events, respectively. Goal awareness preserves information about objectives and system constraints. Lastly, a meta-self-aware system maintains knowledge about its awareness. My opinion on why this framework is insufficient is its inability to serve as a physiologically realistic model according to the brain's neural underpinnings. Additionally, there should be a convergence towards a central idea and unified perspective on the manner, which appears to be lacking. Consequently, this presents a dilemma that may prevent some aspects of the field from advancing forward in the proper direction.

2.5 Implications of Physiological Subjective Experience to the Emergence of Machine Subjective Experience

In Berry and Parker [7], we gave a succinct primer of consciousness and self-awareness(SA). We issued a proposition as to how brain augmentation can influence the arrival of machine agency and self-awareness. Overall, we stated our opinion for (i) why self-awareness must be systematically examined in conjunction with brain augmentation approaches and (ii) how such a merger could become a tool for investigating subjective experiences, namely consciousness. This section will review related works that reinforce our proposal for physiological subjective experience to machine subjective experience.

The successes of the artificial retina and cochlea have lent encouragement to researchers in the general field of brain augmentation [61, 36]. However, in order for brain augmentation to progress beyond conventional sensory substitution to comprehensive augmentation of the human brain, we believe a better understanding of self-awareness and consciousness must be obtained, even if the "hard" problem of consciousness [26] remains elusive. Here we propose that forthcoming brain augmentation studies should insistently include investigations of its potential effects on self-awareness and consciousness. As a first step, it is imperative for comprehensive augmentation to include interfacing with the biological brain in a manner that either distinguishes self (biological brain) from other (augmentation circuitry) or incorporates both biological and electronic aspects into an integrated understanding of the meaning of self. This distinction poses not only psychological and physiological issues regarding the discrepancy of self and other. However, it raises ethical and philosophical issues when the brain augmentation is capable of introducing thoughts, emotions, memories, and beliefs in such an integrated fashion that the wearer of such technology cannot distinguish his biological thoughts from thoughts introduced by the brain augmentation.

A consideration of self begins with the conventional mirror self-recognition test (MSR) [60] that has been successfully executed with Eurasian magpies [142], bottlenose dolphins [145], orca whales [40], human infants typically between 18 and 24 months [2, 151], and notably the Asian elephant (Plotnik et al., 2006). The only primate species reported to pass the Gallup Mirror Test, albeit controversially, were orangutans and chimpanzees [168]. For years, MSR has been the designated litmus test for determining whether a species possesses self-awareness (SA), ultimately raising the question of whether the animal is then a conscious entity due to passing this test [39]. "Mirror self-recognition is an indicator of self-awareness," proclaims Gallup et al. [59]. If indeed so, then the subsequent query to raise is whether self-awareness, the ability to differentiate oneself among others, is a precursor to or derivative of consciousness and whether the mirror test is necessary and sufficient [126].

In light of brain research like the Blue Brain Project[78], BRAIN Initiative [89], and the development of neural prosthetics, the interest in consciousness is steadily growing. Here, we not only encourage the study of and suggest methods for addressing science's "elephant in the room," which asserts consciousness is neither physical nor functional, but also place the Elephas maximus in our proverbial mirror to obtain a perspective toward forming a cohesive alliance between philosophical studies of consciousness and neural engineering's augmentative innovations. As MSR is purposed to grant the animal subject personal physical inspection from an objective viewpoint, resulting in self-cognizance, so shall we take the approach to examine our modern scientific methods in conceptual mirrors, to appraise our consciousness dilemma and propose an assertion for progression in augmentative technologies. Following here is a succinct primer of consciousness and SA. We also issue a proposition as to how brain augmentation can influence the arrival of machine consciousness. Overall, we state our opinion for (1) why SA must be systematically examined in conjunction with brain augmentation approaches and (2) how such a merger could become a tool for investigating consciousness.

2.5.1 Ineffable Consciousness

The first pitfall encountered with consciousness is the inability to derive a functional explanation for what it means to experience. Chalmers [26] lists the "easy" problems of consciousness as "the ability to discriminate, categorize, and react to environmental stimuli; the integration of information by a cognitive system; the reportability of mental states; the ability of a system to access its internal states; the focus of attention; the deliberate control of behavior; the difference between wakefulness and sleep." These phenomena are relatively feasible to exploit and can be described in computational model terms and neural operation derivations. Chalmers then counteracts them with the "hard" problem of lacking competency to explain why and how we have phenomenal experiences when being entertained by a movie, exhibiting a sensation toward classical music, or having feelings when watching a sunset. Explaining how the brain processes visual and auditory signals is trivial compared to how those signals translate to qualia, subjective phenomenal experiences.

2.5.2 Explanatory Gap Dilemma

The term explanatory gap, coined by philosopher Joseph Levine [107], notes our inability to connect physiological functions with psychological experience, thus creating the gap. Although Levine synonymizes consciousness with subjective feelings, the explanatory gap also alludes to reasoning, desires, memory, perception, beliefs, emotions, intentions, and human behavior/action.

Correlating physical brain substrates to thoughts and feelings is the base of a dispute between two parties: materialist reductionists and non-reductionists [156]. Materialists' prevailing view, representative of most neuroengineers, on the matter involves the belief that "when the brain shuts off, the mind shuts off," and the brain is the sole causative driver for consciousness. However, non-reductionists (typically philosophers) embrace a holism approach of mandating that the brain's cortical components are insufficient in capturing consciousness, undertaking the possibility of supernatural properties. It is an inquiry of necessity and sufficiency. The brain may be necessary for mental functions, but is it sufficient? Earlier analytical inspections on conscious experience have implied that an exclusive reductive justification is not satisfactory in delineating its emergence [29, 93, 30, 48]. A novel approach is needed to explain such experience. Our explanatory gap needs an explanatory bridge.

2.5.3 Unraveling Self-Awareness Toward Augmentation

Although many facets of consciousness are difficult to investigate, the development of objective tests for SA could be utilized for brain augmented technologies. With SA comes the sense of agency. Agency imparts a sense of who is the owner of an action/trait, the self, and who represents any entities excluding self, the other(s). Self-other dichotomy processing in the brain is essential to consciousness due to the necessary implications the embodiment of "self" must have to form body ownership. Once an agent gains the ability to discern when its own body is the source of sensory perceptions, it will form body awareness that entails proprioceptive information. We can look to working experiments that attempt to showcase how the brain augments the "self" when necessary to complete a task (Fig. 2.2). Perceptual parametric information builds a premeditated awareness of (1) body part locations and (2) the manipulation of those same parts in space. Body awareness was demonstrated by a machine via Gold and Scassellati [64] who built a robot named

Nico that successfully distinguished its own “self” from “other.” Nico observably achieved self-recognition by completing mirror-aided tasks expending inverse kinematics. Nevertheless, it is believed that Nico lacked consciousness.

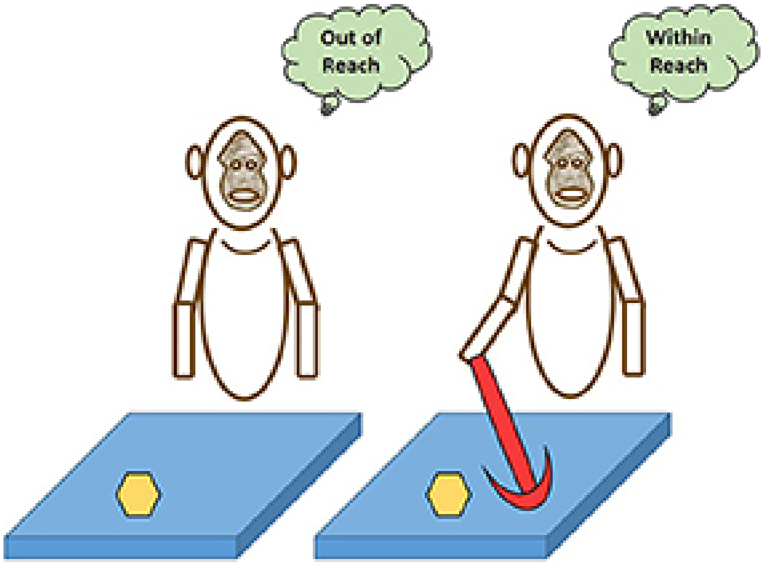


Figure 2.2: **Extension of self-representation.** Here are two depictions of macaque monkeys that exhibit a body using tools as an extension of the “self.” If given a task to retrieve an object (yellow hexagonal shape) that is outside the peripersonal space and the immediate reach of an extended limb (left macaque), the body relies on its physical limitations to define the “self” and its aptitude for the success of the task. However, when an apparatus is introduced (right macaque) that can help achieve the task’s goal, the brain’s neural correlates can augment themselves to psychophysically merge tools that were formerly considered to be of “other” classification into the “self” body schematic and permit optimal behavioral actions to take place [80, 22]. The paradigm for “self” is malleable to accept the dynamic interplay necessary to achieve an aim for a biological function that was once previously unattainable. As tool-use changes the brain’s representations of the body and alters proprioception, we subsequently believe it parallels how enriched brain augmentation can alter an individual’s self-awareness and consciousness.

Before the sense of agency becomes fully refined through experiences over time, there must be a repertoire built for perceptions and actions. Whether action and perception are interdependent or each fundamentally isolated has been the focus of another ongoing debate. It’s not yet concretely understood how the representation of self forms during the initial stages of life. Either an agent first uses perception to motivate their actions in the world or directs their actions to help drive perception of the sensory world, or both occur simultaneously. In either method, bodily awareness is eventually acquired, which contributes to defining subjective cognitive attributes. Two opposing views attempt to solve this problem: the action-oriented theory of visual perception, which suggests that perception results from sensorimotor dynamics in an acting observer [63, 131, 115], and the dual-visual systems hypothesis, which advocates independent streams of perception and action [158, 67, 84, 123]. Self-awareness uses the expectation of impending perceptions and actions to gauge the assimilation of inner experience and external reality. Building a self-aware framework in augmentative technologies requires integrating an expectancy intuition, which can critique based on differences between reality and internal experience. This is our tactic for creating systems with faculties for using perception and action to make predictions of self-sensory states, become self-adaptable to new environmental stimuli, and set objectives for self-improvement.

Crucial for understanding agency is determining how the embodied senses fuse to form self-referential experience [50, 51]. It is our opinion that future advances of brain augmentation hinge on the application of such knowledge. Once we bridge this gap of the unknown, we will be challenged to use computational intelligence to create consciousness artificially and integrate synthetic qualia with that produced in the brain. Presently, artificial devices can create various aspects of consciousness. Artificial perception is made available via cochlear, retinal, and tactile implants. However, they work alone as replacements for sensory organs with consciousness and SA arriving later in the brain's neural processing. Applications for augmenting consciousness would contribute to studies relating to emotions, attention, supplementing memory capacity, personality alteration, experience enrichment, sensory perception enhancement, and hypernormal brain plasticity for self-repair.

2.5.4 Proposed Transition to Machine Consciousness

The marvel of human intelligence is its ability to eclipse physical limitations and overcome our biological constraints to form an ever-evolving existence [87]. One primary goal for reverse-engineering the human brain is to recreate the same functional mechanisms that underlie human consciousness in our software infrastructures, neurorobotic agents, and computational systems. However, prosthetic memory, sensory implants, neurofeedback (EEG Biofeedback), and brain computer interfaces (BCIs) are all working examples of fusing such "intelligent" systems with the brain, leading to conceivable prospects for consciousness-altering devices. Although BCIs commonly target disability treatments and brain function recovery from a lesion, the amalgamation of computational devices with the cortical brain itself [52] may even prompt increasing developments of an operational "exobrain" [12] for the purposes of better understanding how our brain works. For example, in a scenario where a split-brain condition is present within a subject, we now have the option to look toward interfacing artificial exobrains with the cerebrum; such an interface can either serve as a replacement for neurological issues or supplement features the brain does not naturally comprise. If these exobrains have a modicum of manipulability, then we can explore the plausibility of mind transfer from device to organ and vice versa; thus, providing speculation for a conscious machine that can affect how we can perceive, act, express emotion, feel, and adapt. This poses ethical concerns as it opens the door for alterations of an individual's SA when augmentation can modify reasoning skills and subjective judgment. Successful augmentation of the sort might render the individual powerless in discriminating actual characteristics and thoughts from those that are mock and introduced artificially outside the cortex. Combining the precision and information processing speed of a computer with the intrinsic non-computational attributes of a human may provoke discoveries of the mind (e.g., consciousness) that we as humans are currently incapable of resolving. We suggest efforts made toward an augmentative interface between brain and machine that prompts the human mind to think beyond its unknown limits for constructing our explanatory bridge.

2.5.5 Challenges Moving Forward

Many people view an in-depth exploration into consciousness and its emergence as a gamble, considering decades already spent on the matter with a void of consensus [41, 88, 71, 164, 33, 174, 38, 160]. Before we attempt to create another hypothesis, our approach needs to change; it is our suggestion to further refine the constructs and emergence of SA and to use brain augmentation as an instrument for inspection. We need to define an objective test for determining whether an entity is a sentient being. This test, in addition to advances in neural engineering, provides optimism that disputes within the consciousness field can be resolved. Augmentation has a promising future as an enhancement to our brains and will hopefully influence our centuries-old methods of thinking about consciousness toward an answer for science's greatest mystery.

2.6 Models and Applications of Body Representations in Robotics: A Review

Now we will discuss previous attempts made to build body schemas and the applications they were designed for. Beginning with Lewis et al.'s [109] Reference Architecture framework for computing systems, we observed a common approach taken in the development of self-modeling systems; that is, forming an engineered architecture directly from psychology. Without taking into account the neuroscience, this may permit some dilemmas in accurately encompassing all of the facets related to the formation of body schemas. Their attempt to create a model in such a way was purposed to bring explicit structure to the design of self-aware systems in general, paralleling Neisser's [129] levels of self-appraisal. It categorizes various levels of self-awareness capabilities as system benchmarks (e.g., stimulus awareness, time awareness, goal awareness).

The agents' and host systems' tasks and goals determine the benchmark complexity chosen for implementation. However, their template will prove inadequate in one fundamental area: self-modeling to incorporate new features of the system. In support of the model, it does accentuate self-awareness as an ongoing going process of online learning, which we agree with to some extent. It will classify a version of this process as emergence. Lewis et al. [109] also implies that action selection directly affects the agent's ability to learn. We would like to further advance this notion by suggesting that decision making, then subsequently action selection, affects the learning and mapping of the manifold space within the self-model. The Reference Architecture additionally assigns the same goals and methodology for public and private self-awareness. Although the two domains are not mutually disjoint, both physiology and psychology sciences identified them as having different trade-offs, especially considering adaptation. Although Lewis et al.'s [109] self-aware framework proved to be effective in one case-study for a service-selection cloud computing platform, it does not provide the means for the system to learn and adapt at runtime to changing conditions. It was not explicitly stated that their implementation involved an emergent model of the self.

Through mirror perspective-taking, a makeshift humanoid robot named Nico was assembled to demonstrate self-awareness as an emergent property [75] through the goal of developing an architecture that permitted the robot to pass the classical Mirror Test [60]. The architecture is composed of six sub-models describing various levels of self-knowledge that could be obtained from the robot to complete the task: end-effector model, perceptual model, perspective-taking model, structural model, appearance model, and functional model. Many have supported this body of work with affirmative claims that this exhibited the first "self-aware" robot to pass the mirror test. However, opposing views countered those claims with the argument that the system was lacking introspection. Instead of being self-aware, Nico was classified as the first machine with the ability to reference the location of its body part in three-dimensional space by using a reflection. Our perception of this model aligns with the latter opposing claims that this model mainly demonstrated visual recognition. The robot was instructed to maintain three designated arm postures to achieve this recognition, each having 50 training sets and 100 sample tests of different positions in space. The self-knowledge procured through the training sets is then used to make predictions about the body's whereabouts via calibrated kinematics and a stereo vision system. Results indicated that the robot successfully developed a model of its arm based on its visual point of view. Self-observation, rather than self-awareness, as we noted earlier, appears to be the running theme here. It is evident Nico *observed*, but it did not obtain awareness of its experiences.

Self-aware frameworks have also been implemented in collective host systems like autonomous multi-camera networks to coordinate object tracking [147]. What interests us about this framework is the attention given to topological learning for resource adaptation among the cameras, which we feel is necessary and will incorporate in our implementation of the simulated neuromuscular system. Continuous topology monitoring will create an enduring self-model and not solely create a temporary model representation from online learning. Bongard and Lipson [14] further discuss the concepts of self-modeling in robots as self-reflection is stated as a vital aspect for robustness when encountering unexpected changes in the body.

Chapter 3

Self-Recognition: Extension of Mirror Neuron System II for Agency

> *"...the human brain is peppered with mirror neurons and they activate in us exactly what we see in the other person: Their emotions, their movements, and even their intentions." -Daniel Goleman*

3.1 Introduction to Mirror Self-Recognition

We conducted a study to identify and simulate the brain's minimal neural correlates for achieving proprieties of self-recognition and agency, a trait once deemed be a unique characteristic of only humans but has been disproven [3, 81]. In modeling the classical Mirror Test performed on Asian Elephants, we sought to provide a solution to an ongoing inquiry. In addition to the mirror neuron system (known for responding to performed actions of the self and observed actions of the other), which neural patterns are responsible for making an agent aware of its physical characteristics and behaviors? We proposed a framework, based on article entries in the Brain Operating Database System (formerly located at *http://bodb.usc.edu*), which linked systems neuroscience data to testable models and designs for the generalization of high-level concepts required for making a self-other distinction. The framework was applied in a simulation involving three-dimensional shapes, representing figurative anatomical bodies. Each was categorized as either the Agent or Adversary. Over a pre-defined set of runtime iterations of the simulation, the *Agent* trains itself with an adaptive network to optimize its ability to differentiate between the physical characteristics preserved by itself and those of others. This framework's execution served as a working example of identifying and executing the minimal components required for an agent to sufficiently reach one of the early levels of self-recognition through social interactions. Our first step towards the difficult goal of creating computational self-awareness by way of self-recognition was a software simulation of self versus other based on the "gold standard" test for self-awareness in animals and human infants, the Gallup Mirror Test [60]. We sought to successfully implement a system that can pass the Mirror Test from a biological perspective (targeting and modeling specific neurons in the brain). The mirror test assesses an animal's ability to discern its social and behavioral responses based on its reflection in a mirror. Only a select group of non-human species were reported to pass the mirror test and achieve this particular level of self-awareness: orangutans [168], gorillas, dolphins [145], elephants [141], orcas [40], macaques, Eurasian magpies [142], and bonobos. Human infants are unable to pass the test until brain nerves, and supportive tissues develop at an average age of 18 months [9, 2]. Examining this phenomenon in further detail has assisted greatly in forming the basis for this research, which extends towards forming body representations, self-modeling, and self-awareness. DARPA's recommendation [1] for designers of self-aware systems is to contemplate the architecture of a self-aware computer system from three distinct perspectives: (1) an autonomous agent view, (2) an information processing view, and (3) a biological view. A combination of all three perspectives likely serves as the most beneficial instead of only looking into one perspective alone.

The framework was designed to be extensible according to each new brain component linked to self-awareness. One of the initial questions proposed at the beginning of this research was whether the mirror neuron system alone was sufficient to solve self-recognition? Considering that the brain is highly labyrinthine and dense with neural networks, the simple answer is no. If not, then we were left to learn the additional elements that play a role in allowing the brain-mind-body complex to achieve self-awareness. Approaches taken to discover the neural circuitry associated with awareness/recognition connections are not scarce. One is the neural correlations of consciousness (NCC), which are defined as the minimal neuronal mechanisms cooperatively adequate for any specific conscious percept [33]. In this study, we reviewed recent works relating to NCC and applied some of the computational mechanisms found to be aspects of affordance extractions for the recognition of objects. Another supplementary method used is to examine disease conditions in which self-recognition and self-awareness are degraded or perturbed. Autism [187], schizophrenia [155], and psychopathy [150] can serve as for disease models in assisting with characterizing the NCC paradigm. The mirror neuron system (MNS) [148] operates as the central groundwork for our blueprint. Basic self-recognition tasks and mirror self-recognition are jointly the core of the self-awareness complex. Our proposal that the MNS is central to our architecture is founded on the notion that the MNS has been highly functional in self-other distinction recognition. The MNS has been primarily observed in experimental studies involving macaque monkeys as subjects performing various visuomotor tasks. Within the MNS are mirror neurons in the premotor area F5 that fire when a monkey performs a set of actions and observes another monkey performing that same action, if not very similar [149, 58, 13]. To understand the workings of the MNS, we must note the internal models that connect motor control to cognitive perception. Similar to feedback loops and standard robotic motor control, the framework patterned two forms of internal models that play a role in recognizing actions and adjusting the motor system accordingly: inverse and forward [119]. Inverse models were responsible for activity during the observation of actions. A mapping of the intended action and motor commands that encode an action is created, while forward models are responsible for activity during the execution of imitated actions.

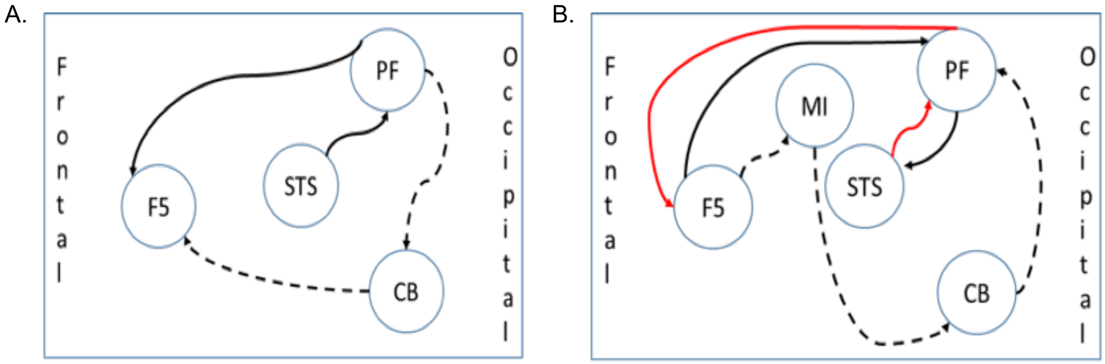


Figure 3.1: (**A**) Inverse Model: Cortical activity routes during observation of actions. The circuit linking STS, PF, and F5 (solid arrows) acts as an inverse model. The cerebellum has this function (dashed arrows). (**B**) Forward Model: Cortical activity routes during the execution of imitated actions. The circuit linking F5, PF, and STS (solid arrows) acts as a forward model to generate a prediction of movement outcome. Alternate routing is made available with the cerebellum (dashed arrows). Red Arrows: Prediction error coding through empirical Bayesian inference [92]. This figure was adapted and enhanced from Miall's concept for linking mirroring and modeling [119].

When a subject (e.g., monkey) performs an action such as throwing a baseball or views another subject perform that same action, neural processes are reported to take place in the posterior parietal cortex (PPC). Fig. 3.1 displays the inverse and forward models incorporating the F5 mirror neurons in their signaling pathways for information transfer. Mirror neurons in the PF area are shown to code for somatosensory components of the observed action [23]. Mirror neurons in the superior temporal sulcus (STS) are tasked with signaling for the visual response of biological motion, body parts (e.g., appendages and faces), and for grasp movements [137]. STS is also noted for perspective-taking. The primary motor cortex (MI) in Area IV serves as a control for voluntary movements. In Fig. 3.1A, the pathway for firing is directed from the STS to PF, and then F5. What was visually seen or conceived in the mind is recorded in the STS, then processed through PF for matching against instantiated goals. A signal is subsequently directed to F5 to induce firing if a match is made for the observed action or the subject performing the intended action. Fig. 3.1 B of the forward model contains links that are in reverse of the inverse model. The motor plan from F5 is converted back to STS, where a sensory action is demonstrated, likely visual.

Prediction error coding is used to refine motor movement to become more accurate through time and experience.

In both models, there are additional connections made to the cerebellum (CB). It was discovered to be an alternate route in the models, yet it performs essentially the same processing functions [120, 189], suggesting diffused connectivity among the cortical regions. For our simulation implementation of this particular brain mechanism, there will likely be a representation of the internal models. Suppose either of the general inverse or forward pathways are obstructed by way of a lesion, for example. In that case, the cerebellum should be able to take over and continue customary overall processing. It has been proposed that self-awareness depends on specific brain regions: the insular cortex, the anterior cingulate cortex (ACC), and the medial prefrontal cortex (mPFC) [138]. The insula is presumed as the necessary substrate for nerve impulse awareness [31, 32]. Emotional awareness, facial self-recognition, and the overall conscious experience have been linked to the ACC [35, 94]. Self-referential, self-reflective thought processing, and the projection of future self is associated with the mPFC [132, 167, 157]. On the contrary, many neurologists believe self-awareness is a product of a disseminated assortment of networks in the brain. In a study conducted on a human patient that suffered from herpes simplex encephalitis [138], it was revealed that the insular cortex, ACC, and mPFC are not a requirement for most properties of self-awareness. The patient, given the name R, had brain damage extending the basal forebrain, anterior inferior parietal lobe, medial temporal lobe, amygdala, and hippocampus. Brain damage was not found in the hypothalamus, thalamus, basal ganglia, and occipital and parietal lobes. After being probed with extensive tests, results concluded that Patient R maintained a sense of self-agency, self-recognition, and judgment. Experimentally justified, self-awareness is likely to transpire from the brain's distributed networks, including the thalamus, cerebellum, and brain stem. This justification will make a neural inspired self-aware framework rather difficult. Additionally, there are other components to consider in our initial framework. BA5 cells of the PPC should be implemented for the purpose of coding for non-dynamic kinematics. Also to be included is the anterior cingulate, referred to as sensory pain neurons, because it signals when a subject receives a touch stimulus such as being poked with a needle.

After identifying some of the key cortical regions, we looked towards encoding first-person perspective and third-person perspective in the brain. One aspect of self-awareness is the Theory of Mind (TOM), a principle used to recognize, predict, and justify the actions of both the self and other subjects. When considering the self, a first-person perspective is used and a third-person perspective is used for the other. An external study involving subjects under PET investigations revealed the brain regions activated when a subject imagines a frame of mind from either perspective [153]. The right inferior parietal, precuneus, and somatosensory cortex are involved in distinguishing perspectives. More specifically, a first-person perspective will show a strong left-hemispheric regional cerebral blood flow(rCBF) increases in the inferior, precentral gyrus, superior frontal gyrus (SMA proper), the occipitotemporal junction (MT/V5), and anterior insula. The cerebellum and precuneus were activated in the right hemisphere. The third-person perspective shows bilateral rCBF increases in the precuneus. On the left side, activation was detected in the precentral gyrus, superior frontal gyrus (pre-SMA), and occipitotemporal junction (MT/V5). On the right side, the inferior parietal lobule and frontomarginal gyrus were both activated [153]. The diffusion of areas listed here is yet another testament to the disparity of the brain activity to region ratio.

Now that we have established a few (relative to the scope of the brain as a whole) physiological components that have been found to lead to self-recognition possibly, their functions and connections should be placed in a black-box model detailing how and where neural communication links are made. Our framework from the conceptual high-level will ensure that the general concepts for attaining self-recognition are met. Fig. 3.2 shows the dated platform for what was required for self-recognition. Its objectives are to 1) recognize and generate the desired action from training,

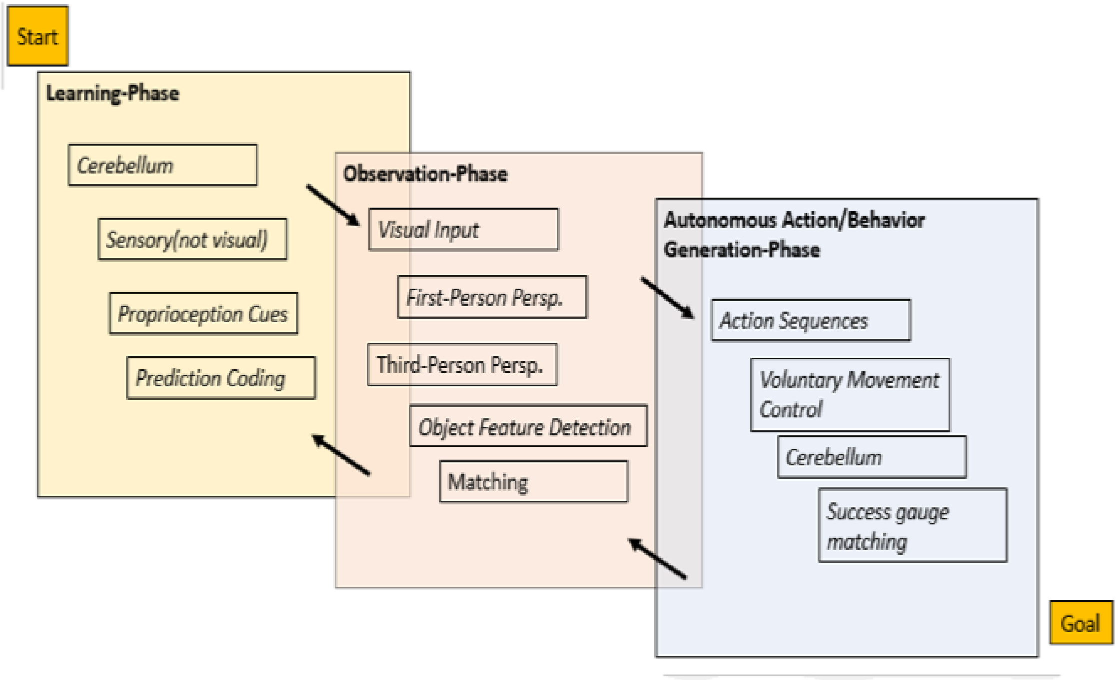


Figure 3.2: Bidirectional circuit map for coordinating visual body image and motor body image. Three phases were implemented: Learning, Observation, and Autonomous Action/Behavior Generation.

and 2) successfully coordinate the visual body image and motor body image. The themes to keep in mind when constructing this platform are cognition - the way the body is conceptualized, visual - the way bodies are sensed and perceived, and motor - the actual control and active sensing of bodily movement. The model consists of three coalescing phases of the main process occurring when a subject's brain attempts to match what is perceived through sensory inputs against internal representations. Black arrows indicate the information flow transfer through this bidirectional circuit map. Our map begins at the Learning-Phase (L-P) at the cerebellum. The cerebellum receives information about the positions in the space of the joints and the body from proprioceptors. Proprioceptive cues are sensory elements indicating factors about the body. Such cues for a robotic agent would include acceleration sensors, temperature sensing, gyroscope, touch sensor, etc. In this phase, the Agent can relate between the visual-body-image that it wants to see achieved and the motor-body-image, as previously demonstrated by Steels and Spranger

[165]. Following is the Observation-Phase (O-P), where the self-other distinction is made, mainly in the right inferior parietal lobe, precuneus, and somatosensory areas. The concluding stage is Autonomous Action/Behavior Generation-Phase (AABG-P). This phase is not finite as the map shows a feedback route for adaption by updating the system with more precise information in reverse. AABG-P provides feedback information to previous phases on whether the success of the action occurred or not. If goals involving motor plans can be accurately attained in the case of the mirror test at AABG-P, then the Agent can distinguish itself in a reflection. An additional list of information collected to assist with our task is presented in Table 3.1. Brain Operating Principles (BOPs), Summaries of Empirical Data (SEDs), and Summaries of Simulation Results (SSRs) are metrics for ensuring that our architecture is complete in including the brain's conceptual activities at the neuron level.

3.1.1 Agent Self-Recognition Test

Many tests can be used to gauge self-awareness. Two of those tests include tests of self-recognition and self-agency. Self-recognition tests seek to gauge whether the Agent can interact with its intermediate environment based on the visual scene of the mirror. An agent should have the ability to use mirrors for spatial reasoning [75]. Our simulated Agent should also readily recognize itself as a separate entity passing the classical Mirror Test. In self-agency, agents must differentiate between themselves and others that it resembles to reach the target goal location or successfully single itself out among like members. Agency also needs to entail body ownership. There must be a successful match between self, intended action, and perceived action. Here, the brain theoretically creates a representation of itself while incorporating a conscious awareness of intentionality. How the subject attributes an agent to himself or another agent will determine the level of self-agency engaged. For the software implementation, we have used a fundamental version of the self-recognition test.

Table 3.1: Brain Operating Principles

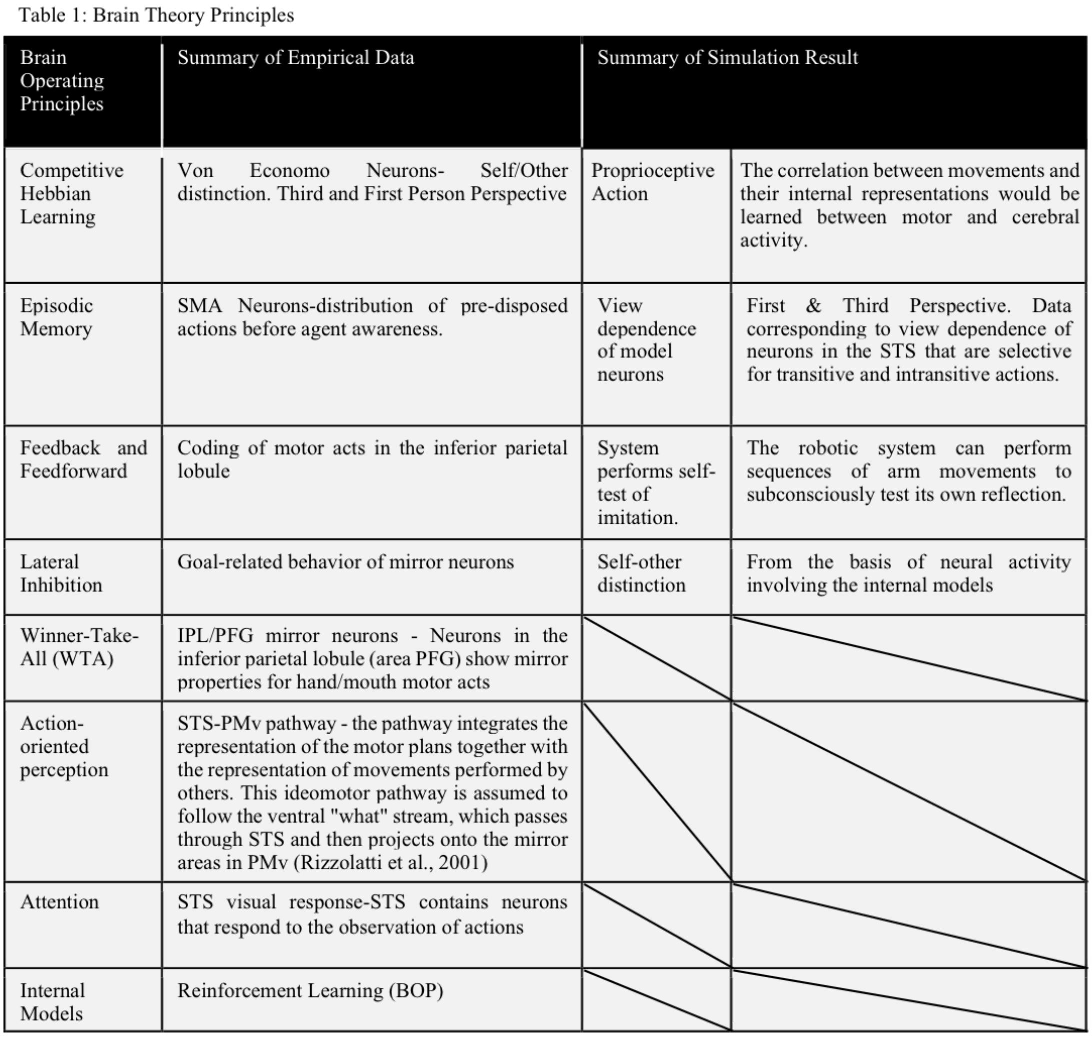


Table 1: Brain Theory Principles

Brain Operating Principles	Summary of Empirical Data	Summary of Simulation Result	
Competitive Hebbian Learning	Von Economo Neurons- Self/Other distinction. Third and First Person Perspective	Proprioceptive Action	The correlation between movements and their internal representations would be learned between motor and cerebral activity.
Episodic Memory	SMA Neurons-distribution of pre-disposed actions before agent awareness.	View dependence of model neurons	First & Third Perspective. Data corresponding to view dependence of neurons in the STS that are selective for transitive and intransitive actions.
Feedback and Feedforward	Coding of motor acts in the inferior parietal lobule	System performs self-test of imitation.	The robotic system can perform sequences of arm movements to subconsciously test its own reflection.
Lateral Inhibition	Goal-related behavior of mirror neurons	Self-other distinction	From the basis of neural activity involving the internal models
Winner-Take-All (WTA)	IPL/PFG mirror neurons - Neurons in the inferior parietal lobule (area PFG) show mirror properties for hand/mouth motor acts		
Action-oriented perception	STS-PMv pathway - the pathway integrates the representation of the motor plans together with the representation of movements performed by others. This ideomotor pathway is assumed to follow the ventral "what" stream, which passes through STS and then projects onto the mirror areas in PMv (Rizzolatti et al., 2001)		
Attention	STS visual response-STS contains neurons that respond to the observation of actions		
Internal Models	Reinforcement Learning (BOP)		

3.1.2 Experimental Design

A testing environment was created to showcase the effectiveness of our proposed framework in replicating the mirror test. An Agent vs. Adversary approach was taken. The hypothetical subjects created were five three-dimensional shapes: square/cube, rectangle, sphere/circle, triangle, and cylinder. The color of the Agent was always yellow, for simulation consistency and control. The Agent's shape is selected by the user at initial runtime, leaving the total number of possible Agents to be n=5, each having their characteristics and affordances. On the other hand, the

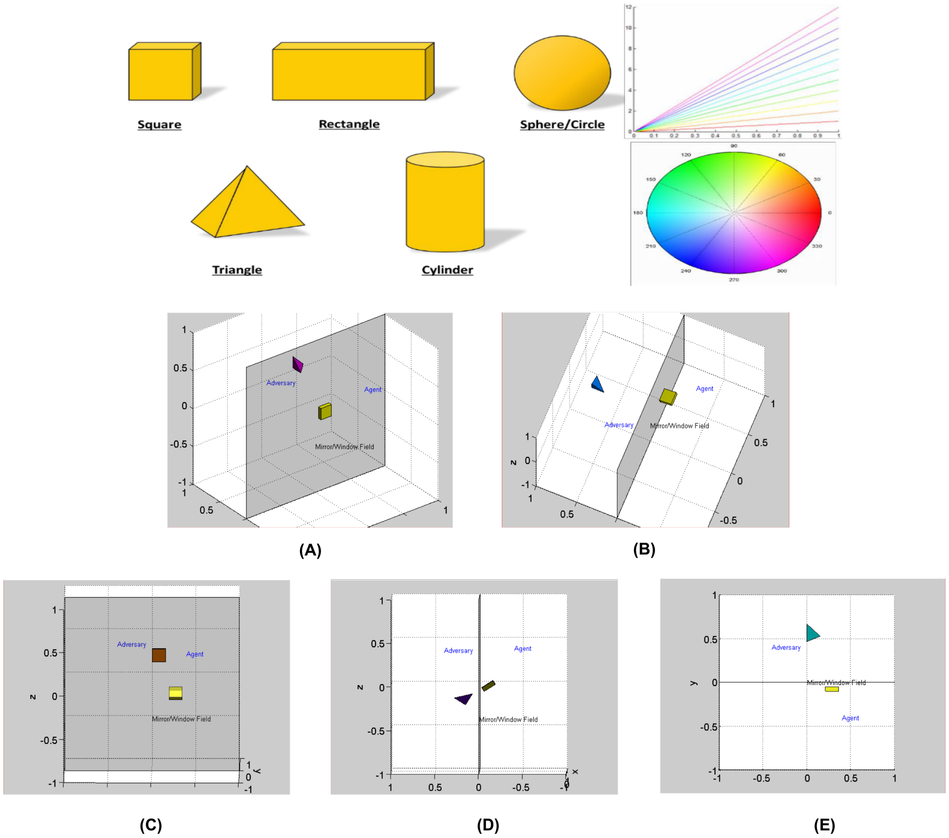


Figure 3.3: **Agent versus Adversary.** **Top row**: Five 3-D shapes were used as either the Agent or Adversary, with the color of the Adversary bring randomly chosen among the 12 HSV color hues. The Agent's color is always yellow. Shapes included were square, rectangle, sphere, triangle, and cylinder. **Bottom row**: Activity Field. Square Agent – Purple Triangle Adversary b) Square Agent – Blue Triangle Adversary c) X-axis view d) Y-axis view e) Z-axis view

Adversary's shape is randomly selected along with a random selection of 12 color hues. Fig. 3.3 gives a depiction of how the shapes are resembled in simulation and the colors chosen from the 360-degree HSV (hue, saturation, value) color wheel. The number of possibilities for the Adversary subject is 60 (12x5). When selecting the Adversary for the Agent to match against, the code is given a 30% chance to auto-generate a subject that resembles an exact representation of the Agent in both aesthetics and movement. The remaining 70% chance is the likelihood of the Adversary being one of the 60 possibilities. As shown in Figs. 3.3A-E, the environment for our subjects is referred to as the Activity Field, featured with a three-dimensional axis space with standard units on the x-y-z plane and a gray-colored margin that represents either a *Mirror* or a *Window* depending on how an Agent perceived its Adversary. The *Window* will signify that the Adversary's features and movement do not match from the Agent's perspective. *Mirror* signifies that the Agent visualizes an equivalence of its features and movements. The Agent is inevitably tasked with discriminating the perception field as *Mirror* or *Window*. Since the Agent's shapes are code-generated and do not have an actual visual perception as a human would have eyes, each Adversary shape's body parameters were created and passed to the Agent's domain as it would perceive it through a similar method to that of the Pinhole Camera Model. Several kinds of information are processed as an image for the Agent: 1) Geometric for axis positions, points, lines, and curves, 2) Photometric for color intensity, and 3) Movement direction and speed.

3.1.3 Results

Self-awareness can make a computational system more robust and self-repairing over an extended time period. This is, in retrospect, the expected hypothesis for our Agent behavior when prompted to test for self-recognition. Just as an infant's and toddler's brain continuously evolves by going through a process of pruning synapses for optimization, so should our system. As a result, the simulation required a sort of adaptive network for learning. Hebbian learning and Perceptron learning are common schemes for strengthening and weakening synapses for accurate neural firing. Better suited methods are the incorporation of Simulated Annealing and Hill Climbing techniques. These techniques facilitate our system to progressively meet their objectives in uncertain and dynamic environments. To determine which algorithm to use for our self-recognition test, the pros and cons were first assessed. Hill-climbing search only looks one step ahead at a time to decide if any successor is better than the current state the Agent is in. The disadvantage of hill-climbing is its inability to allow backtracking since it does not have the capacity to recall the previous state it has been in, due to its single state memory. Therefore, there was an issue with the state getting stuck in local minima, and the system not fully optimizing. Simulated annealing had slightly more momentum as it can escape the local minima. If the system was provided enough runtime, then a complete and fully optimal solution can be given. With simulated annealing, the system gradually degrades its repertoire of “bad” actions through a random search that accepts adaptions to increase objective realization. Thus, making it the preferred algorithm. Three main scripts were programmed to emulate our framework: Learning, Observation, and Action. At the start of simulation runtime, the program is given a specific number of iterations as the maximum opportunities for the Agent to correctly identify whether what it is currently viewing is a representation of itself or the Adversary. With each iteration, a new Adversary is randomly chosen for the Agent to examine. And also, with each iteration, the Agent is technically more “aware” of its self-properties than before. Using the mechanisms mentioned throughout this report, we should expect a familiar learning trend to an aging human infant when interacting with a mirror. In the beginning, the Agent virtually knows nothing about itself and is not certain of what traits and properties make up its being as an entity. With each passing iteration, the Agent becomes increasingly reinforced of what is the self and can expeditiously recognize it.

The range for self-awareness to be achieved in a child is 15-24 months. Scaling this range down to our IDE simulation time in seconds shows mimicry of brain development. The average life span of a human is approximately 82.5 years, which is 990 months. Taking the quotient of the months of initial awareness and the lifespan gives us a lower and upper bound value range for our system to approximate at what moment in time we should expect it to become fully aware. The lower bound is 1.5% and the upper bound is 2.4% of the system time span. For example, if our system runs for 100 seconds, we should predict that self-recognition is wholly demonstrated within 1.5 and 2.4 seconds from program start time, depending on the simulation processing rate. Several executions of the simulation were conducted to observe the effect of trial runs on the length of time full self-recognition occurs. As the Agent learns, observes, and acts in response to the Adversary, the accuracy of the Agent’s sense of self improves over time. Prediction with regards to movement, first-person perspective, third-person perspective, color, and proprioceptive cues are the main aspects the Agent must closely identify to differentiate the self from other. As mentioned earlier, the method of approach uses simulated annealing. The next iteration of this simulation test will use more conventional synaptic plasticity. With each trial run iteration, the accuracy of the Agent's perception of self is fed to the annealing script, which starts with an initial 'temperature' of 100 and maximization factor $-x^2$. Assuming that our Agent begins the simulation without any correct information about itself, our first iteration value would be -10000 (y-axis), as shown in Fig. 3.4. The stopping condition is at the threshold value level of 0, which the system attempts to achieve. Here we examine cases where there are 100, 200, 400, and 800 training sets. In Fig. 3.4, the following learning curve plot symbols and their representations are used:

- blue dot: value is accepted because the new solution is better than the current solution
- green circle: the new solution is not only better than the current solution, but also better than the best overall curve solution
- cyan dot: the new solution is somewhat worse than the current solution, but is accepted to keep in temporary memory for checking later
- red cross: the new solution is rejected because it is much worse than the current solution.

The trends we observed within our plots are 1) the closer a 1:1 ratio of green circles to blue dots as seen over successive trials indicate earlier stages of self-recognition and 2) as the iterations continue, the fewer green plots appearing on the curve show the system getting closer towards the threshold value with less chance of finding another overall best solution. Noticeably, what is seen in the latter parts of the plots, where an elongated plateau is observed, are more concentrations of all other symbols besides the green marker. In Fig. 3.4a, the 100 training sets give us a best

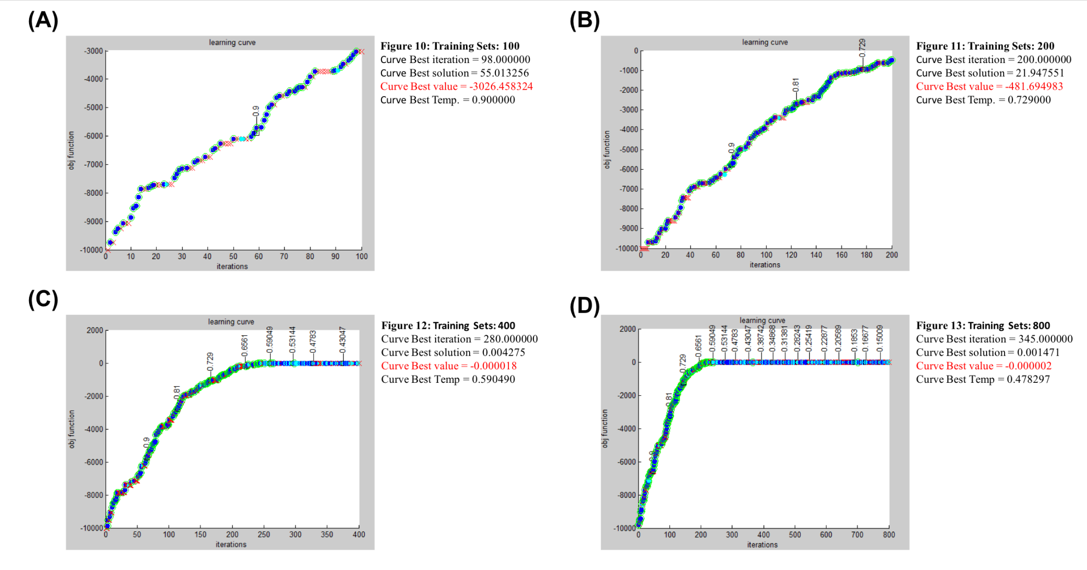


Figure 3.4: Sampled training set results for trial iterations. Results simply indicated longer trial durations produced a greater quantity of successful recognition solutions.

value of -3026.45, far from value 0. Fig. 3.4b's training set of 200 has the value -481.69. Fig. 3.4c with 400 training sets has a best value of -0.000018. And Fig. 3.4d, shows 800 training sets has a best value of -0.000002. As expected, we concluded that the longer the system can perform with trial runs to improve the accuracy of the Agent's self-other distinction, the more likely it will approximately reach the threshold level, signaling self-recognition. In the future, we seek to update this model with features that address the limitations (i.e., number of trials does not describe the effects of first exposure to a mirror) of the latest version. We are interested in viewing the system's performance if it contained prior experience with theoretical conspecifics. If possible, we would like to compile available data sets on imitation by elephants and compare them to the results of the updated version of this system that uses more neural computing over machine learning principles.

3.2 Extension of Mirror Neuron System, 2 (MNS2) for Agency

Building on the FARS (Fagg–Arbib–Rizzolatti–Sakata) (Fagg and Arbib 1998) and MNS2 [13] models, we sought to incorporate additional neurophysiological and anatomical data that highlights a compilation of vital anatomical regions that are necessary to the mirror neuron circuit's ability to predict and interpret actions that are both performed (Action) and withheld (Inaction). The goal was to construct a neurobiological simulation of the operant conditioning method to train our computational system for the Inaction and Action task in Bonini et al. [16, 17]. We address the integration of reinforcement learning and temporal difference learning in achieving results showing the anticipation of an agent's action based on a sensory cue from both the environment and intrinsic expectations. Interactions between the primate's (found in Macaca nemestrina and Macaca mulatta) cortical and subcortical regions have been identified and simulated to achieve the desired visuomotor sequences. Part I of this study particularly focused on how the system learns to perform the task. This phase develops the desired synaptic weights and encoding for each neural population identified in the Bonini experiments. Several examinations were conducted on macaque mirror and motor neurons, while the primate was instructed to perform visuomotor tasks of grasping an assortment of objects according to some external incentives. Bonini et al. [16, 17] presented scenarios that have allowed us to take an introspective look into multiple cognitive themes relating to action recognition, motor mapping for grasping, and prediction. Bonini et al. [17] show explicitly that while the majority of macaque ventral premotor neurons are silent (no discharge) when the monkey refrains from grasping an object, there are also other neurons that fire both when the monkey performs action and inaction tasks with an object. A movement is classified as an action when any muscle reflexes have followed the cue stimulus to form the grasp and an inaction when no observable movement in the monkey's hand or arm. The experiment paradigm in Bonini et al. [17] can be explicitly depicted in Fig. 3.5 (Left). Depending on the experiment's initial setup and the fixation point's placement, the monkey can infer whether the task context is an Execution or Observation. An action or inaction condition is randomly chosen, which is signaled by the audio tone's frequency. High and low-frequency tones of 1200 MHz and 300 MHz, respectively, were the only audio options used in the experiment. Most of the trials were completed with some form of light. The purpose of the Dark condition in the experiment was to ensure that all the recorded motor responses were present in the dark as well, concluding that they cannot be simply caused by the monkey's hand visual feedback.

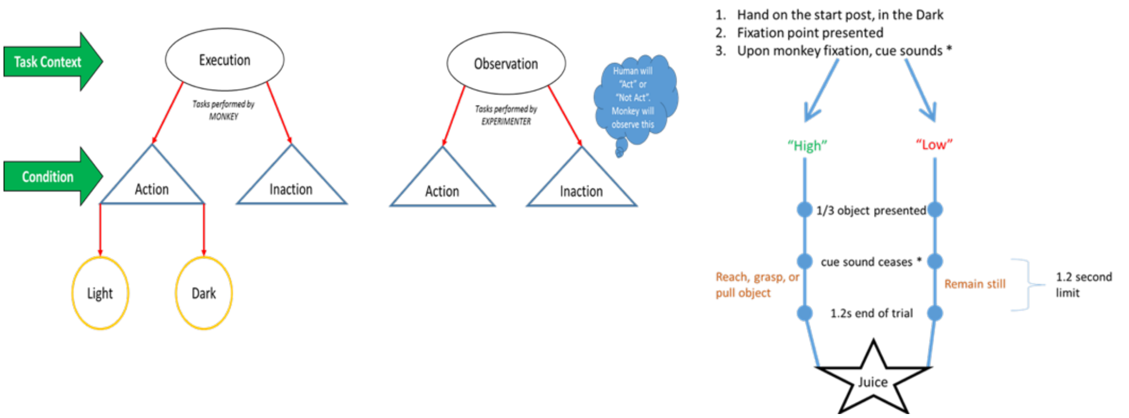


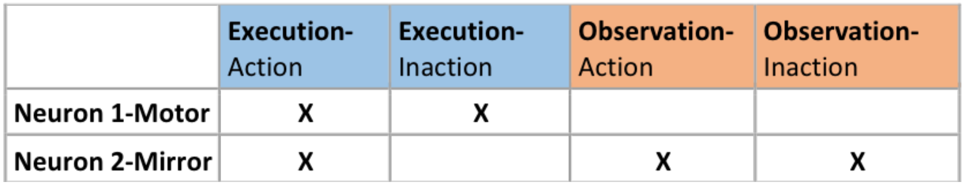
Figure 3.5: Left: Action-Inaction Paradigm. Experimental structure of the Action-Inaction task. Two task contexts were selectively chosen at the beginning of each trial. Within each context, a random selection of the action and inaction condition is made. An execution task context. Right: Epochs of interest. (*) indicates when the monkey becomes aware of 1) whether it will be acting or not and 2) whether to voluntary grasp/not grasp.

The experiment presented several epochs of interest that assisted our efforts in formulating a sound model: baseline, object presentation, pre-go/no-go cue signal, and lastly, post-go/no-go cue signal. Fig. 3.5 (right) outlines the sequential steps that occur throughout a trial. After the cue sounds, the third sequence branches off into two streams of either Action or Inaction. Although 663 area F5 grasping neurons were recorded, there were advantages in showcasing neurons individually. Only two types were selected, with a third in the supplementary paper. Table 3.2 logs these two neurons and in what conditions they were reported to fire. Neuron 1 and Neuron 2 were respectively showcased in the paper as a motor neuron, which fired when the monkey both grasped and refrained from grasping an object, and a mirror neuron that fires during grasping of execution and observation tasks and when there is refraining from grasping during observation. In other words, out of the four total conditions that the neurons can operate in, mirror neurons will *not* discharge when the Agent itself is performing the inaction. This is the case where I vouched more individual neuron tests should have been done because it may be challenging to claim that all mirror neurons do not fire in Execution-Inaction based on the performance of only one neuron. On the other hand, this behavior was seen at the population level as well. This observation does not appear overwhelmingly convincing that every mirror neuron here fires the exact same way.

All of the total neurons observed by Bonini [17, 16] can be divvied into distinct categories:

- All (n=663) neurons discharged during action execution
- 188 – 28.35% also fired during action observation: Mirror Neurons
- 105 – 15.8% also became active during the inaction condition relative to both baseline and the object presentation epoch
    - 26 out of 105 were motor (purely) neurons and discharged exclusively during the inaction condition of the execution task
    - 79 out of 105 were mirror neurons. Most (42 out of 79) discharged exclusively during the observation-inaction task.

Table 3.2: Neurons 1 and 2 fired in only certain conditions out of the four designed in the experiment.



	Execution- Action	**Execution-** Inaction	**Observation-** Action	**Observation-** Inaction
Neuron 1-Motor	X	X		
Neuron 2-Mirror	X		X	X

3.2.1 Learning with Operant Conditioning

Examining why and how to bring about certain behaviors within the brain must involve using the experimental psychology approach. To understand the full scope of why we have chosen to use specific algorithms for this particular model, it would be best first to explore the paradigm of operant conditioning and its intended purposes. By now, we should already be familiar with the action/inaction tasks that the monkey was trained to do. But how exactly it was trained appeared to be a negated detail in the official publications of the study conducted by the University of Parma's Brain Center for Social and Motor Cognition. This is a mark in our work where we believe some novelty may result. Usually, in a visuomotor task, electrophysiology data are not collected (or rather reported) before or while the primate is being trained. This brings up the question as to how do mirror neurons (MNs) develop over time to code for changes an Agent needs to make? Do MNs initially have the ability to encode specific action repertoires, or do they eventually gain this trait over time? This is just one question of many we hope that future electrophysiology tests will provide answers to. In the meantime, we have attempted to train our computational system with neural subpopulations using the same method that the actual primates in the experiment were trained on: operant conditioning (OC).

OC, somewhat different from the more widely-known classical conditioning method, resolves to goal-oriented behavior. In OC, the Agent under observation learns to behave in a way as to obtain rewards and avoid punishments. It fits the mold for learning by trial and error. OC is also viewed as a more aggressive method. In contrast, classical conditioning is associated with passiveness because it generates a behavior within the Agent that naturally emerges due to an unforeseeable connection between a stimulus and reinforcement. But it can also be argued that some classical conditioning traits are incorporated in the present experiment because the system is also learning to predict important events and not just learning the outcome of behaviors. Prediction will be a key theme throughout this study and the thesis topic. Delving more into the OC method of Applied Behavior Analysis, we have determined that the Inaction/Action task uses positive reinforcement (presenting a motivating stimulus, e.g., juice reward is given for touch start) and negative punishment (removal of desired stimulus or something "good," e.g., juice). Since the terms reinforcement and punishment will be used repeatedly, to distinguish them feasibly, it is best to remember that reinforcement is a process used to help increase the probability of behavior and punishment is the process of allowing a consequence to occur after a behavior to decrease the frequency of that behavior in future trials. We have just established the type of training method used in the Part I-Learning stage and its associated responses (i.e., operants) instrumental in the action/inaction task. But there was another element to our training procedure that we considered and implemented into the model: timing. The timing of reward and punishment also affects learning in ways that can significantly alter the Agent's ability to make definitive associations and the rate of learning some arbitrary task. Our model incorporates a varying sequence schedule to strengthen neural synaptic weights for associative learning. Schedules of reinforcement have been shown to have different effects on an Agent's behavior. Two primary schedule types were examined as potential time template structures to utilize. We have evaluated both continuous scheduling and intermittent scheduling. The first being continuous schedules, which reinforce every instance of the desired behavior with a reward. Modeling a human's reaction with continuous scheduling provides two main benefits: 1) emphatic associations will be made between the Agent's desired behavior and the reinforcements received, and 2) there will be a rapid increase in desired behaviors.

Once a behavior occurs to the desired frequency, then intermittent scheduling can be introduced. Intermittent schedules, also referred to as partial reinforcement scheduling, further encompass four other scheduling variations for behavior maintenance based on ratios: the number of responses, intervals, and time. Fixed ratio (FR) schedules are touted as the most effective mechanism for teaching new behaviors. It promotes a heightened behavior rate immediately before the expected reinforcement is received. After receiving the reward, behavior pauses momentarily and will eventually steadily increase as the time nears closer to receiving the reward again. Fixed interval (FI) schedules consist of high and low rates of behaviors and a brief pause after reinforcements. Still, it differs from fixed ratio schedules in that the rate of behaviors is generally lower because reinforcement is only given after some delayed quantum of time after the behavior occurs. One detrimental consequence of fixed intervals is that behaviors can eventually cease to exist if there is a case when reinforcement is no longer present. Variable ratios (VR) is another intermittent scheduler that is best suited for maintaining newly acquired behaviors. Here, the Agent is not knowledgeable of the amount of responses that are required to receive reinforcement. As a result, the Agent will repetitively engage in the behavior until the reward is given. One advantage of doing such will make the behavior more resistant to obsolescence. And lastly, there are variable-interval schedules. Similar to variable ratios, variable intervals (VI) also generate steady behavior rates because the Agent is unaware how much time needs to pass for reinforcement. However, the difference is that behavior rates are low due to the structure being based on the amount of time within the quantum instead of the number of reward responses. Simply put, VI behavior is reinforced after an average amount of time passes. For example, a VI-15 seconds schedule will reinforce the first instance a response is given at the average length of 15 seconds.

We have concluded that out of all the schedule of reinforcements mentioned, one supersedes the rest as it provides the most accurate template structure for our learning model. Fixed interval scheduling gives us the capability to set up the reinforcement in a manner that replicates the timing of reward/punishment delivery. Within FI, responses are reinforced after a fixed amount of time. The characteristic that sets FI apart from the rest is that it refrains the reinforcement from being automatically obtained after the established quantum. FI only makes the reinforcement available while the Agent is still required to respond appropriately to receive the reinforcement.

In addition to a temporal structure, we were also set up the experiment to recreate how associations are built. OC is primarily based on three events occurring within a trial. The events include the stimulus (S), response (R), and outcome (O). There are multiple ways in which they can be arranged to suit the task constraints. OC produces a faster learning rate if the interval space is relatively short in the beginning. But as the experiment progresses, the interval length may increase if there is prior evidence of an association that is successfully learned. According to the OC steps taken, as made explicit by Luca Bonini via external conversations, we have compartmentalized the overall OC training into five tiers to recognize and implement the phases necessary for training.

Within each tier, the primate is trained to perform certain sub-sequences of the overall task. They are the following:

- Tier 1: Start Position.
  - Monkey randomly touches start position. Reward is given, quickly.
  - Allow monkey to increase holding time before delivering reward. Reward is delayed
- Tier 2: Motor Part & Go-Tone
  - Monkey is trained to reach or grasp a target as soon as brief HIGH tone is played.
  - Reward is delayed until monkey grasps/holds object for desired time (.8s).
  - Other objects are introduced. Shaping procedure takes place for monkey to grasp objects in appropriate way.
- Tier 3: Sound Duration (a criterion of duration 0.8-1.2s is achieved)
  - Sound duration is progressively increased, making false start errors. Monkey does not get rewarded. Monkey learns to stay while sound is played, go when it stops.

- ■ Tier 4: No-Go Tone
  - – LOW tone introduced, Reward delivered as soon as sound stops. Monkey immediately associates new sound with “reward for not acting”
  - – Delay between the end of sound and reward delivery is slowly increased, up to 1.2s (max time)
  - – Alternate with Go-trials with final parameter, to reinforce previous steps of training.
- ■ Tier 5: Sound Duration (optional)
  - – Similar to Tier 3, trains system for sound duration of the LOW tone.

The tiers work together serially. Once the system has adequately learned tier 1, activity proceeds to Tier 2, Tier 3, etc. The metric for determining whether a tier has been sufficiently learned is dependent on the success rate. The learning process usually involves a type of examination afterward to gauge how well the Agent learned it. This will be the Part II-Performance stage. Instead of partitioning the experiment by the tiers mentioned above, Part II will alter the measure of success by the formal experimental steps in which the neural firing patterns and data were collected. The brevity of these experimental steps is expressed as follows: 1. Monkey subject, sitting in a chair, focuses attention on an LED. 2. A ready signal (turning on the LED) indicates the start of the trial. Monkey subject responds by placing its hand on a touchpad and fixating on the LED, indicating movement may begin. The LED is positioned such that it appears superimposed on top of the object to be grasped. The object is lighted in such a way that it is also visible. 3. LED changes color (GO signal). Monkey subject responds by removing his hand from the touchpad and reaching towards the object. Reaching phase: monkey subject preshapes his hand in anticipation of making contact with the object. 4. Contact with the object is made, and the monkey subject secures a grasp. Manipulatory movement is made (pulling or pushing the object in some direction). The resulting position is held for a random delay period.

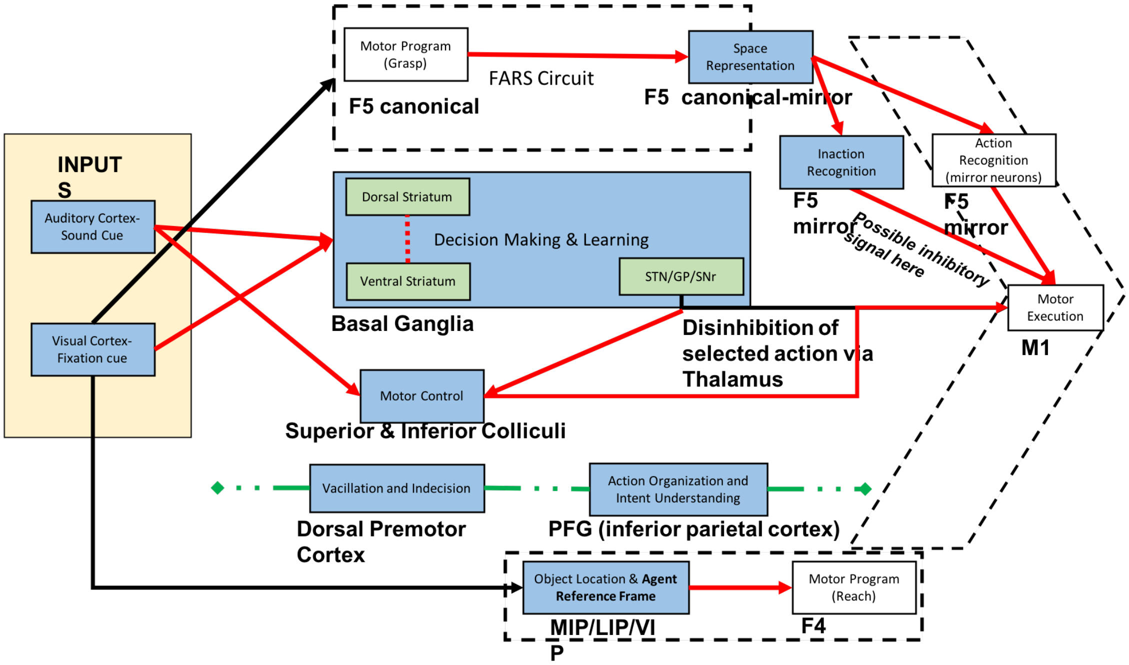


Figure 3.6: Augmented MNS2 model for Action and Inaction Task. Updates included 1) Emphasis on Basal Ganglia for learning, 2) Action and Inaction Encoding, 3) Space-Dependent Representation, and neural pathways for 4) Self versus Other pathways. Red arrows denote new connections from previous models. A new sensory modality for auditory input was added. And a Basal Ganglia unit, including the dorsal and ventral striatum, was another feature added to enhance decision making.

3.2.2 Experimental Implementation

The updated version of an MNS implementation, shown in Fig. 3.6, may be labeled “How the brain generates predictive motor representations of action based on ‘decision’ for Self inaction and visual responses to Others’ inaction.” A possible theme for this study is in questioning whether this problem is a matter of focus of attention: an IS vs. IF. “IS” places the Agent’s attention in the present moment of what is actually taking place currently. “IF” places the Agent’s attention on what possible actions can take place in the potential proximate future. In Bonini et al. [17], experimental data showed that when the macaque viewed an inaction event by the human, the same neural patterns fired here as there is firing when the monkey views the grasp in motion.

In this upcoming version of MNS, using data from Bonini's study, we want to show that motor representation can be predictively encoded when an agent executes or observes the negation of an action. Rightly so, it is necessary to provide a thorough analysis of the neural processing taking place in iterative steps of F5 neurons and other neurons that contribute. Other neurons than the F5 neurons shall either be hypothesized or conclusive to be a part of this process analysis presented in this report. In light of this, let's discuss what has to be tentatively added in F5 and elsewhere to extend the F5 mirror and non-mirror (motor) F5 neurons mentioned in the study. The objectives for the new additions are listed below. Two approaches for making model changes are to 1) reveal new brain regions that play a role in the process, and 2) discover how the already model-implemented brain regions might work differently considering the context. With that said, some proposed high-level model changes from both Bonini papers [16, 17] include:

- Adding a Prediction class: Classify MNs as either Action or Inaction MNs. They are not entirely distinct, but they do encode differently through a reactive and predictive context.

- Space-representation class: Implement region that is responsible for discriminating action in peri-personal and extra-personal space.

Neural patterns are different for action and inaction conditions in the varied spaces [21]. The objectives we felt needed to be implemented are based on the FARS and DAJ ('95) models. We sought value in incorporating areas such as the caudate, striatum, SNr, and thalamus. Several updated model proposals to consider should include the following:

1. Classify (mirror neurons) MNs as either Action or Inaction MNs—relay when they do and don't fire.

2. Model Inaction MNs (potential) inhibitory function.

3. Showcase Four Discharge Patterns among the F5 mirror neuron Predictive and Reactive for both Action and Inaction conditions.

4. Implement Canonical-Mirror Neuron region, in addition to the separable Canonical and Mirror Neuron [16]

5. Incorporating the IT cortex back in the model.

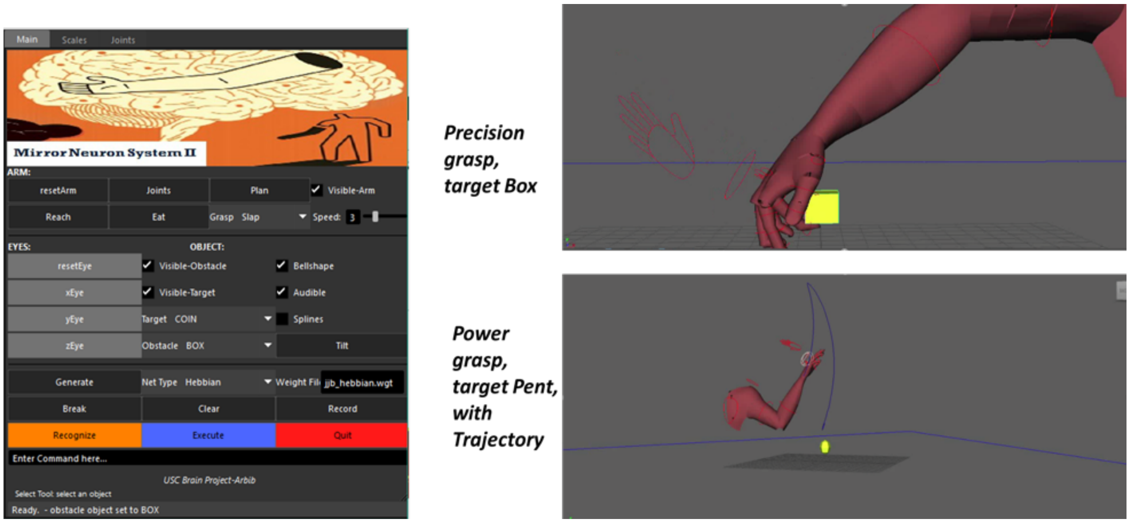
6. Incorporating the basal ganglia to deal with the motivation of motor control and learning. BG works with the thalamus and cerebral cortex to help make decisions and shift between activities. This will more than likely assist in our efforts in distinguishing pathways for Action and Inaction.

7. Incorporate the Inferior Colliculus cortex as it is used in MNS for receiving and training for auditory cues.

8. Incorporate the striatum (caudate nucleus, putamen), which receives inputs from all cortical areas and, throughout the thalamus, project to frontal lobe areas (prefrontal, premotor, and supplementary motor areas) which are purposed with motor planning. The neural circuits involving these regions (i) provides information for both automatic and voluntary motor responses (ii) assist in predicting future events, reinforcing wanted behavior and suppressing unwanted behavior, and (iii) are involved in shifting attentional sets, movement initiation, and spatial working memory.

9. Showcase the firing pattern for extra-personal and peripersonal space [16]

- Here, we can somehow show that predictive discharge occurs earlier and more frequently when neurons are tested during observation of actions performed in the monkeys' extrapersonal rather than peripersonal space.



☑ Visible-Arm
Grasp Slap
Speed: 3
☑ Visible-Obstacle
☑ Bellshape
☑ Visible-Target
☑ Audible
Target COIN
☐ Splines
Obstacle BOX
Net Type Hebbian
Weight Fil jjb_hebbian.wgt
Enter Command here...
Select Tool: select an object
Ready. - obstacle object set to BOX

Figure 3.7: Left: Graphical user interface for the simulation settings. Users have the option to select the training weight files using Hebbian and backpropagation, backpropagation through time (BPTT), and BPTT with Hebbian networks. Variation of type of arm movement can be made for a Reach or Eat trajectory, with grasping actions ranging from Natural, Side, Power, Precision, and Slap. Objects used for the target and obstacle span several basic shapes like a coin, box, and rectangular sheet, to name a few. Right: Here are two frames of the field of view of the simulation in action. The top frame depicts the trained in-motion arm performing a precision grasp of the box with the index and thumb fingers. The bottom frame shows an attempt for Power grasping the Pent shaped object, following the blue-line trajectory.

In terms of the outlook of both simulation and visualization, a setup of trials shall consist of an observation of what it is like for MNs to acquire the ability to distinguish self and other's action/inaction, therefore needing to incorporate a learning mechanism. Several initiating stimuli have been considered suitable in bringing about an agent's inaction (e.g., Not grasping a coffee cup). As you are in the process of reaching, you may be inhibited to grasp by:

1. Thought- You quickly recall the cup is empty of contents, therefore you cease your action to get a drink.

2. Vision- You see that the cup is empty, therefore there is no need to pick up the mug to drink. Grasp ceases.

3. Sensory- You feel that the cup is absurdly hot before touching. Grasp ceases to avoid burn.

4. Auditory- You hear someone yell “Stop” or “Don’t” in reference to not touch the cup. Grasp ceases.

Granted, *Thought* and *Somatosensory* may involve more in-depth simulation efforts, but it would be worthwhile to mention what factors encompass the entire phenomenon of inaction. A potential simulation and visualization of such stimuli would be to use Auditory and Vision in a “grasping” scenario of some sort. The simulated monkey will have to learn to associate cue sounds with inaction/action events with a high success rate, as mentioned in the experimental paper. Using the model to show correct grasps can be learned to be predicted by an observer in the absence of physical motion if enough appropriate stimuli are available to hint at the Agent’s goal intent. The accuracy of this prediction will gauge whether our model is sound or not. Fig. 3.6 shows the initial implementation of the software based on MNS2 that taught the simulated arm how to properly grasp various objects according to the size, shape, and affordances. The goal of this portion of the model, as shown previously by Bonaiuto [13] was to demonstrate through different training weight files how an object’s physical characteristics can alter action types for reaching grasping tasks.

3.3 Discussion

We are particularly interested in how our potential findings might be practical to the routines and architecture design used by autonomous systems to gather information (via senses) about their environment and form perceptions that prompt ideal behaviors for a given task. If our hypothesis shows the evidence that we foresee, there is an opportunity for an implementable

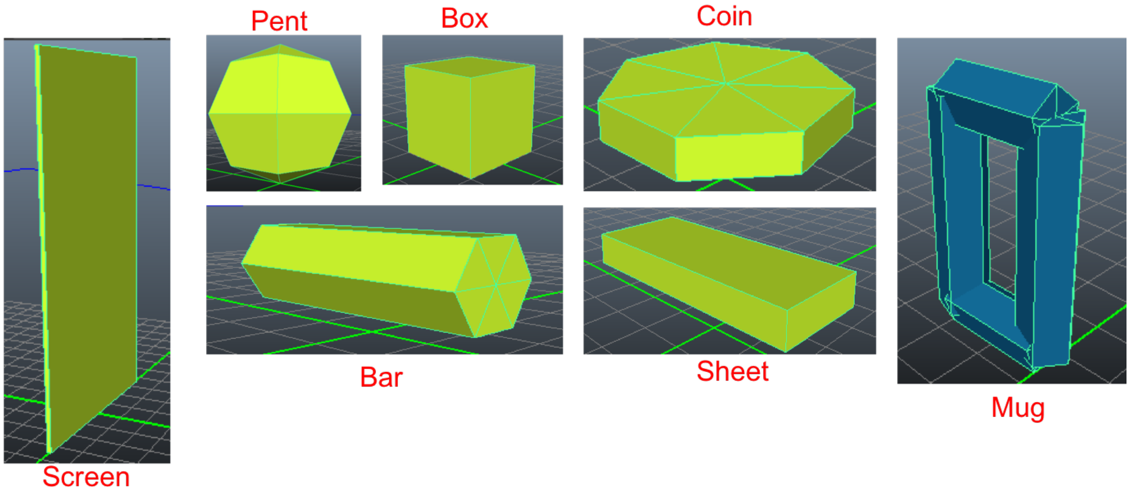


Figure 3.8: Obstacles for Interactions.

framework to be created for a diverse set of technologies to use that exploit the usage and benefits of sensory afferents. Machine learning and computational models will be used, as they are often the fundamental bases for introducing vital and fundamental breakthroughs in neuroscience on many topics. Interfacing the developed self-aware software framework with the lab's hardware robotic limbs will be one of the final results of demonstrating the working solutions. Lastly, this thesis intends to show how afferent information traverses through the nervous system, is perceived by the body, and integrates to create behavior. The experimental outcome will display how the role of sensory afferents is indeed necessary toward the development of a 'self' model that has the capacity to construct predictive motor control capabilities for the contrast inner experience and external reality.

It is not enough to solely implement self-awareness into a system. Another goal includes the pairing of self-awareness to enable more effective behaviors and resulting actions via a process called self-expression. To truly be autonomic, a computing system needs to know itself, its limits, its constraints, its capabilities, and its constrictions. In bio-inspired applications, such knowledge can be obtained and maintained through continuous high-dimensional sensory inputs like proprioception, vision, auditory signals, and tactile senses. Then a model of the self can be built using those same afferent values. What self-awareness means for machine and computational means is not well-defined and remains to be an open field with competing and contrasting theories. As self-awareness is a concept and term that is inherently subjective in nature, it is understood why there is such dissonance in the field among experts. Many facets of the field remain to be ambiguous due to a lack of agreement on how to define self-awareness and in which manner is it appropriate to duplicate it. Furthermore, we'll contribute insights from the MNS to develop a unified methodology and framework for creating the underlying principles for an emerging self-aware system with more concentration on exploiting perceptions and influx of sensory information. Following, we will address the *(i)* importance of sensory information in building the brain and the representation of self as an object, *(ii)* why sensory information is coincidentally an unexplored field in both self-aware systems and sensorimotor research, and *(iii)* why self-awareness must be holistically examined.

Chapter 4

Sense: Quantifying High Dimensional Feasible Sensory Sets

> *"The senses are gateways to the intelligence. There is nothing in the intelligence which did not first pass through the senses." -Aristotle*

4.1 Abstract

We introduce the concept of trajectory-specific sensory manifolds. They are the unique multi-dimensional and time-varying combinations of afferent signals that obligatorily emerge during a limb movement. We use the example of muscle spindles (i.e., the muscle's proprioceptors for length and velocity) that arise during movements of an arm (a planar 2-DOF 6- muscle model) during the production of straight, curved and oscillatory hand movements. Through the use of parallel coordinates, we visualize the high-dimensional evolution of the afferent signaling across muscles and tasks. We demonstrate that a given movement gives rise to a distinct sensory manifold embedded in the 12-D space of spindle information that is largely independent of the choice of muscle coordination strategy. Given that muscle lengths and velocities are fully determined by joint kinematics, such manifolds provide a rich set of information to use in its control.

4.2 Introduction: How the Body Builds the Brain

Sensorimotor control research, both past and ongoing, has made efforts to predominately provide evidence for how the brain shapes the body [72, 162]. However, the counterpart to these works, how the body shapes the brain [28, 5, 135] is not as extensively considered. Often not taken into account are sensory states and their effects on building the brain's body awareness which is necessary for involuntary and voluntary behavior. Here we investigate the flow of information underlying limb movements, and explore its significance to perceptual learning. We begin our work in this area by defining the concept of a feasible sensory set for a given movement. Our study follows the work of [101, 176, 177] that developed a theoretical framework for all possible body accelerations, activations and torques for a given tasks (e.g., feasible acceleration, activation, torque or force sets). By extension, we can also speak of a feasible sensory set (FSS). In the case of muscle spindle afferents—which sense muscle lengths and velocities—the FSS defines the set of sensory signals that can emerge for a given limb posture or movement. In particular, given that muscle lengths and velocities are geometrically defined by joint angles and angular velocities [179] then a given posture or limb movement will also prescribe the spindle afferent signals. Given a motor task, and a definition of available sensory information, we sought to define the associated manifold of spindle afferent signals that define its FSS.

4.3 Experimental Methodology

We used a simplified tendon-driven arm model where simulated muscles pull on tendons that cross, and therefore actuate, kinematic Degrees of Freedom (DOFs). The planar model had six muscles, two links (upper arm and forearm) and two DOFs (Shoulder Flexion/Extension (SFE) and elbow flexion/extension (EFE). A combination of paired antagonistic muscles formed the tendon routing of a right arm appendage: deltoid anterior (monoarticular shoulder flexor), deltoidmposterior (monoarticular shoulder extensor), biceps brachii (biarticular elbow flexor), triceps brachii

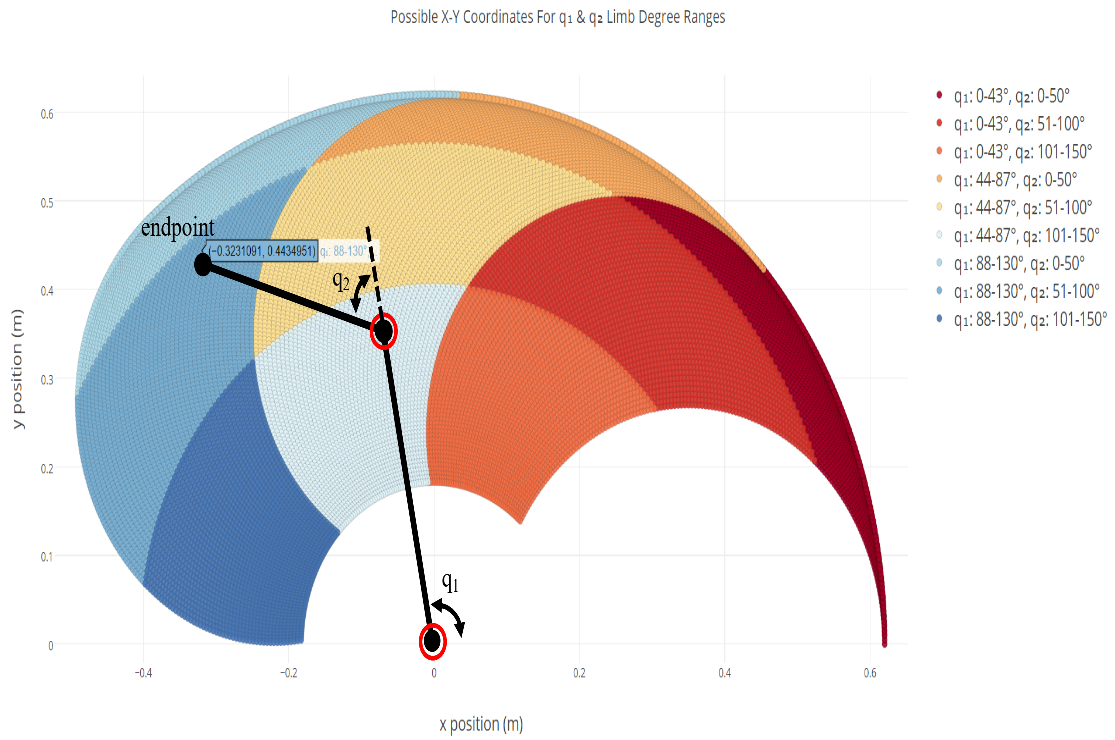


Figure 4.1: For the Static Case, all possible x-y coordinates for q_1 and q_2 degree ranges. Parameters q_1 and q_2 were constrained within ranges of motion 0-130° and 0-150°, respectively. Location of the SFE joint remains fixed at the origin (0, 0)

(biarticular elbow extensor), brachialis (monoarticular elbow flexor), and anconeus (monoarticular elbow extensor). The study was partitioned in two parts beginning with kinematic calculations of an arm during a task to derive limb joint angles and endpoint locations, and then applying those metrics to the spindle model for observation of afferent signaling. Incorporating modeled muscle parameters of optimal fiber lengths L_o, change in angle δq, and constant moment arm values r from upper extremity analyses [83] allowed the initial computation of tendon excursion (change in length of musculotendon) values, as shown in Eq. 4.1.

$$\delta s = r\delta q \tag{4.1}$$

Table 4.1: Simulated limb and musculotendon parameters.

Receptor Type	**Axon Fiber**	**Fiber Name**	**Transducer Modality**
Muscle Spindle Primary[a]	Aα	Ia	Muscle length and speed
Muscle Spindle secondary[b]	Aβ	II	Muscle strength

a. Classification of afferents and their respective function for detecting deformation of muscle tissue and transducing those signals into electrical responses. Fastest conduction speed and fiber diameter, Aα (72-120 m/s) has the thickest myelination.
b. Aβ (36-72 m/s) possess thinner myelinated axons. [90]

We investigated how limb movements affect two chief elements of muscle afferentation for muscle length/contractile velocity. Using inverse kinematics [159], a Static Case was used to find all possible discrete positions our modeled arm can achieve. Variations in proprioceptive signaling are shown to be dependent on task constraints as studies have revealed active movements tending to report more accurate proprioceptive approximations [54, 193], which led us to incorporate a case with continuous arm movement in dynamic settings. The Dynamic Case consists of specific trajectories the arm follows over a set time frame that can be modified via the Speed Factor parameter. As the Speed Factor increases so does the velocity of the movement under observation. Within the Dynamic Case, we evaluated our arm limb moving in several tasks starting with the Circle trajectory (in counterclockwise direction) as illustrated in Fig. 4.2A. The Straight Linear trajectory task consisted of five distinct pathways on a plane for the arm to follow, each perturbed at the slope by a 0.1 decrement (Fig. 4.2B). The Oscillatory trajectory represented sinusoidal movement with an angular frequency of 6π, amplitude of .05m, and .35m vertical shift (Fig. 4.2C). And lastly, we defined the symmetrical lobe Lemniscate (i.e., "figure of eight") trajectory [146] (Fig. 4.2D) using the mathematical expressions shown in Eq. 4.2 and Eq. 4.3. It must be noted that the derived configuration spaces only disclose exclusive θ values for q_1 and q_2 despite the possibility of a multiplicity of joint angles producing the same end-effector position.

$$x = \frac{\alpha cos(t)}{(1 + sin^2(t))} \tag{4.2}$$

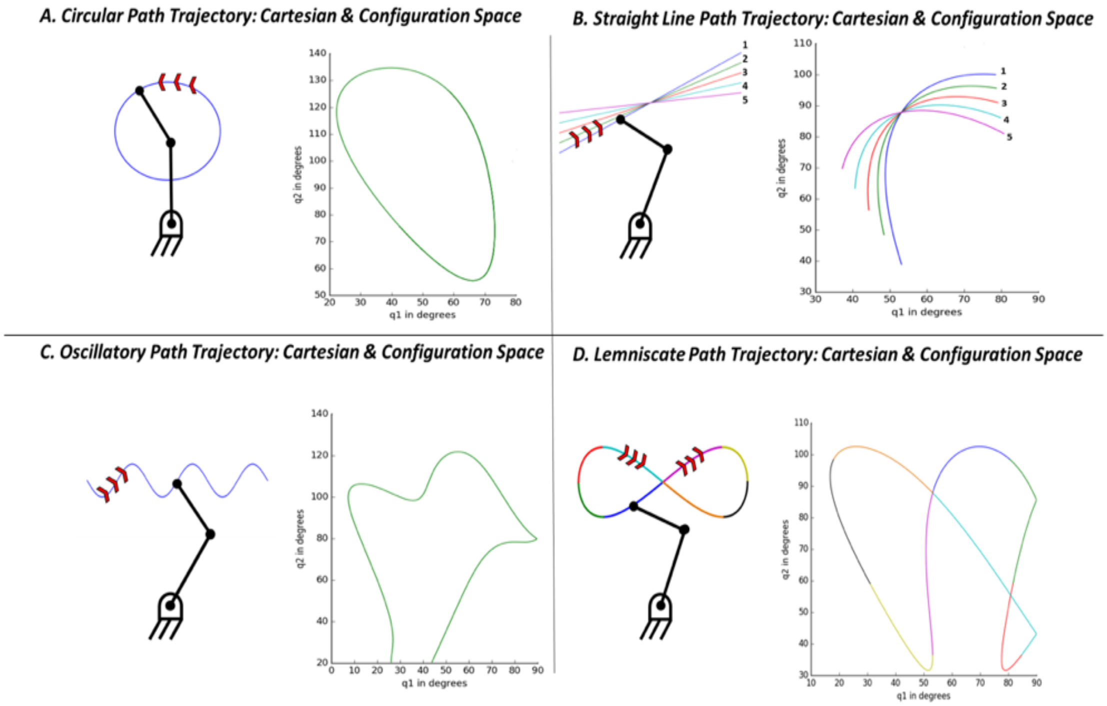


Figure 4.2: Cartesian space and Configuration space of arm movement in directions indicated by the red cursors for (**A**) the arm limb in action during the Circle Trajectory task in the counterclockwise direction. Configuration space illustrates the joint angles for 360 distinct postures. (**B**) Arm limb in action during the Straight Linear Trajectory task in a left to right direction. Line 1 trajectory, in blue, sustains a slope of .5. Line 2 trajectory, in green, sustains a slope of .4. Lines 3-5 follow according with a negative .1 gradient. Configuration Space illustrates the joint angles for 1,000 distinct postures. (**C**) Arm limb in action during the Oscillatory Trajectory task in a left to right direction. Configuration Space illustrates the joint angles for 1,000 distinct postures from the leftmost to rightmost point along the trajectory. (**D**) Arm limb in action for Lemniscate Trajectory task with symmetrical lobes. The depicted path is partitioned according to color scheme for mapping the end-effector location in Cartesian coordinates to the joint angles illustrated in the Configuration Space, which illustrates for 1,000 distinct postures.

$$y = \frac{\alpha sin(t)cos(t)}{(1 + sin^2(t))} \tag{4.3}$$

After solving for the joint and limb kinematics, we utilized a computational sub-model to simulate the biological spindle as observed in mammalian muscles, namely that of the cat [121, 122], which has also been used in human simulations [163, 103]. Action potentials in pulses per second (pps) were generated for primary (Ia) and secondary (II) afferents based on the interactions of the intrafusal fibers (chain, bag1, bag2). The first analysis that we performed examined whether afferent signals are dependent on muscle velocity throughout a task. We varied the Speed Factor in the system by a combination of values ranging in ascending speed: 0.0005, 0.005, 0.05, 0.5, and 1. One-way analysis of variance (ANOVA) of the measured spindle signals under these varying velocities tested whether there was a significant difference between the group output values. Velocities were categorical and set as the independent variable while the spindle firings served as the continuous dependent variable. Our second analysis developed the high-dimensional sensory space for Ia and II afferent signaling to extract the sensory afferent sets for the Dynamic tasks.

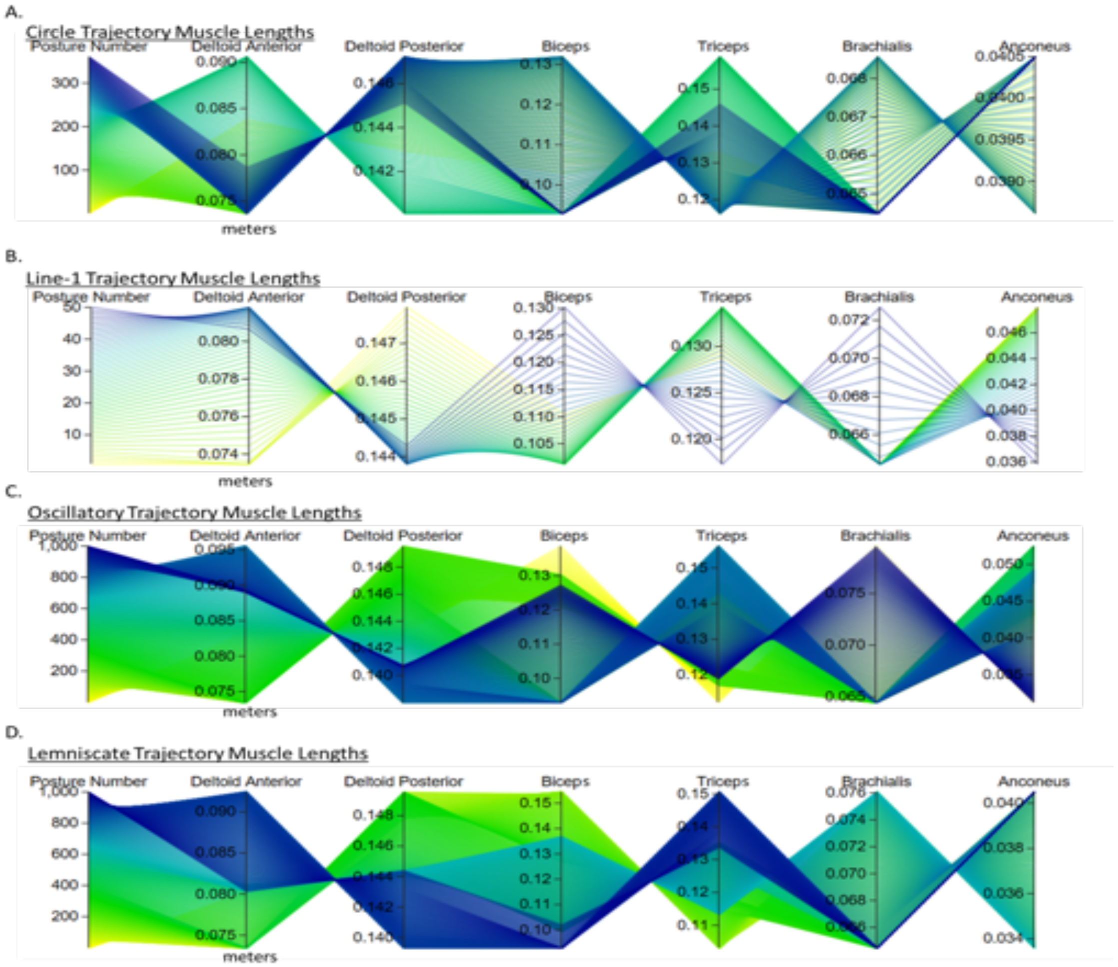


Figure 4.3: Six-dimensional representation of change in muscle length along four trajectories of the Dynamic Case. Color gradient depicts initial postures(yellow), intermediate postures (green), and concluding postures (blue). (**A**) Muscle length values (meters) during Circle Trajectory task. 360 postures were examined ranging from Posture 1 at 0 radians to Posture 360 at 2π radians. Direction of movement along the trajectory is counter- clockwise. (**B**) Muscle length values during Line 1 Trajectory task. 50 postures were examined ranging from Posture 1 at the leftmost point on the line to Posture 50 at the rightmost point. Lines 2-5 follow the same paradigm of movement sequences just with an altered slope. (**C**) Muscle length values during Oscillatory Trajectory task. 1,000 postures were examined ranging from Posture 1 at the leftmost point on the sinusoidal wave to Posture 1,000 at the rightmost point. (**D**) Muscle length values during Lemniscate Trajectory task for 1000 postures.

4.4 Results

4.4.1 Kinematics Assessment

Parallel coordinates were used to clearly illustrate the multi-dimensional change in muscle lengths for each posture during the tasks of the Dynamic Case. Such assessment was conducted to verify the efficacy of our model to ensure isometric, concentric and eccentric contractions according to physiological expectations [106, 183]. As shown in 4.3 we sampled n postures along each trajectory (i.e., task) and integrated δs from the initial posture. The three pairs of antagonistic muscles showed the expected concentric and eccentric contractions along their respective continuous trajectories. The muscle lengths were differentiated to derive the obtain their respective velocities and accelerations, which served as direct input parameters to each spindle model.

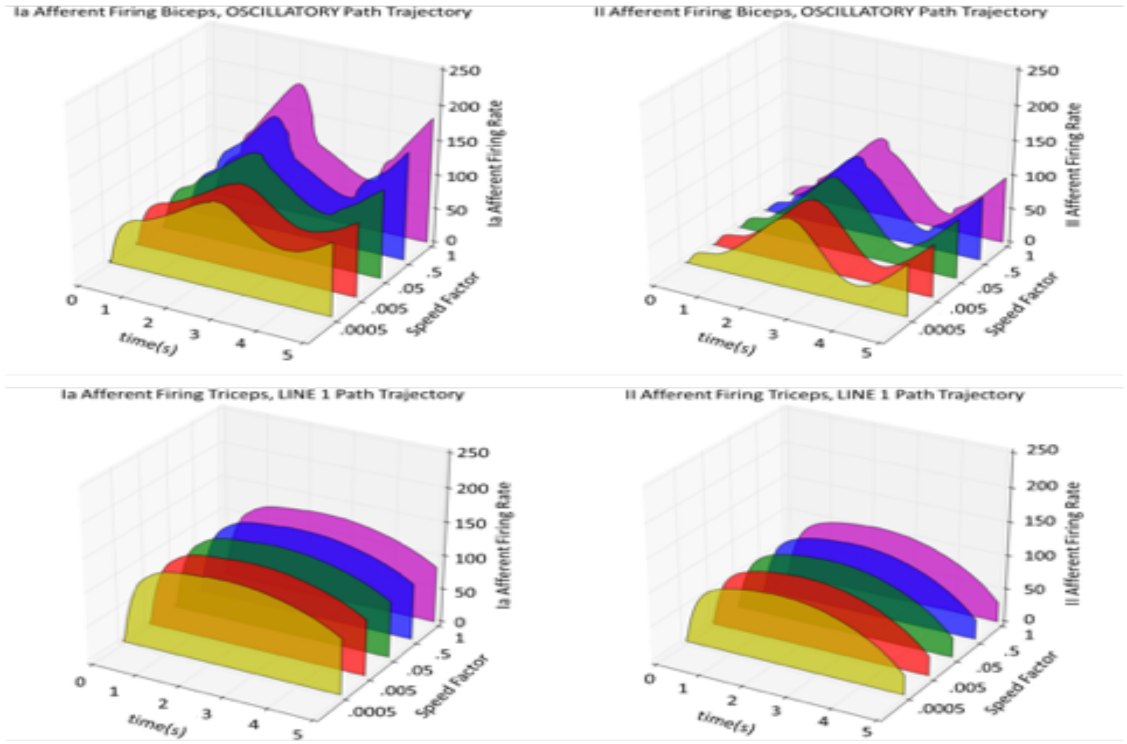


Figure 4.4: Velocity speeds versus afferent signals in Group Ia (left) and II (right). Five values were used for the Speed Factor, with value 1 signaling the fastest speed across the task. **Top row**: Afferent firing in the biceps muscle for the Oscillatory task shows slight oscillations with increasing speed. **Bottom row**: Afferent Firing for the triceps muscle shows a uniform and smooth signal throughout the span of Speed Factors. After

4.4.2 Afferent Signaling Dependent on Muscle Velocity

We expected that increased muscle velocities would affect spindles, and therefore, "body sense" in a heavily nonlinear way. Fig. 4.4 provides a sample of our observations. We detected the presence of perturbations in the Circle, Oscillatory, and Lemniscate; prospectively owing to the curvature of the trajectories which can induce abrupt changes in velocities. These finding corroborate observations from [121]. ANOVA tests revealed p-values for each muscle's Ia and II afferent signals in each task of the Dynamic Case (Table 4.2). For those cases where $p \leq 0.05$, we rejected the null hypothesis that there was no difference between the defined sets of velocities and the resulting afferentation in each task. We detected significance in only particular groups of muscles in the Circle and Oscillatory trajectory tasks. The Straight Linear trajectory demonstrated no significant difference across all six muscles for each Speed Factor value (Fig. 4.4), while the Lemniscate trajectory showed evidence for all muscles having affected afferentation.

4.4.3 Sensory Bounds According to Task Constraints

We also used parallel coordinates to describe Group Ia and II signals in the Dynamic Case with a reasonable duration of 5 seconds. Our sampling frequency (fs) was set at 10 kHz (10,000 samples/second), resulting in 50,000 time samples throughout each trajectory. Fig.4.5 presents the high-dimensional correlated relationships among afferent signals. To read these parallel coordinates, please note: 1) each axis is likely to have a different scale depending on the range of values reported for that muscle, 2) adjacent dimensions are more easily interpretable than non-adjacent dimensions, and 3) a web-based view provides the ability to interactively analyze subsets of activities of single muscles (as shown in Fig. 4.5 for the Line 1 & 5 trajectories). Therefore, we can explore the multivariate comparisons, patterns, and sequences that are unique to each muscle and trajectory. For example, for the Group II afferents of the Line 5 trajectory, we isolated the signals on the triceps muscle between 50-150pps. This revealed the associated firing rates for

Table 4.2: Velocity Significance in Afferents (ANOVA P-values)

	Fiber Type	**Circle**	**Line 1-5 ($\overline{X}$)**	**Oscillatory**	**Lemniscate**
Deltoid A.	Ia	0.049	0.249	<0.001	<0.001
	II	0.830	0.861	<0.001	<0.001
Deltoid P.	Ia	0.538	0.999	<0.001	<0.001
	II	0.999	0.999	0.087	<0.001
Biceps	Ia	<0.001	0.981	<0.001	<0.001
	II	0.213	0.999	<0.05	<0.001
Triceps	Ia	<0.001	0.753	<0.001	<0.001
	II	<0.01	0.950	<0.001	<0.001
Brachialis	Ia	0.362	0.999	<0.001	<0.001
	II	0.850	0.999	0.072	<0.001
Anconeus	Ia	<0.001	0.474	<0.001	<0.001
	II	<0.001	0.967	<0.001	<0.001

other muscles as related to the triceps: deltoid anterior 0-5 pps, deltoid posterior 80-130 pps, biceps 35-60+ pps, brachialis 0 pps, and anconeus 185-250+ pps. Similar introspections of signal bandwidth can be made for all other muscles. We also have the ability to trace and correlate any subset of physiological with kinematic variables such as velocities, accelerations, muscle lengths and stretch, and limb position in space.

To gain insight into the robustness of spindle afferents, we performed Monte Carlo simulation [144] for each of the six muscles with variation of the gamma static and dynamic fusimotor drive values. Within 100 trial iterations, boundary limits on both $\gamma_{dynamic}$ and γ_{static} were set to inclusively span 70 and 150 pps. Maximum standard deviation between any given set of the observed points approximated to 20pps, consequently resolving to a 10% deviation estimate of the signal as $\gamma_{dynamic}$ and γ_{static} were constrained at a constant rate of 100pps.

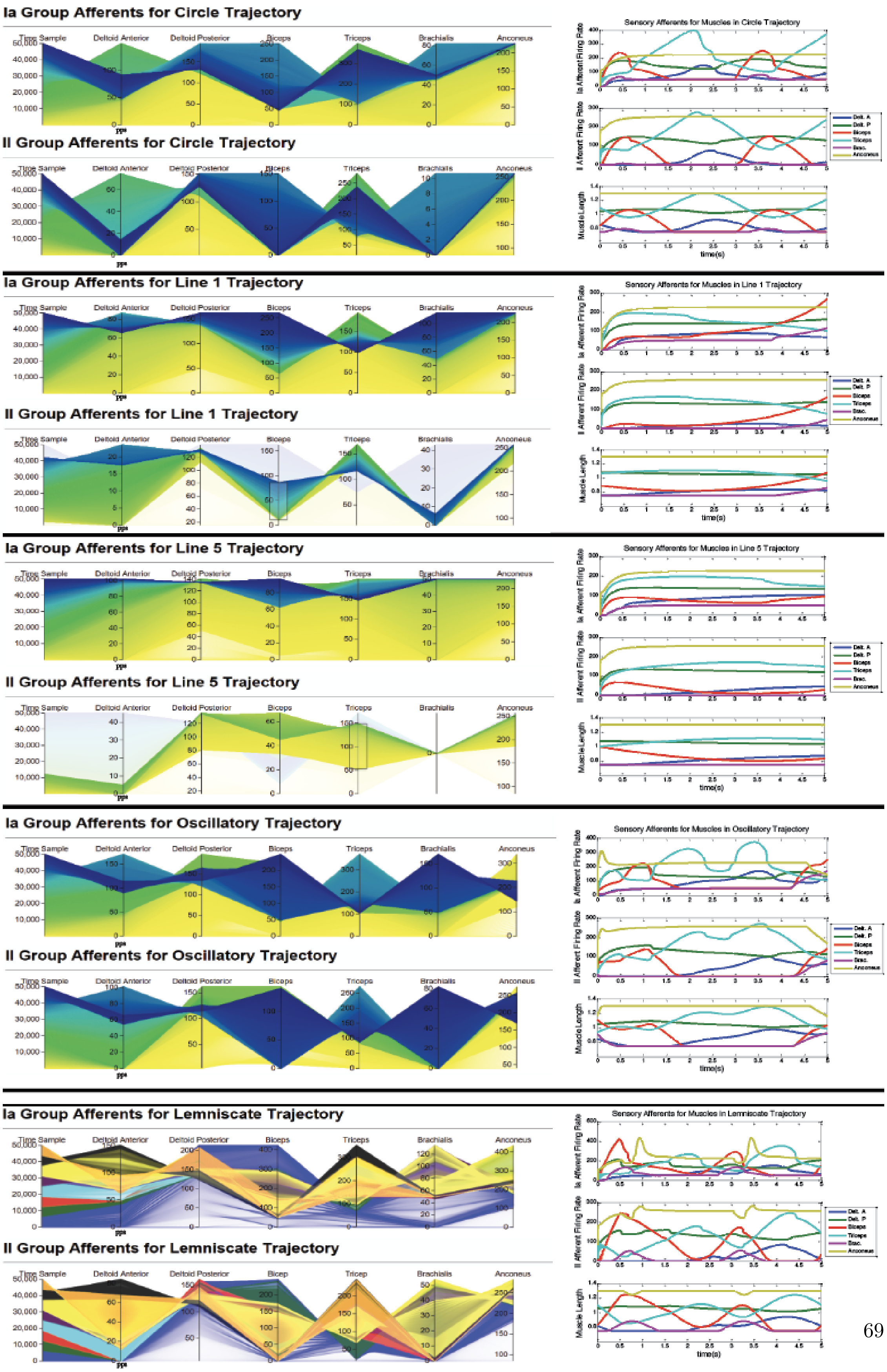

Ia Group Afferents for Circle Trajectory
II Group Afferents for Circle Trajectory
Ia Group Afferents for Line 1 Trajectory
II Group Afferents for Line 1 Trajectory
Ia Group Afferents for Line 5 Trajectory
II Group Afferents for Line 5 Trajectory
Ia Group Afferents for Oscillatory Trajectory
II Group Afferents for Oscillatory Trajectory
Ia Group Afferents for Lemniscate Trajectory
II Group Afferents for Lemniscate Trajectory
Time Sample
Deltoid Anterior
Deltoid Posterior
Biceps
Triceps
Brachialis
Anconeus
pps
Sensory Afferents for Muscles in Circle Trajectory
Sensory Afferents for Muscles in Line 1 Trajectory
Sensory Afferents for Muscles in Line 5 Trajectory
Sensory Afferents for Muscles in Oscillatory Trajectory
Sensory Afferents for Muscles in Lemniscate Trajectory
Ia Afferent Firing Rate
II Afferent Firing Rate
Muscle Length
time(s)
Delt. A
Delt. P
Biceps
Triceps
Brac.
Anconeus

Figure 4.5: Primary and secondary afferent space for the Circle, Line 1, Line 5, Oscillatory, and Lemniscate trajectory tasks marked by 50,000 samples in a time interval of 5 seconds from starting position to ending position along the prescribed trajectory. **Left side**: Parallel coordinates showing the activation of each group of muscle during a sampling range along the trajectory. **Right side**: Spindle afferentation of each muscle according to the range 0.75-1.3 of the optimal lengths. The parallel coordinate afferents for the Lemniscate trajectory is mapped with the color segments used in Fig. 2D. It can now be observed which location along the Lemniscate trajectory produces a certain afferentation value.

This allows us to quantify how fusimotor activation, naturally, affects both motor capabilities and body sense. Furthermore, we present the importance of how each movement leads to a very specific set of sensory information. We can then propose the concept of feasible sensory manifolds, FSS, associated with each movement task.

4.5 Discussion

How do our sensory signals shape the motor choices we make in daily life? In this paper, we addressed the body sense that arises from muscle spindle afferents, to enable future studies investigating how those same sensory signals affect the representation of our physical self and the actions we make. As per equation 4.1, we know that muscle excursions and velocities are completely determined by the time history of joint angles (so long as muscle tone prevents any muscle from being slack in any posture). Thus, every limb movement is associated with a unique set of specific sensory states, the FSS. As such we must consider how the nervous system obtains and processes sensory data to create a body sense that interacts with explicit or implicit internal models of the body, and external influences on the body. There are several perspectives on how sensory data (mostly visual) leads to perceptual states: the action-oriented theory of perception, which suggests that perception is the result of sensorimotor dynamics in an acting observer [63, 115, 131] and the dual-visual systems hypothesis, which advocates for independent streams of perception and action [67, 66, 84, 123, 158, 175]. Recognized predominantly as the Perception-Action Cycle [128], various methods developed from this framework may be utilized to replicate the decision making that occurs during the process of acquiring sensory modalities regarding the external world [113]. In the context of neuromechanics, we posit that sensory data obtained in any moment is dependent on the kinematic posture, position, and action task of the respective limb producing the sensory stimuli. Our present study delved into the consequences to sensory systems towards representation of high-dimensional observability to complement the controllability of muscle- driven limbs; specifically, within the mammalian muscle spindle. Our methods for obtaining these results can be employed towards systems such as robotics and brain-machine interfaces (BMIs) that are optimized on the limits of simulated neural drive obtained from sensory inputs. This concludes the first phase towards focusing on the categorization of sensory states. Results for this phase are primarily based on the results documented by Berry et al. 2017 [8].

Chapter 5

State: Sensory Afferent Organization to Classification of Actionable States

"Thoughts make the plan. Actions make the man." -Unknown

5.1 Abstract

High-dimensional proprioceptive signals like those from muscle spindles are thought to enable robust estimates of bodily states. Yet, it remains unknown whether spindle signals suffice to discriminate limb movements. Here, we used a 4-musculotendon, 2-joint limb model to simulate muscle spindle II and Ia signals (length and velocity, respectively) during repeated cycles of five end-point trajectories in forward and reverse directions. We find that cross-correlation of the 8D time series of raw firing rates (four Ia signals, four II signals) cannot discriminate among most movement pairs (only 29% by one measure). However, projecting these signals onto their 1st and 2nd principal components greatly improves discriminability of movement pairs (82% by that same measure). We conclude that high-dimensional multi-muscle proprioceptive ensembles can usefully discriminate limb states—but only after minimal pre-processing. Importantly, this may

explain the documented subcortical pre-processing of afferent signals, such as cutaneous signals processing by the cat's cuneate nucleus.

5.2 Introduction

Physical behavior in vertebrates is made possible by hierarchical neuronal systems that send motor commands from the central nervous system to muscles on the basis of sensory information coming from the peripheral nervous system. Motor function has received much attention given the relative ease with which the activity of $\alpha-$motoneurons and muscle can be measured and associated with physical behavior. In contrast, the emergence of somatosensory 'percepts' (i.e., the transformation from spike trains from an ensemble of mechanoreceptors to a neural impression useful to the control of movement) has proven much more challenging to understand. This is because the action potentials from mechanoreceptors on the skin, muscles and joints are not easily isolated or recorded [170], and the somatosensory percepts they elicit in the central nervous system cannot be readily inferred.

The lack of understanding of the physiological bases of somatosensory percepts is particularly problematic to the study and theories of sensorimotor control [112]. In particular, the somatosensory percept of proprioception, also called kinaesthesia, provides the sense of self-movement and body configuration/position [49]. Rigorous neurophysiological work on mechanoreceptors has led to the fundamental tenet of sensorimotor control that muscle spindles (whose II and Ia afferent fibers encode the length and velocity of each muscle) provide necessary, if not sufficient, limb configuration information for adaptable, accurate, and robust control of limb movement. This is supported by the geometrically obligatory relationship between joint angles and muscle lengths, but also challenged by the facts that muscles often span multiple joints and that spindle signals can be modulated independently of joint angles by $\gamma-$motoneuron drive to their intrafusal fibers. We are not aware of conclusive evidence of this tenet, which is adopted to the point that other mechanoreceptors also affected by joint angles (i.e., synovial capsule, ligaments and skin) and Golgi tendon organs are considered secondary for reasons detailed in the Discussion. However, this has not been demonstrated experimentally because spindle afferent recordings from numerous limb muscles in peripheral nerves or dorsal root ganglia cannot be obtained during large limb movements.

Therefore, we performed a computational experiment to assess the utility of muscle spindle afferents to provide usable limb configuration information. A minimal requirement for utility is the statistical notion of discriminability. Discrimination test are employed in sensory evaluations and analyses. Discriminability has been used to test how raw and processed signals from skin mechanoreceptors on the fingertips can be used to distinguish among different edges and textures to inform manipulation [133, 154]. In our case, we performed pair comparison tests to evaluate the extent to which raw and processed ensembles of Ia and II spindle afferents signals, during five distinct limb movements (Fig. 5.1), could discriminate among the five limb movements that produced them.

5.3 Experimental Methodology

The computational design of the simulated tendon-driven system, the trajectories selected for inspection, and the modified spindle afferent model will be described. Then we'll detail the methods of pre-processing and filtering used to reduce the dimensions of afferent signals. Our pre-planned trajectories produced afferent signals that were compared in inter-class contexts in then processed in data series estimation, pattern identification, and unsupervised machine learning algorithms on the resulting afferents to reveal their spatial and temporal dynamics. Lastly, we'll conclude with a review of how the feature selection and extraction techniques were implemented to determine which relevant spindle model features maintained substantial effects in classifying one trajectory from another within sensory space.

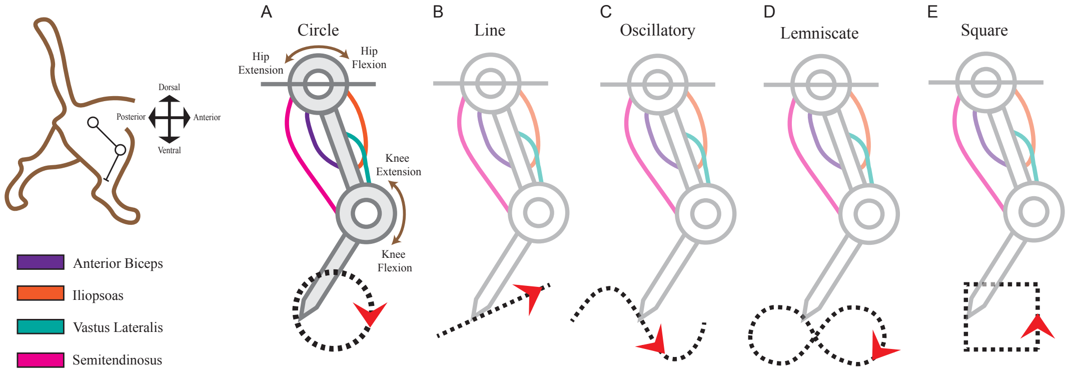


Figure 5.1: **- Limb kinematics were derived from distinctive trajectory types.** The 2-joint kinematic model, which was fitted with four muscles found in the cat musculoskeletal hindlimb structure encompassed a muscle group that included anterior biceps, iliopsoas, vastus lateralis, and semitendinsosus. The following five planned trajectories, with the red arrow indicating its direction of motion, were selected for comparative analysis. (**A**) Circle trajectory maintained motion progression in clockwise direction. (**B**) Line trajectory completed a full cycle by . (**C**) Oscillatory trajectory replicates the sine wave curve oscillations. (**D**) Lemniscate trajectory is a polar curve that is usually referred to by the name "figure-eight". (**E**) Square trajectory maintained motion progression in counterclockwise direction.

5.3.1 Kinematic Model Structure and Parameters

We constructed a simplified tendon-driven leg model, represented as the feline hindlimb, with a pivot at the hip joint. In tendon-driven anatomies, tendons are responsible for permitting muscles to act on vertebrate limbs and actuating the kinematic Degrees of Freedom (DOF) [181]. The planar model consisted of four muscles, two links, and two DOFs (Hip Flexion/Extension and Knee Flexion/Extension) connecting the thigh and shank, as shown in Fig. 5.1. For simplicity we excluded actuation of the foot (i.e., paw), which is normally included in a feline hindlimb model and would be more representative of the actual feline. Lengths of the thigh and shank segments were set to 90 mm and 100 mm, respectively, with musculature comparable to the muscle-joint interactions and parameter data resulting from system identification analyses [74] that were based on mathematical properties of skeletal muscle formulated by Zajac [192].

To imitate the useful dynamics of the cat's hindlimb mobility, we captured the movements of the leg as generated by 4 muscles: *Anterior Biceps (AB)*, *Iliopsoas (IL)*, *Vastus Lateralis (VL)*, and the *Semitendinosus (SM)*. Table 5.1 summarizes the parameters we used in the musculotendon structure which contained parameters of maximal length as L_{max}, constant moment arm values as r, and optimal length values (L_O) per muscle at the reference angle. Fig. 5.1 depicts the tendon routing of AB, IL, and VL as unifunctional joint muscles. AB and IL are acting in paired antagonistic form on the hip. VL activates knee extension movements, while SM serves as a bifunctional joint muscle acting on both the hip and knee. According to Harischandra and Ekeberg [74], the resting (neutral) posture of the hip at 65° and the knee at 100° maintained mono-articulated muscles at a length of 85% of L_{max} and 75% for bi-articulated muscles.

Table 5.1: Simulated limb and musculotendon parameters.

Muscle Name	L_{max} **(mm)**	**Angle Movement**	**Moment Arm (mm)**	**Reference Angle** L_O
Anterior Biceps	70	Hip Extension	30	85%
Iliopsoas	70	Hip Flexion	-44	85%
Vastus Lateralis	50	Knee Extension	9	85%
Semitendinosus	70	Hip Extension Knee Flexion	30 -38	75%

Optimal lengths (L_O) of each muscle at the reference angle were set to 85% and 75% of L_{max} for unifunctional and bifunctional muscles, respectively.

5.3.2 Trajectory Planning

Arbitrary shapes were selected as pre-planned trajectories in two-dimensional planar space for the end-effector limb positions. All trajectories were performed in closed loops and mathematically expressed as parametric functions of time, t, to obtain the x and y coordinate locations. The cat limb executed five point-to-point movements that will be further referred to as *task representations*. Each task representation contained a total of 200 equidistant points per cycle on the trajectory. One full cycle lasted for a time frame of one cycle/second.

The first task representation is the Circle trajectory (Fig. 5.1A) which prompted the limb to perform uniform circular motion within a 5 mm diameter in the clockwise direction. The end-effector's total distance traveled approximates to 15.71 mm. Next, the Line trajectory (Fig. 5.1B) positioned the end-effector on the path of a straight line to simulate smooth, uninterrupted movement along a ramp. Relative to the horizontal plane, the line segment retained a 50% incline at 26.57°steepness. Its midpoint position was at the 100 mm y-intercept on the Cartesian plane. The total distance travelled for one cycle of the Line trajectory was 12 mm.

The Oscillatory trajectory (Fig. 5.1C) is a sinusoidal wave forming a path of a smooth periodic oscillation. Using Eq. 5.1 as a function of time, the amplitude A was set to 20 mm with a frequency, f, of 10Hz. The angular frequency, w, expressed in radians at run-time Eq. 5.2 along with zero phase shift, φ.

$$\omega = 2\pi f \tag{5.1}$$

$$y_n(t) = A\sin(\omega t + \varphi) \tag{5.2}$$

The Lemniscate trajectory (Fig. 5.1D) created two symmetrical and uniform-sized lobes to form a shape resembling the "figure-of-eight" curve [146]. The curve was formed using parametric curves from Eq. 5.3 and Eq. 5.4.

$$x_n = 4 \times 10^{-2} \sin(5 \times 10^{-1} t) \tag{5.3}$$

$$y_n = 2 \times 10^{-1} \sin(t) + 11 \times 10^{-2} \tag{5.4}$$

Lastly, we prescribed the Square trajectory (Fig. 5.1E) as a proximity comparison to the Circle trajectory movement. Considering squares and circles are topologically equivalent shapes, we expected to view closer spatiotemporal similarities in the sensory space between these two shapes over others. However, squares differ in their non-continuity and finite lines of reflectional symmetry which also might reflect symmetry within the afferent manifolds. To what extent will the afferent signals reflect these features in the observed kinematics and make them muscle activities and joint motions distinguishable is one facet of the experimental outcomes we sought to observe.

The limb joints on the planar limb actuate as revolute joints with links capable of rotating around it. The 2 links comprise of an end effector which maintains the foot position at the end of the shank link and also the end of the articulated body. While the hip position remained affixed as the root joint, we calculated the tracing of the end effector position across each of the five trajectories using inverse kinematics. For each trajectory, the 200 target positions in the Cartesian space were selected as inputs for the inverse kinematics algorithm and the limb pose (i.e., state) required for the target position were derived to determine the joint angles at the hip and knee, q_1 and q_2 respectively.

Inverse kinematic solutions are generally not unique, and are sometimes dependent on the initial joint coordinated q_0, which typically defaults to value 0. However, the θ values for q_1 and q_2 of the limb were successfully obtained despite the possibility of a multiplicity of joint angles producing the same end-effector position. Given the desired limb's end-effector positions, for each time step across the trajectory at instance i, the segment link lengths, l_1 and l_2, and the coordinate positions, x_1 and x_2, were recorded to calculate variables c and s in Eq. 5.5 and Eq. 5.6, respectively. Joint angles q_1 and q_2 for each segment were then iteratively derived using equations Eq.5.7 and Eq.5.8.

$$c = \frac{(x_i^2 + y_i^2 - l_1^2 - l_2^2)}{(2 l_1 l_2)} \tag{5.5}$$

$$s = \sqrt{1 - c^2} \tag{5.6}$$

$$q_1 = \sin^{-1} \frac{y_i(l_1 + l_2 c) - x_i l_2 s}{x_i^2 + y_i^2} \tag{5.7}$$

$$q_2 = \cos^{-1} \frac{x_i^2 + y_i^2 - l_1^2 - l_2^2}{(2l_1 l_2)} \tag{5.8}$$

Once the limb's joint angles are calculated, a Jacobian matrix can be generated to determine the relationship between simulated limb's joint parameters and the end-effector velocities. The change in joint angles are then used as inputs for the muscle spindle model to obtain raw sensory afferents for each trajectory.

5.3.3 Muscle Spindle Afferent Data Collection

In a similar method that was used in Chapter 4, the joint and limb kinematics were solved using a computational sub-model to simulate the biological spindle as observed in mammalian muscles, namely that of the cat [121, 122], which has also been used in human simulations [163, 103]. Action potentials in pulses per second (pps) were generated for primary (Ia) and secondary (II) afferents based on the interactions of the intrafusal fibers (chain, bag1, bag2). Fusimotor activation and the property changes in induces within the spindle model is represented by contractile elements (CE). The spindle model operates from a set of parameterized inputs that included that included L_o as optimal muscle lengths, L_{ce} as muscle length normalized to L_o, V_{ce} as the rate of change in muscle length (i.e., velocities), A_{ce} as muscle length acceleration, Fs as sampling frequency, $\gamma_{dynamic}$ as dynamic gamma drive, and γ_{static} as static gamma drive.

The model produced only two outputs, which were non-linear firings of the primary afferent potential and secondary afferent potential modalities in the spindle, Ia and II respectively. As stated in [121], the generation of afferent potential reflects the stretch of the intrafusal fiber model's sensory zone. Afferent potential primary derived based on Eq. 5.9 where T/K^{SR} is the calculated stretch in the sensory region of each intrafusal fiber, L_N^{SR} is the sensory region threshold length, L_0^{SR} is the sensory region rest length, and G is a constant that indicates the numerical relationship between intrafusal fiber's sensory region to primary afferent firing. Afferent potential secondary derived based on Eq. 5.10 where X is the percentage of the secondary afferent located on the sensory region and $L_{secondary}$ is the secondary rest length.

$$AfferentPotential_{Ia} = G \times \left[\frac{T}{K^{SR}} - (L_N^{SR} - L_0^{SR})\right] \tag{5.9}$$

$$\begin{aligned} AfferentPotential_{II} = G \times \Bigg\{ & X \times \frac{L_{secondary}}{L_0^{SR}} \times \left[\frac{T}{K^{SR}} - (L_N^{SR} - L_N^{0})\right] \\ & +(1-X) \times \frac{L_{secondary}}{L_0^{PR}} \times \left(L - \frac{T}{K^{SR}} - L_0^{SR} - L_N^{PR}\right) \Bigg\} \end{aligned} \tag{5.10}$$

Both of the afferent firing model's output firings were collected as raw data to be to be statistically analyzed for useful features that would indicated the current state of the limb.

5.3.4 Comparison of Inter-class Trajectory Context

In order to evaluate the discriminability of afferent signals against task-actions, the trajectory types must be compared extensively. The five trajectories selected for inspection are cycles of shapes and curvatures that aren't typically associated with the natural gait of a feline hind limb: Circle, Line, Oscillatory, Lemniscate, and Square. For this reason, there is an increased likelihood for indisputably discern variations despite noise that may be present with a data set's dimensionality, resolution, and sparsity. In our initial simulation executions, we observed that sensory afferent outputs of the muscles varied significantly depending on the initial conditions and the direction the limb moves in to complete the cycle. Therefore, we ensured that the simulated limb traversed each of the trajectories in two opposite directions: Reverse (REV) and Forward (FWD). For example the Circle-FWD, which indicates the limb traversed the Circle trajectory moving in the Forward direction, was compared in series to Circle-REV, Line-FWD, Line-REV, Oscillatory-FWD, Oscillatory-REV, Lemniscate-FWD, Lemniscate-REV, Square-FWD, and Square-REV. All possible combinations of trajectory comparisons totaled to 45 correlation pairs in both the raw data set and pre-processed (i.e., PCA) data set. The combination set did not include pairs that evaluated a trajectory-direction against each other.

5.3.5 Spatial, Spatio-Temporal, Pre-processing of Muscle Spindle Afferent Data

To test the presence of discriminability across tasks, the afferent data sets were evaluated within 3 pattern constraints: spatial, spatio-temporal, and pre-processing from dimensional reduction.

Spatial Analysis

We first evaluated the spatial patterns using the K-means++ algorithm. Since the standard K-means algorithm does not guarantee to find the optimum, an alternative, K-means++ chooses initial centers on a justifiable upper bound within cluster sum of squares objective. The approach is initiated by separating the k initial cluster centers, spatially.

Overall the formal objective is to determine:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \left\| \mathbf{x} - \boldsymbol{\mu}_i \right\|^2 = \underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i \tag{5.11}$$

where μ_i is the mean of points in S_i. This may also be shown to be equivalent to minimization of the squared deviations of points, as shown by:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \frac{1}{2|S_i|} \sum_{\mathbf{x},\mathbf{y} \in S_i} \left\| \mathbf{x} - \mathbf{y} \right\|^2 \tag{5.12}$$

For an initial set of k means $m_1^{(1)}, ..., m_k^{(1)}$, the algorithm proceeds by alternating between the assignment step and an update step, until convergence.

Spatio-Temporal Analysis

A useful statistical measure to use that identifies significant correlations among multiple trajectories with spatial and temporal components is cross correlation. It compares the time-series of afferent data across tasks, and is represented as the ratio in Eq. (5.13), where n is the total number of data point indices recorded per task cycle. This is suitable for measuring well two variables move in relation to each other. Both x_i and y_i are the individual spindle afferent sets, Ia and II, respectively. A temporal shift delay, phase lag τ, of the output cross correlation, R_{xy}, measure is applied to determine where the correlation of the data is maximized, as shown in Eq. (5.14).

$$R_{xy}(\tau) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{5.13}$$

$$\tau_{estimated} = \underset{\tau \in \mathbb{R}}{\arg\max}(R_{xy}(\tau)) \tag{5.14}$$

To retrieve the correlation coefficients, local sums can be calculated in an alternative way to normalize the cross-correlation. Using normalized cross correlation follows a general procedure by [108, 73] in Eq. (5.15).

$$\gamma(u,v) = \frac{\sum_{x,y}(f(x,y) - \overline{f}_{u,v})(t(x-u, y-v) - \overline{t})}{\sqrt{\sum_{x,y}[f(x,y) - \overline{f}_{u,v}]^2 \sum_{x,y}[t(x-u,y-v) - \overline{t}]^2}} \tag{5.15}$$

We can treat the combined group of muscle modalities within the afferent data as a template and image and calculating the cross-correlation in the spatial or the frequency domain. The implementation closely follows the formula from [108], where f is the image, $\bar{t}$ is the mean of the template, and $\overline{f}_{u,v}$ is the mean of $f(x,y)$ in the region under the template.

Data Pre-processing Analysis

Principal component analysis (PCA) is often used for dimensional reduction on multi-dimensional data, which assists with visual interpretation. PCA can identify the principal components that are able to distinguish the Ia and II modalities and which represent the most variations between groups. However, PCA is not particularly useful in accurately defining clear boundaries between different clusters in the data. The combined use of PCA with clustering methods helps us understand the cluster size and distribution of the spike trains associated with each task.

5.4 Results

5.4.1 Raw Multi-Dimensional Afferents Are Value Bound, But State-Indiscriminable

We first evaluated the task-dependency of the raw spindle afferent distribution. By plotting the averaged 200-point afferent distribution of all muscles during each of the five trajectory cycles, we can observe the spatial relationship among the Ia and II modalities. Fig. 5.2A displays the full comprehensive view of the five 8-dimensional sets of the spindle afferents for the average of the trajectories in the Forward and Reverse directions. For example, the mean of the resulting afferents were calculated between the time series of the Circle trajectory in the Forward and Reverse direction to obtain a single representation for that specific task. For both the Forward and Reverse plots, there were no spatially discernible clusters that could indicate a state association. However, there was a recognizable afferent separation of the Iliopsoas group muscle from the remaining group. This gave an indication that the sensory sets were likely to have a value boundedness characteristic, or having finite limits based on the musculotendon stretch that occurs throughout the designated type cycle in the performance of a task. K-means++ clustering analysis was predictably unable to adequately differentiate one trajectory from another. Although the data

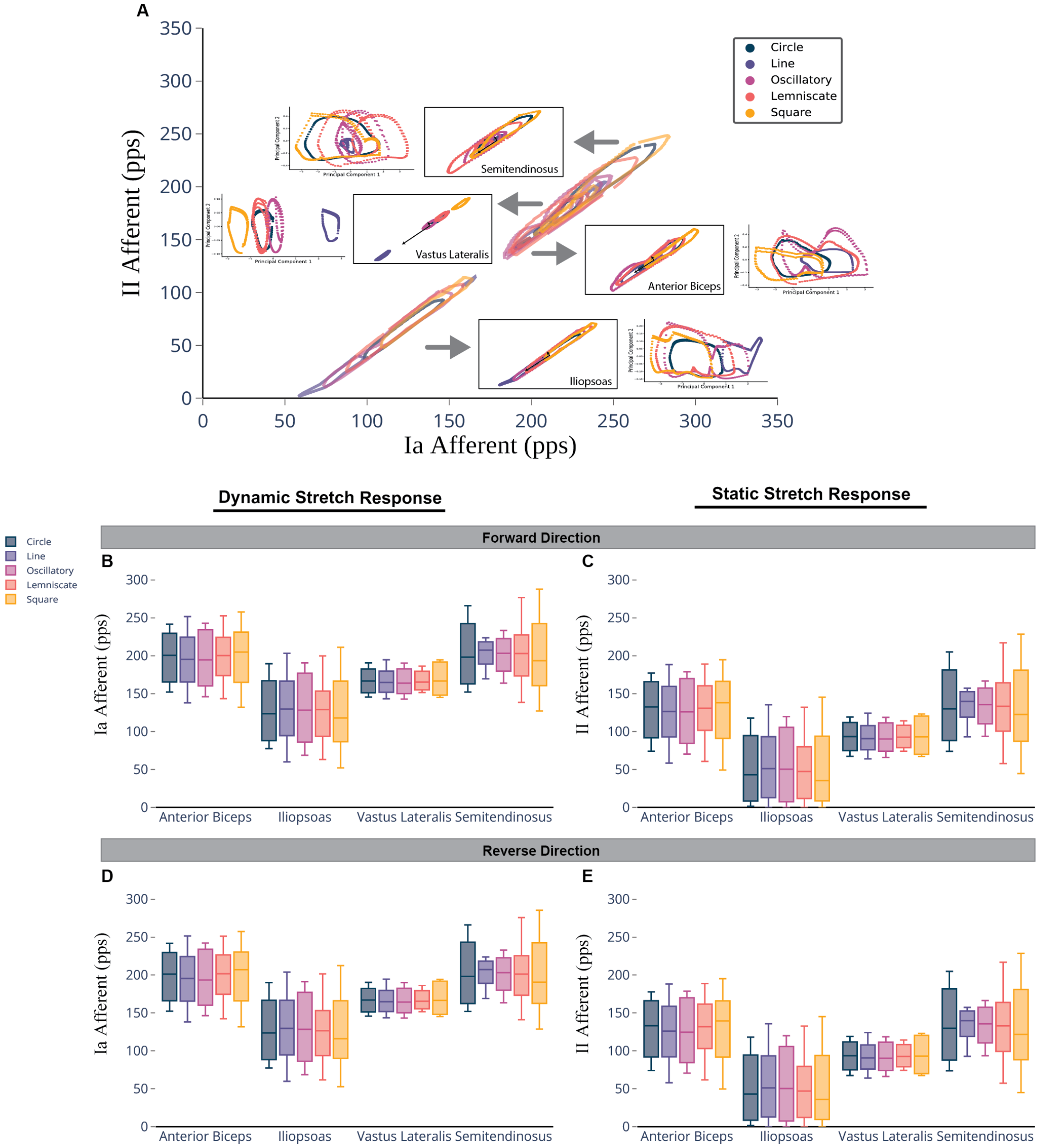


Figure 5.2: **Spindle afferent population data for five distinct trajectories.** (**A**) Full comprehensive view of all sensory signals. The projections for principal components 1 and 2 for each muscle are guided by the gray arrows. Overall patterns of dynamic stretch response for the Ia modality, are shown the left side, with (**B**) and (**D**). Overall patterns of the static stretch response for the II modality.

contained equal-volume clusters without outliers to represent each task, the spherical attribute and overlapping cluster radii of the data were further evidence that K-means was undesirable for spatial classification. Assuming the number of clusters K is initially unknown, the estimated value for the number of clusters is $K = 3$ was a grossly underestimate of the true number of clusters $K = 10$ (i.e., 5 trajectories in 2 directions). Since K-means clusters data points purely based on their geometric closeness of Euclidean distance to the assigned cluster centroid, this analysis fails at determining state discriminability among the raw afferent set. However, we observed the data structure had unique boundaries of target values that were respective of the muscle group.

Across all trajectories the spatial distribution in the dynamic stretch response (Ia) for the Forward and Reverse trajectories is shown in Figures 5.2B-E. Across both modalities, the Vastus Lateralis spanned the minimal range of afferent spikes in contrast to the other 3 muscles, lending to the expectation that the Vastus Lateralis may produce less accuracy overall for task classification in the raw data set. The maximum, median, and minimum values for each trajectory were averaged together in their respective muscle group to get a sense of the variations of finite space associated with that particular muscle. Bounded ranges for the dynamic stretch response (Ia), Fig. 5.2B and Fig. 5.2D, are explicitly listed in Table 5.2.

Table 5.2: Bounded ranges of Ia afferent activity, measured in pulses per second (pps), for muscle groups averaged across Forward and Reverse directions.

Muscle Name	**Maximum** (p)	>	**Median**	>	**Minimum** (q)
Anterior Biceps	249.07	>	199.4	>	142.22
Iliopsoas	199.33	>	125.22	>	64.15
Vastus Lateralis	191.18	>	165.5	>	145.77
Semitendinosus	257.17	>	200.55	>	150.50

Following is the static stretch responses (II), where the spatial distribution is shown in Fig. 5.2C and Fig. 5.2E, are listed in Table 5.3.

Table 5.3: Bounded ranges of II afferent activity, measured in pulses per second (pps), for muscle groups averaged across Forward and Reverse directions.

Muscle Name	**Maximum** (p)	$>$	**Median**	$>$	**Minimum** (q)
Anterior Biceps	185.66	$>$	130.99	$>$	62.55
Iliopsoas	130.16	$>$	45.38	$>$	.28
Vastus Lateralis	119.855	$>$	92.01	$>$	67.73
Semitendinosus	194.83	$>$	132.08	$>$	72.53

5.4.2 Pre-processing Suggests Observable Correlations in Sensory and Motor Maps

In order to improve our analysis beyond the limitations of K-means, we invoked techniques for dimensional reduction on the 8-D high dimensional space to a low-dimensional representation, which we assumed would retain some meaningful properties of the original raw afferent data. Interestingly, not only did we find strong groupings in the pre-processed data sets but there were observable correlations that exist within the spatiotemporal dynamics of sensorimotor space. In Fig. 5.3A, the top 3 principal components are plotted the five trajectories. The projections overlap each other significantly and are tightly clustered along the same plane. Most of the explained variance is captured in the first 2 principal components. PC1 captures the most variation at 70.22%, PC2 follows with 28.81%, and PC3 captures 0.58%. Fig. 5.3B shows the breakdown for each component with their individual and then the overall cumulative values. When the principal components for each individual trajectory were plotted separate from one another, we were able to perceive discernible shapes that weren't visible, but possibly obscured, in the raw data set. In Figs. 5.3C-G, PC1 and PC2 revealed projections that closely resembled the prescribed trajectories and task in the joint kinematic space. Fig. 5.3C, associated with the Line trajectory, reveals a non-straight line with slight curvature. Fig. 5.3E captures the full revolution of the Oscillatory task. One half of the task's revolution does not completely trace over the other half, unlike the Line, but overall afferent response still reveals the sinusoidal shape. The Square trajectory roughly resembles the the planned trajectory, except the sides aren't quite equilateral and roughly resembles a parallelogram. Some distortion is acceptable here and not indicative of any errors in the reduction of data. In fact, the results of near-identifiable shapes emerging from the principal components were surprising and not expected, considering the raw data presented clusters oval-like shapes.

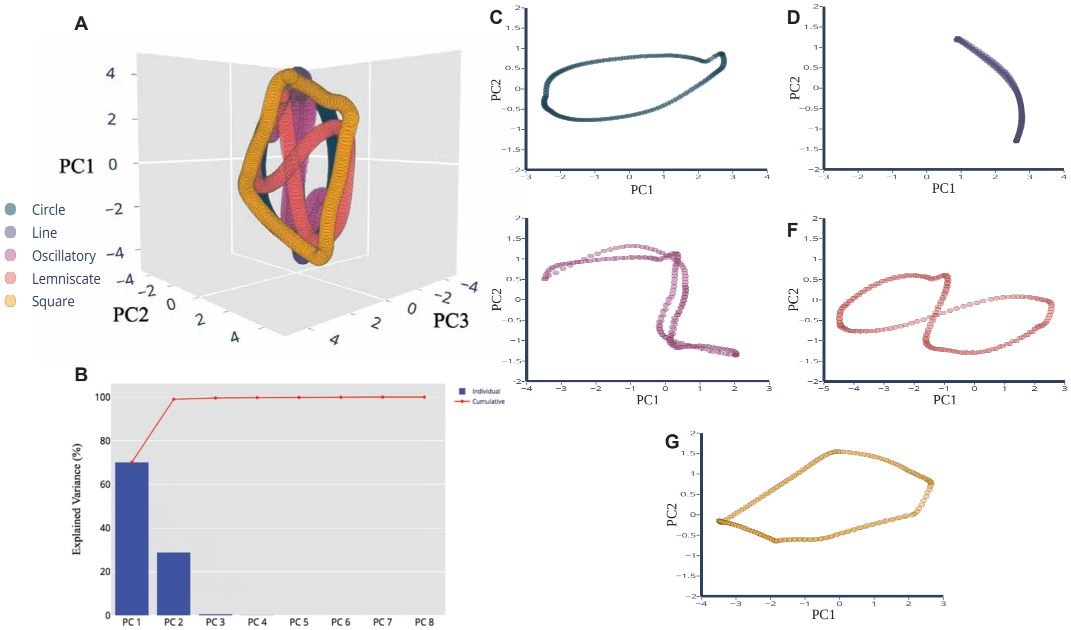


Figure 5.3: **PCA Dimensions of Afferents Reveal Distinct Shapes.** Principal component plots of pre-processed data revealed shapes that are quantitatively correlated to the planned trajectory cycles in kinematic space. (**A**) A three-dimensional PCA plot shows the cluster of samples based on their similarity, revealing distinctive shapes in space. (**B**) PCA scree plot of the variance explained by each of the 8 individual principal components are shown here in blue, with cumulative percentages show in red. The first 3 PCs explain 99.61% of the variance. The two-dimensional plots of the (**C**) Circle, (**D**) Line, (**E**) Oscillatory, (**F**) Lemniscate, and (**G**) Square shapes show more distinction in visual appearance of the trajectory when the PC1 and PC2 variables were plotted together.

Since the pre-processed data was able to be visualized with 2-3 principal components there was no need to consider other dimension reduction techniques such as T-distributed Stochastic Neighbor Embedding (t-SNE) and multidimensional scaling (MDS). Two or three principal components are usually sufficient for our plotting purposes whereas for classification or modeling purposes, the number of significant components was can be properly determined using metrics such the explained variance. Here, we were able to conclude that there is a presence of near-approximate quantitative correlations of joint kinematic and sensory space of the muscle spindle. The next experimental findings further use these top three components to determine their usability for state classification.

5.4.3 Correlation Index Reveals Markers of Action Discriminability, Classification

Before pre-processing the afferent manifolds to detect useful features, cross correlation was performed on the raw data set to retrieve the correlation coefficient or index value that measures similarity in movements of two time-series sets of data relative to each other. To our dissatisfaction, cross-correlation analysis, as computed from Eq. (5.13), did not provide sufficient discriminability among the five states when comparing the raw spindle manifolds. A positive 50%, the measure of chance, was set as the threshold for verifying discriminability among the span of possible cross correlation values where the value -1 indicates a perfect negative correlation, +1 indicates the perfect positive correlation, and 0 is no correlation between the paired tasks. Essentially, $R_{xy}(\tau) \geq 0.5$ indicates less discriminability among the tasks and $R_{xy}(\tau) < 0.5$ indicates more discriminability. Assessments for cross correlation were divided into 5 sensory afferent groups: combined muscles set (all four muscles combined), Anterior Biceps, Iliopsoas, Vastus Lateralis, and Semitendinosus. For each of the $n = 45$ possible trajectory combinations and pairwise comparisons we plotted their correlation coefficients, $R_{xy}(\tau)$, spatial scatter visualization for both the raw afferents and pre-processed afferents as shown in in Fig. 5.4. All raw data correlations for the combined muscle were set at the 8-D (i.e., 4 muscles x 2 afferents) high dimensional space while the individual muscles were compared in 2-D space. All pre-processed data correlations for the combined muscle were set at the 3-D space while the individual muscles were compared in 2-D space.

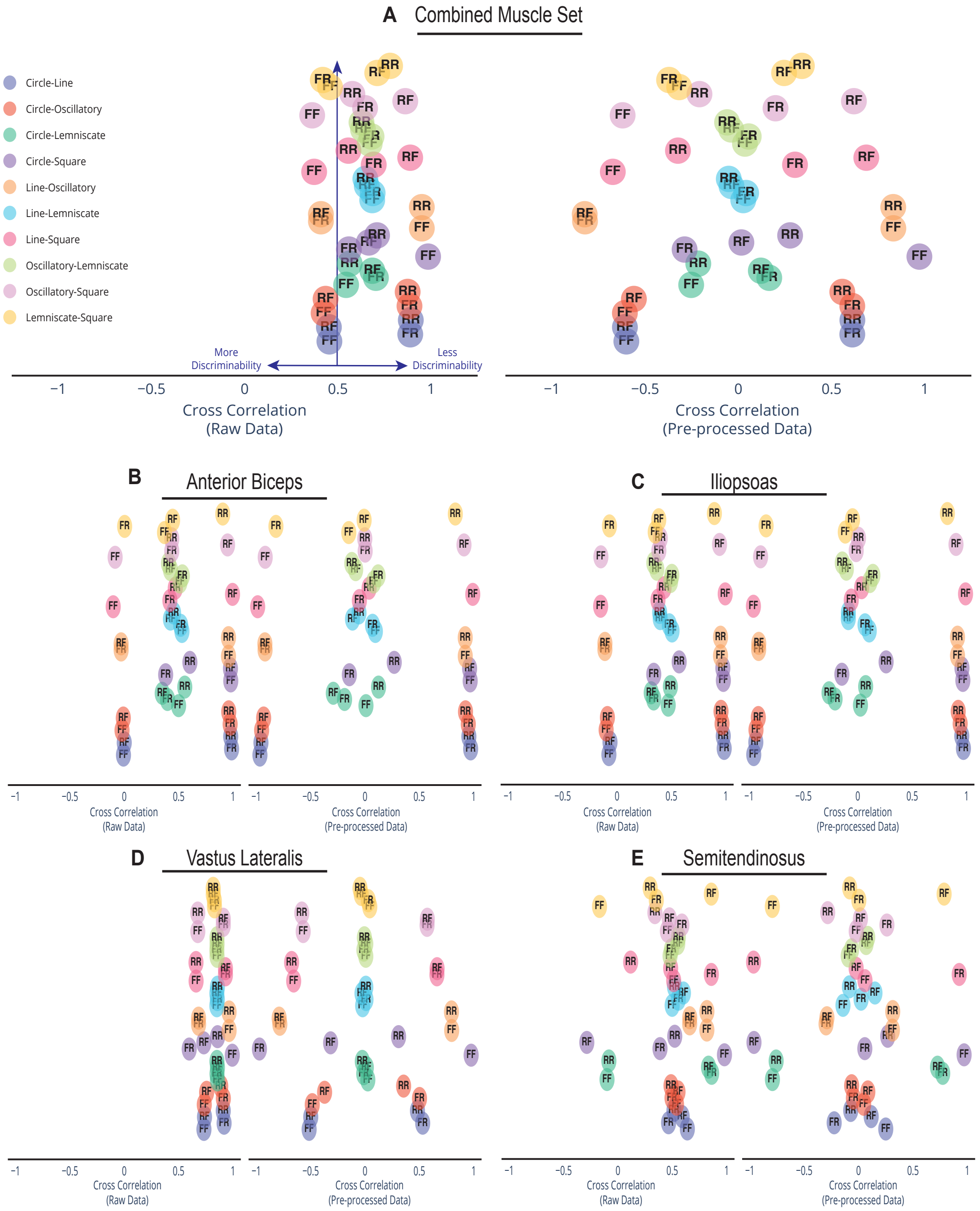
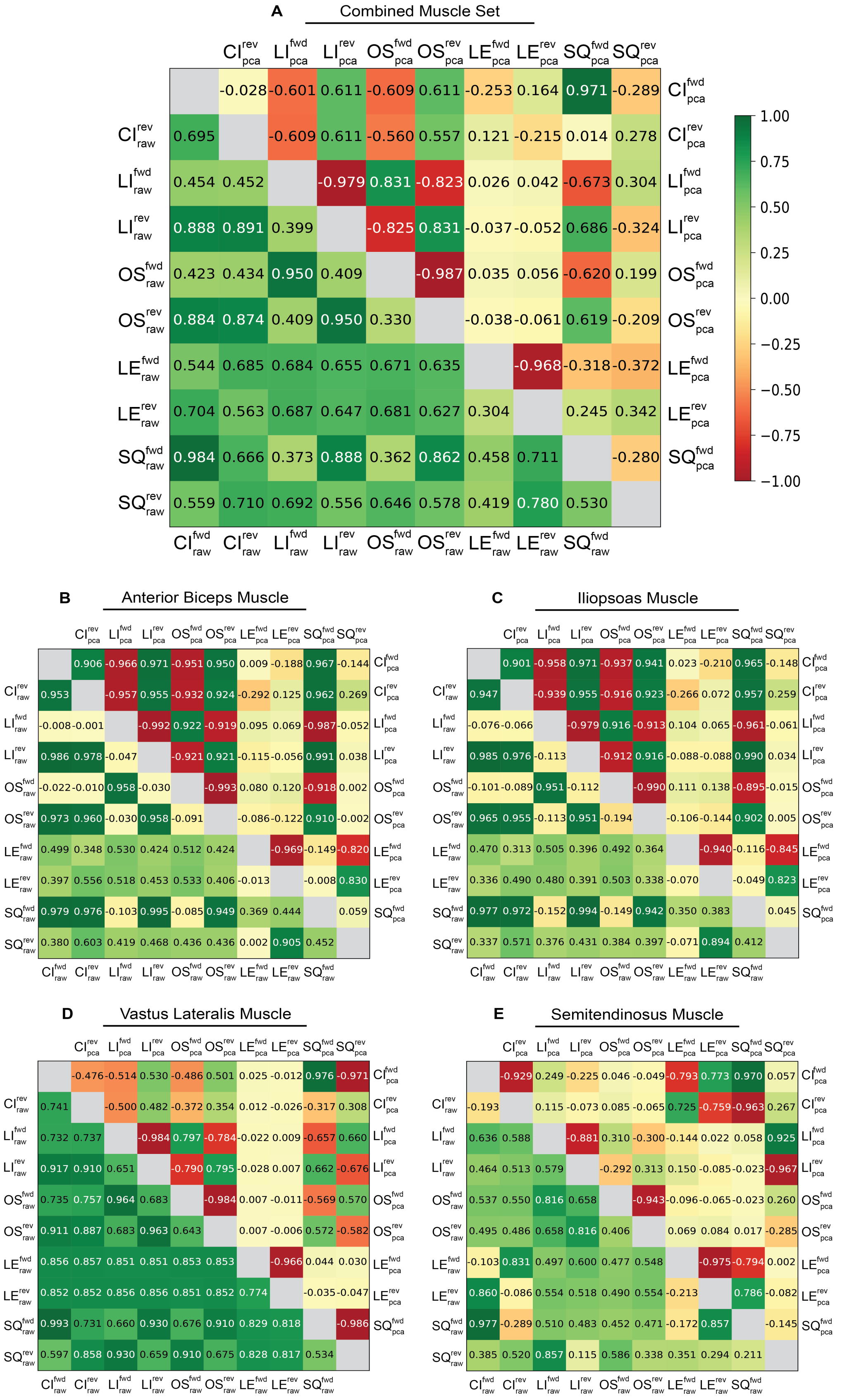


Figure 5.4: **Spread of Discriminability Within Cross Correlation Scatter.** Cross Correlations of all possible trajectory combinations ($n = 45$) for the raw information and pre-processed afferent signal information, which are plotted on the left and right, respectively. Each trajectory pairing has an assigned direction. An R label is for Reverse Direction and an F label is for Forward Direction. For example, the Circle-Line pairing with label FF indicates Circle going Forward and Line going Forward. The Line-Square pairing of RF indicates Line going in Reverse and Square going Forward. Cross-correlations were plotted for (**A**) the combined set of the 4 muscles in the cat limb, (**B**) Anterior Biceps, (**C**) Iliopsoas, (**D**) Vastus Lateralis, and (**E**) Semitendinosus.

Let α be the span or bandwidth of the detected correlations in the cluster and β be the total space of possible correlation values, where α/β is the percentage covered by the correlations values. Raw data correlations in the combined muscle set (Fig. 5.4A) show a tight cluster within a 32.1% (i.e., where α is .646 and β is 2) of the full correlation range. However, that range expands to 98% (1.958/2) in the pre-processed set as more pairings move away from being less discriminable to more discriminable. This form of expansion was not only evident in the combined muscle grouping but also in the individual muscle groups (Figs. 5.4B-E). Out of the four muscles, the Vastus Lateralis (Fig. 5.4D) contained the most compact clustering in the raw set with the maximal expansion, spanning a minimal 23% (.459/2) and expanding to 98% (1.96/2) in the pre-processed set. We highlight the compact-to-expansion dynamic that occurs from raw to pre-processed afferents to show the usefulness of pre-processing in giving each task more distinction and separability to enhance classification. Furthermore, the usefulness of cross-correlation is additionally investigated in this study in the context of a state classifier. We find the display of confusion matrices as heat maps particularly useful here because of the ability to describe the performance of our classification model. You can observe any patterns in value for one or both variables by observing change in color gradients of cell colors change across each axis in Fig. 5.5. The differences in discriminability vary significantly by each matrix. Our combined muscle set reports 29% discriminability among the 10 possible trajectories in the raw 8-D set and drastically increases to 82% in reduced 3-D pre-proccessed set, as shown in Fig. 5.5A. For the Anterior Biceps muscle (Fig. 5.5B), cross correlation reports 60% discriminability in the raw 2-D set and increases to 73% in 2-D pre-proccessed set. The Iliopsoas muscle (Fig. 5.5C) reports 66% and 73% discriminabilities, Vastus Lateralis (Fig. 5.5D) reports 0% and 78% discriminabilities, and Semitendinosus (Fig. 5.5E) reports 49% and 89% discriminabilities, for the raw and pre-proccessed set, respectively. The difference in correlation between the two groups (i.e., Raw and PCA) was determined to statistically significant ($p = 0.001$), according to the Wilcoxon signed-rank test, for the combined muscle group and individual muscle sets.



A Combined Muscle Set

		CI^{rev}_{pca}	LI^{fwd}_{pca}	LI^{rev}_{pca}	OS^{fwd}_{pca}	OS^{rev}_{pca}	LE^{fwd}_{pca}	LE^{rev}_{pca}	SQ^{fwd}_{pca}	SQ^{rev}_{pca}	
		-0.028	-0.601	0.611	-0.609	0.611	-0.253	0.164	0.971	-0.289	CI^{fwd}_{pca}
CI^{rev}_{raw}	0.695		-0.609	0.611	-0.560	0.557	0.121	-0.215	0.014	0.278	CI^{rev}_{pca}
LI^{fwd}_{raw}	0.454	0.452		-0.979	0.831	-0.823	0.026	0.042	-0.673	0.304	LI^{fwd}_{pca}
LI^{rev}_{raw}	0.888	0.891	0.399		-0.825	0.831	-0.037	-0.052	0.686	-0.324	LI^{rev}_{pca}
OS^{fwd}_{raw}	0.423	0.434	0.950	0.409		-0.987	0.035	0.056	-0.620	0.199	OS^{fwd}_{pca}
OS^{rev}_{raw}	0.884	0.874	0.409	0.950	0.330		-0.038	-0.061	0.619	-0.209	OS^{rev}_{pca}
LE^{fwd}_{raw}	0.544	0.685	0.684	0.655	0.671	0.635		-0.968	-0.318	-0.372	LE^{fwd}_{pca}
LE^{rev}_{raw}	0.704	0.563	0.687	0.647	0.681	0.627	0.304		0.245	0.342	LE^{rev}_{pca}
SQ^{fwd}_{raw}	0.984	0.666	0.373	0.888	0.362	0.862	0.458	0.711		-0.280	SQ^{fwd}_{pca}
SQ^{rev}_{raw}	0.559	0.710	0.692	0.556	0.646	0.578	0.419	0.780	0.530		
	CI^{fwd}_{raw}	CI^{rev}_{raw}	LI^{fwd}_{raw}	LI^{rev}_{raw}	OS^{fwd}_{raw}	OS^{rev}_{raw}	LE^{fwd}_{raw}	LE^{rev}_{raw}	SQ^{fwd}_{raw}		

B Anterior Biceps Muscle

		CI^{rev}_{pca}	LI^{fwd}_{pca}	LI^{rev}_{pca}	OS^{fwd}_{pca}	OS^{rev}_{pca}	LE^{fwd}_{pca}	LE^{rev}_{pca}	SQ^{fwd}_{pca}	SQ^{rev}_{pca}	
		0.906	-0.966	0.971	-0.951	0.950	0.009	-0.188	0.967	-0.144	CI^{fwd}_{pca}
CI^{rev}_{raw}	0.953		-0.957	0.955	-0.932	0.924	-0.292	0.125	0.962	0.269	CI^{rev}_{pca}
LI^{fwd}_{raw}	-0.008	-0.001		-0.992	0.922	-0.919	0.095	0.069	-0.987	-0.052	LI^{fwd}_{pca}
LI^{rev}_{raw}	0.986	0.978	-0.047		-0.921	0.921	-0.115	-0.056	0.991	0.038	LI^{rev}_{pca}
OS^{fwd}_{raw}	-0.022	-0.010	0.958	-0.030		-0.993	0.080	0.120	-0.918	0.002	OS^{fwd}_{pca}
OS^{rev}_{raw}	0.973	0.960	-0.030	0.958	-0.091		-0.086	-0.122	0.910	-0.002	OS^{rev}_{pca}
LE^{fwd}_{raw}	0.499	0.348	0.530	0.424	0.512	0.424		-0.969	-0.149	-0.820	LE^{fwd}_{pca}
LE^{rev}_{raw}	0.397	0.556	0.518	0.453	0.533	0.406	-0.013		-0.008	0.830	LE^{rev}_{pca}
SQ^{fwd}_{raw}	0.979	0.976	-0.103	0.995	-0.085	0.949	0.369	0.444		0.059	SQ^{fwd}_{pca}
SQ^{rev}_{raw}	0.380	0.603	0.419	0.468	0.436	0.436	0.002	0.905	0.452		
	CI^{fwd}_{raw}	CI^{rev}_{raw}	LI^{fwd}_{raw}	LI^{rev}_{raw}	OS^{fwd}_{raw}	OS^{rev}_{raw}	LE^{fwd}_{raw}	LE^{rev}_{raw}	SQ^{fwd}_{raw}		

C Iliopsoas Muscle

		CI^{rev}_{pca}	LI^{fwd}_{pca}	LI^{rev}_{pca}	OS^{fwd}_{pca}	OS^{rev}_{pca}	LE^{fwd}_{pca}	LE^{rev}_{pca}	SQ^{fwd}_{pca}	SQ^{rev}_{pca}	
		0.901	-0.958	0.971	-0.937	0.941	0.023	-0.210	0.965	-0.148	CI^{fwd}_{pca}
CI^{rev}_{raw}	0.947		-0.939	0.955	-0.916	0.923	-0.266	0.072	0.957	0.259	CI^{rev}_{pca}
LI^{fwd}_{raw}	-0.076	-0.066		-0.979	0.916	-0.913	0.104	0.065	-0.961	-0.061	LI^{fwd}_{pca}
LI^{rev}_{raw}	0.985	0.976	-0.113		-0.912	0.916	-0.088	-0.088	0.990	0.034	LI^{rev}_{pca}
OS^{fwd}_{raw}	-0.101	-0.089	0.951	-0.112		-0.990	0.111	0.138	-0.895	-0.015	OS^{fwd}_{pca}
OS^{rev}_{raw}	0.965	0.955	-0.113	0.951	-0.194		-0.106	-0.144	0.902	0.005	OS^{rev}_{pca}
LE^{fwd}_{raw}	0.470	0.313	0.505	0.396	0.492	0.364		-0.940	-0.116	-0.845	LE^{fwd}_{pca}
LE^{rev}_{raw}	0.336	0.490	0.480	0.391	0.503	0.338	-0.070		-0.049	0.823	LE^{rev}_{pca}
SQ^{fwd}_{raw}	0.977	0.972	-0.152	0.994	-0.149	0.942	0.350	0.383		0.045	SQ^{fwd}_{pca}
SQ^{rev}_{raw}	0.337	0.571	0.376	0.431	0.384	0.397	-0.071	0.894	0.412		
	CI^{fwd}_{raw}	CI^{rev}_{raw}	LI^{fwd}_{raw}	LI^{rev}_{raw}	OS^{fwd}_{raw}	OS^{rev}_{raw}	LE^{fwd}_{raw}	LE^{rev}_{raw}	SQ^{fwd}_{raw}		

D Vastus Lateralis Muscle

		CI^{rev}_{pca}	LI^{fwd}_{pca}	LI^{rev}_{pca}	OS^{fwd}_{pca}	OS^{rev}_{pca}	LE^{fwd}_{pca}	LE^{rev}_{pca}	SQ^{fwd}_{pca}	SQ^{rev}_{pca}	
		-0.476	-0.514	0.530	-0.486	0.501	0.025	-0.012	0.976	-0.971	CI^{fwd}_{pca}
CI^{rev}_{raw}	0.741		-0.500	0.482	-0.372	0.354	0.012	-0.026	-0.317	0.308	CI^{rev}_{pca}
LI^{fwd}_{raw}	0.732	0.737		-0.984	0.797	-0.784	-0.022	0.009	-0.657	0.660	LI^{fwd}_{pca}
LI^{rev}_{raw}	0.917	0.910	0.651		-0.790	0.795	-0.028	0.007	0.662	-0.676	LI^{rev}_{pca}
OS^{fwd}_{raw}	0.735	0.757	0.964	0.683		-0.984	0.007	-0.011	-0.569	0.570	OS^{fwd}_{pca}
OS^{rev}_{raw}	0.911	0.887	0.683	0.963	0.643		0.007	-0.006	0.572	-0.582	OS^{rev}_{pca}
LE^{fwd}_{raw}	0.856	0.857	0.851	0.851	0.853	0.853		-0.966	0.044	0.030	LE^{fwd}_{pca}
LE^{rev}_{raw}	0.852	0.852	0.856	0.856	0.851	0.852	0.774		-0.035	-0.047	LE^{rev}_{pca}
SQ^{fwd}_{raw}	0.993	0.731	0.660	0.930	0.676	0.910	0.829	0.818		-0.986	SQ^{fwd}_{pca}
SQ^{rev}_{raw}	0.597	0.858	0.930	0.659	0.910	0.675	0.828	0.817	0.534		
	CI^{fwd}_{raw}	CI^{rev}_{raw}	LI^{fwd}_{raw}	LI^{rev}_{raw}	OS^{fwd}_{raw}	OS^{rev}_{raw}	LE^{fwd}_{raw}	LE^{rev}_{raw}	SQ^{fwd}_{raw}		

E Semitendinosus Muscle

		CI^{rev}_{pca}	LI^{fwd}_{pca}	LI^{rev}_{pca}	OS^{fwd}_{pca}	OS^{rev}_{pca}	LE^{fwd}_{pca}	LE^{rev}_{pca}	SQ^{fwd}_{pca}	SQ^{rev}_{pca}	
		-0.929	0.249	-0.225	0.046	-0.049	-0.793	0.773	0.970	0.057	CI^{fwd}_{pca}
CI^{rev}_{raw}	-0.193		0.115	-0.073	0.085	-0.065	0.725	-0.759	-0.963	0.267	CI^{rev}_{pca}
LI^{fwd}_{raw}	0.636	0.588		-0.881	0.310	-0.300	-0.144	0.022	0.058	0.925	LI^{fwd}_{pca}
LI^{rev}_{raw}	0.464	0.513	0.579		-0.292	0.313	0.150	-0.085	-0.023	-0.967	LI^{rev}_{pca}
OS^{fwd}_{raw}	0.537	0.550	0.816	0.658		-0.943	-0.096	-0.065	-0.023	0.260	OS^{fwd}_{pca}
OS^{rev}_{raw}	0.495	0.486	0.658	0.816	0.406		0.069	0.084	0.017	-0.285	OS^{rev}_{pca}
LE^{fwd}_{raw}	-0.103	0.831	0.497	0.600	0.477	0.548		-0.975	-0.794	0.002	LE^{fwd}_{pca}
LE^{rev}_{raw}	0.860	-0.086	0.554	0.518	0.490	0.554	-0.213		0.786	-0.082	LE^{rev}_{pca}
SQ^{fwd}_{raw}	0.977	-0.289	0.510	0.483	0.452	0.471	-0.172	0.857		-0.145	SQ^{fwd}_{pca}
SQ^{rev}_{raw}	0.385	0.520	0.857	0.115	0.586	0.338	0.351	0.294	0.211		
	CI^{fwd}_{raw}	CI^{rev}_{raw}	LI^{fwd}_{raw}	LI^{rev}_{raw}	OS^{fwd}_{raw}	OS^{rev}_{raw}	LE^{fwd}_{raw}	LE^{rev}_{raw}	SQ^{fwd}_{raw}		

Figure 5.5: **Confusion matrices of raw and pre-processed spindle afferent data.** For (**A**) the combined set of muscles and spindle modalities (four tendons as shown in Fig. 5.1), cross correlation analysis was performed on the raw afferents across all five trajectories; CI: circle, LI: line, OS: oscillatory, LE: lemniscate, SQ: square. The direction of each task is label either *fwd* for forward and *rev* for reverse. The lower half of matrix entries correspond to the *raw* collection of pps signals. Upper half entries represent the resulting principle components from PCA, labeled as *pca*. Three principal components were selected here. While in individual muscle analysis such as (**B**) Anterior Biceps, two principal components were selected. The same cross correlation was performed for the remaining muscles which include (**C**) Iliopsoas, (**D**) Vastus Lateralis, and (**E**) Semitendinosus.

5.5 Discussion

In this chapter, we focused on using high-dimensional proprioceptive signals from muscle spindles to enable robust estimates of trajectories as pre-defined tasks and bodily states. Before our trials and analysis, it remained unknown whether spindle signals suffice to discriminate limb movements. We obtained three main findings that we discovered in the post analyses. The first is identifying that raw multi-dimensional sensory sets of the muscle spindle are value bound, but are still able to maintain state-indiscriminability. Secondly, pre-processed data shows high correlation of spatio-temporal maps between sensory and motor space. Thirdly, the correlation index revealed markers of sufficient discriminability and classification among spindle afferents. Our findings closely match with similar results from Rongala et al. [152], where biological data on cuneate nucleus neuron recordings in adult cats were obtained and modeled to study generalizable tactile representations. Their work highlights that the cuneate nucleus forms the first interface for the sense of touch in the brain. We conjecture this would be similar for proprioceptive sensory afferent pathways. Triangle matrices of correlations, similar to our analysis, demonstrated how weighted learning in the cuneate nucleus resulted in decorrelated responses between neurons of the same stimulus. Essentially this means the data were less "confused" with another and more discriminable. Altogether, our findings indicate that sensory afferents from the muscle spindle can adequately supply the nervous system with features of discrimination to distinguish one task from another, only if there is suitable forms of pre-processing or filtering to reduce the overwhelming amount of sensory manifolds flooding the nervous at a given time during the performance of an action or task. This is a desirable and necessary result for our dynamic model of body representations or body schemas in neuro-robotics systems.

Chapter 6

Sensory-Motor Gestalt: Sensation and Action as the Foundations of Identity, Agency, and Self

"To understand is to perceive patterns." -Isaiah Berlin

6.1 Abstract

Body movement and proprioception are inextricably linked. Movement produces continuous high-dimensional ensembles of afferent information that provide an *internal proprioceptive body representation* and its relationship to the environment. Motor function is amenable to recording and interpretation and has been relatively well studied. However, we do not yet understand how physiological proprioceptive afferents contribute to internal body representations, neuromuscular control, and even a sense of agency and self. Proprioceptive and motor signals have often been seen as separate, and to be combined mainly to close feedback loops for neuromuscular control. In contrast, 'active sensing,' is an emergent concept for dynamically blending sensory and motor signals. We extend and formalize active sensing into an integrative approach——born out of a neuromechanical perspective—that sees proprioceptive and motor signals as integral parts of the same functional and perceptual continuum we call the *Sensory-Motor Gestalt.* The Sensory-Motor

Gestalt combines formalisms of physics, state estimation, biomechanics, differential geometry, and physiology to understand the emergence of the self in the context of proprioception and motor actions in the physical world. Proprioception, by defining *body state*, defines feasible (continuous or discrete) motor actions compatible with that state and the environment. Conversely, motor actions produce subsequent, often predictable, body states. This syntactical relationship leads to an epistemological continuum that spans body state, feasible behavior, agency, identity, and sense of self in organisms and robots.

6.2 Introduction of Sensory-Motor Gestalt: Origin and Definition

Our computational model of the self begins with *Gestalt Theory.* Gestalt (pronounced gə-'shtält), a concept originating in Austria and Germany, roughly translates to shape, form, configuration, and unified whole. XX-century German psychologist Max Wertheimer utilized this definition to originally present the Gestalt laws (or principles) of grouping for pictorially detailing how the human eye perceives visual elements [186]. These laws are fundamental rules illustrating how humans recognize elements and objects in their visual scene as organized patterns with meaning. The Gestalt theory of the mind and brain intends to form an understanding of how humans and animals 1) comprehend what they are perceiving and 2) obtain meaning from the world with disordered visual stimuli.

In its original formulation, Gestalt laws aim to reduce complex visual scenes into simpler, less complex shapes to can convey an image's meaning in a single formation instead of disparate smaller elements. Hence by being a critical aspect of the perception of patterns into a coherent whole for context and meaning, Gestalt plays an important role in combining epistemology (i.e., study of knowledge and how does one come to "know") and ontology (i.e., study of what is the nature of the self) [70]. This paper proposes underlying mechanisms for brain-body dynamics to merge proprioceptive and motor elements into an epistemological continuum from sensory and proprioceptive input, to state of the body, to feasible motor action, to useful behavior, to the sensory consequences of action— and then on to more abstract notions of agency, identity and sense of self in organisms and robots.

In this study, we focus on proprioception as spike trains from muscle spindles (II and Ia) and Golgi tendon organs (GTOs). They primarily encode muscle fiber length and velocity, and tendon tension, respectively. These proprioceptive signals are known to inform various perceptual modalities of body state (e.g., postures, movements, forces, limb stiffness, alertness). Recent exponential growth in literature relating proprioception to subjective experiences (Figure 6.1) may suggest that the debate about the emergence of the self is advancing. We hypothesize that Gestalt laws can be applied to organizing these physiologically-tenable proprioceptive signals to construct a *totality* of what is perceived as the active body. We seek to do the same for motor actions by developing a mathematical description of the set of plausible motor actions conditional on proprioceptive signals.

Our prior simulation work characterized the high-dimensional, non-linear, time-varying manifolds of muscle spindle afferents (Ia and II encoding, roughly, muscle fiber contraction velocities and lengths, respectively) that emerge during movement of a planar multi-muscle limb [8]. We provided initial evidence that high-dimensional muscle spindle proprioception defines limb states that reflect the consequences of motor actions. We now extend that work by emphasizing that each limb state, by its physical nature, only has a well-defined set of feasible motor actions. This results in a formal sensory-to-motor-to-sensory map that defines both the current proprioceptive states and feasible motor commands (i.e., plausible motor actions) that will lead to new (but expected) proprioceptive signals. The Sensory-Motor Gestalt applies to both biological and engineered agents where the

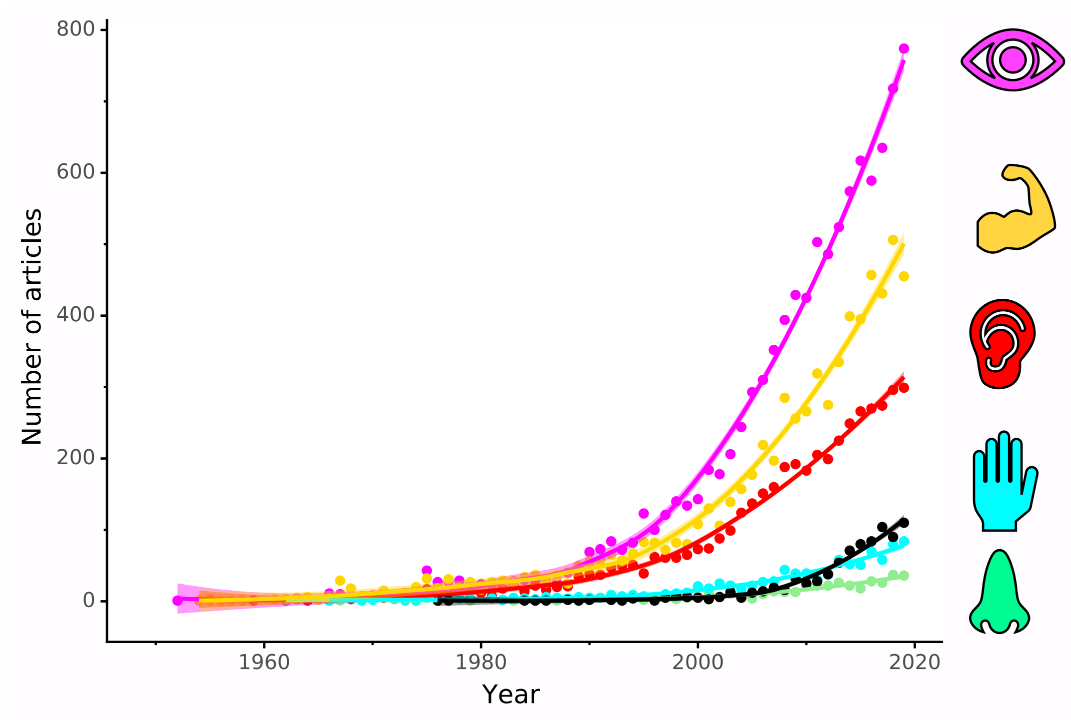


Figure 6.1: Published article count per year that indicate association between subjective experiences and sensory modalities. This chart similarly models a previous search conducted by [45], in which the number of articles published mentioned the words "awareness" or "consciousness" in conjunction with words denoting each sensory modality: "Visual" or "vision" (magenta), "proprioception" or "movement" (yellow), "auditory" or "audition" (red), "touch" or "tactile" (cyan), "olfaction" or "olfactory" (green), and "multisensory" (black). Our PubMed search extended the year range from 1950-2019 and added proprioception. Along with vision, proprioception showed a significant increase in documented work.

concepts of state, observability, and controllability are intimately related; therefore, providing a basis for constructing an artificial core of state, agency, identity and ultimately self.

To our knowledge, this is a first attempt to formally apply the laws of Gestalt to the encoding of the sense of agency, identity and self via proprioception. This article first builds the concept of Sensory-Motor Gestalt from the generic Gestalt theory. We then interpret the Sensory-Motor Gestalt in the form of mathematical encoding for each of the core laws, which may be integrated to form the sensorimotor self. How the self is directly related to sensorimotor experiences of neuromuscular systems is then explored while providing sample platform applications to support Sensory-Motor Gestalt, both biological and in bioinspired robots.

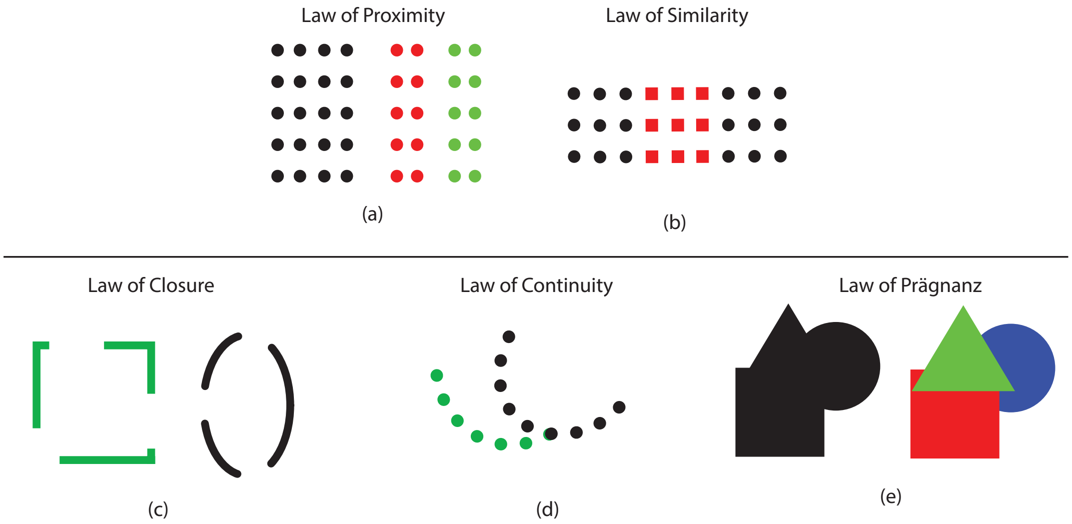


Figure 6.2: Gestalt laws of perceptual organization for topological manifold data can be applied using these core laws: (**a**) **Law of Proximity** aims to group elements together based on spatial closeness. Black dots, red dots, and green dots are perceived as separate groups due to the nearness of columns. (**b**) **Law of Similarity** groups the elements of black dots and red squares as separate sets, although spatial distance between each element is consistent. Shape, orientation, and color are the distinguishing factors. (**c**) **Law of Closure** prompts pattern perception of a green square and black oval despite the non-continuous outline and presence of gaps. (**d**) **Law of Continuity** perceives the figure as a green dotted line and a separate black dotted line due to the observed fluid connection of continuity and direction. (**e**) **Law of Prägnanz** (i.e., pithiness, conciseness, or Good Form) takes the abstract shape, as depicted on the left, and perceptually reorganizes them into a simple, more recognizable forms as is depicted on the right with the colored circle, triangle and square.

6.3 Sensory-Motor Gestalt: Applying Gestalt Laws to Sensorimotor Function

“The whole is greater than the sum of its parts” is the popular adage Gestalt psychology is best known for. It emphasizes the fact that although a sensory experience can be disassembled into individual components (i.e., stimuli), the way in which those components coalesce together generates properties and qualities of the whole that only exist independently of their components. Stimuli patterns presented as a whole often prompts a more meaningful perceptual response. As alluded to earlier, Gestalt theory is typically associated with the visual sense and visual perception (e.g., object and shape recognition, coloring, arrangement of parts) that is used to process graphic designs and images (Figure 6.2). Rarely has Gestalt theory been applied to other sensory modalities such as haptic [27], auditory, and olfactory senses which can all be topographically represented on a multi-dimensional space in the depiction of manifolds (i.e., coherent and continuous lower-dimensional subspaces embedded in a higher-dimensional space). Interestingly, the functional mechanisms of Gestalt laws are active in other cortical areas of the brain and not solely in the visual processing centers. As the brain's neural processing is responsible for stitching together the visual scene of the external world in the primary sensory cortices and also seamlessly binding raw multisensory information to project a single unified experience, we theorize there are benefits in extrapolating ideas of Gestalt laws of perceptual grouping from vision to other modalities of the body such as the somatosensory system. Gestalt theory typically consists of five core laws that govern the fundamental organization of perception: Laws of Proximity, Similarity, Closure, Continuity, and Prägnanz.

6.3.1 Law of Proximity

Within visual perception, objects in space or points on a plane that are near or proximate to each other have a tendency to be grouped together in a single unified set. Conversely, points that are further apart have a lesser likelihood to be viewed as conjoined (see Figure 6.2a). This law is useful for organizing information with increased speed and efficiency. There are several ways in which proprioceptive information can also be processed to yield proximity metrics. One of the earlier attempts to address the Law of Proximity is the Pure Distance model [99], which attempts to quantify visual proximity grouping in dot lattices with an attraction function that measures the probability distribution of grouping.

Several algorithms can process proprioceptive stimuli in this manner. Consider our prior work [8] on the simple case of spindle model output of a single muscle fiber, which is 2-D Ia and II afferent spike trains over time. Figure 6.3 shows a higher-dimensional case of a simulated human arm. When examining proprioceptive signals that are encoded as spike trains in units of pulses per second (pps), we are presented with unlabeled sample points $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$, where n is the set of observations, that can be further mathematically expressed to form representations. Since the notion of proximity is to associate observed points by measurement of Euclidean distance, then a standard unsupervised algorithm such as the K-Means clustering (i.e., a simplified version of vector quantization) proves to be sufficient for revealing underlying data structure.

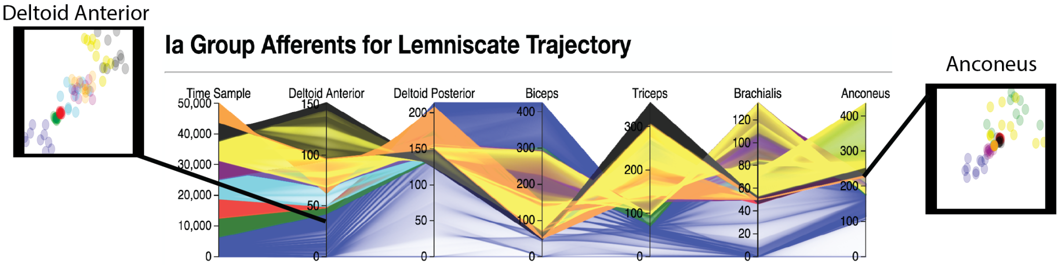


Figure 6.3: Spike trains from spindle afferents produce an evolutionary high-dimensional time-varying manifold of raw afferent information that is distinct for different arm movements. Using parallel coordinates, we show the Ia Group Afferent in 50,000 time samples for the case for a 6-muscle, 2-joint simulated planar arm performing the Lemniscate (figure-of-eight) trajectory with the end point [8]. The coordinates are colored according to the segmented locations within the duration of the Lemniscate trajectory. The shadow boxes to the left and right of the manifolds are scaled-down sample snapshots of the data for the Deltoid Anterior and Anconeus muscles, respectively; ultimately revealing their specific cluster ranges.

6.3.2 Law of Similarity

Elements (e.g., points) that are similar in visual appearance in at least one degree with alike components are more likely to be grouped and organized together perceptually. The Law of Similarity generally spans the attributes of orientation, texture, color, and shape (see Figure 6.2b). There are ways to apply this law to the manifolds produced by proprioceptive signals. Considering the contours and curves that emerge from the collection of proprioceptive manifolds (e.g., Figure 6.3), shape is the most applicable attribute when measuring for similarity. Shape dimensions, such as curvature and elongation, can be perceived as integral dimensions and also used for comparison for similarity. In a similar fashion that [18] quantifies geometric similarity of anatomical surfaces and morphological identification, we can apply statistical analysis when viewing the Ia and II stimuli as a collection of discrete or continuous points on an anatomical surface. Measures of similarity of each afferent signal across various tasks can be applied across the collected time-series data using signal processing. Comparable to the Law of Proximity, K-means clustering may also be used here if measuring 'similarity' of clusters by its relation to Euclidean distance of data points.

$$\cos\theta = \frac{\overrightarrow{x} \cdot \overrightarrow{y}}{\| \overrightarrow{x} \| \, \| \overrightarrow{y} \|} = \frac{\sum_1^n x_i y_i}{\sqrt{\sum_1^n x_i^2} \sqrt{\sum_1^n y_i^2}} \tag{6.1}$$

In our example, let x and y be two vectors of afferent spike trains, Ia and/or II. The cosine similarity function is a measure of similarity that can be used to compare afferent signals in the inner product space. Using the cosine measure, we have Eq. (6.1) where $\overrightarrow{x} \cdot \overrightarrow{y} = \sum_1^n x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$ is the dot product of the two vectors. A cosine similarity, $\cos\theta$, value closer to 1 indicates a higher propensity for perceptual clustering along the manifold. The convolution function would be another choice that quantifies similarity over time for all possible lags between signals.

Another option is cross correlation. It compares the time-series of afferent data across tasks, and is represented as the ratio in Eq. (6.2), where n is the total number of data point indices recorded per task cycle. Both x_i and y_i are the individual spindle afferent sets, Ia and II, respectively. A temporal shift delay, phase lag τ, of the output cross correlation, R_{xy}, measure is applied to determine where the correlation of the data is maximized, as shown in Eq. (6.3).

$$R_{xy}(\tau) = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2 \sum_{i=1}^n (y_i - \overline{y})^2}} \tag{6.2}$$

$$\tau_{estimated} = \arg\max_{\tau \in \mathbb{R}}(R_{xy}(\tau)) \tag{6.3}$$

Magnitude-squared coherence is similar to correlation except that signals are compared in frequency ω, instead of time space, as shown in Eq. (6.4), which values satisfy $0 \leq C_{xy}(\omega) \leq 1$. $S_{xy}(\omega)$ represents the cross-sprectral density between x and y, while $S_x(\omega)$ and $S_y(\omega)$ are the autospectral densities for their respective signals.

$$C_{xy}(\omega) \triangleq \frac{\|S_{xy}(\omega)\|^2}{S_x(\omega)S_y(\omega)} \tag{6.4}$$

$$\begin{aligned} S_{xy}(\omega) &= \int_{\infty}^{\infty} R_{xy}(t)e^{-j\omega t}dt \\ &= \int_{\infty}^{\infty}\left[\int_{\infty}^{\infty} x(\tau)\cdot y(\tau+t)d\tau\right] \end{aligned} \tag{6.5}$$

Lastly, Kullback–Leibler (K-L) divergence is a means to quantify the likelihood that the statistics of a given process are similar to that of another, Eq. (6.6). Probability distributions P and Q are measured in comparison to reveal the relative entropy. This is particularly useful because it measures how much information is lost when we approximate distributions.

$$D(P \parallel Q) = \sum_{x \in X} P(x)\log\left(\frac{P(x)}{Q(x)}\right) \tag{6.6}$$

6.3.3 Law of Closure

The Law of Closure is the tendency to complete unfinished or partially obscured objects. Here, incomplete figures are seen as complete or whole as depicted in Figure 6.2c. Warshall's Algorithm [185] may address this through its approach in computing the transitive of a node relation in a graph. We can envision, that as clusters are being formed via other laws, state nodes will eventually emerge from the aggregate data. To establish state transitions from one afferent cluster to another, the Warshall algorithm can determine whether a vertex j is 'reachable' from another vertex i for all vertex pairs within the graph. This measure of reachability will serve as the transitive closure, indicating directions and where paths exist for point-to-point movement across the manifolds.

This law states that, given available information, there is the expectation (based on prior personal experience) of closure when a fragmented version is presented. Bayes' Rule is a formal way to represent such expectation in the case of visual information, visuomotor perception [97], and now proprioception. Bayes' rule states that we can obtain the posterior distribution (the probability of a given body state given current proprioceptive input $p(x_{true}|x_{sensed})$ by taking into account the likelihood distributions of the prior (i.e., the cumulative information from prior experience) and the evidence (i.e., the current proprioceptive input x_{sensed}):

$$p(x_{true}|x_{sensed}) = p(x_{sensed}|x_{true}) * \frac{p(x_{true})}{p(x_{sensed})} \tag{6.7}$$

where $p(x_{sensed}|x_{true})$ is the likelihood of a particular proprioceptive input x_{sensed} when the perceived body state really is true. This then allows the inference of the current body state given past experience and incomplete or polluted proprioceptive inputs.

6.3.4 Law of Continuity

Objects and points that are co-linear and follow the same direction will be grouped together as a whole (see Figure 6.2d). We can construct proficient continuations between neighboring local environments. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) identifies outliers as noises. The Mean-shift algorithm, Eq. (6.8), actually includes them in the cluster despite differences of the data point. DBSCAN also does not require a pre-set number of clusters, and discovers arbitrarily shaped clusters. These are key facets for analysing proprioceptive data. Given the manifolds of afferent information for natural movements are usually continuous, then the Law of Continuity would naturally apply as the manifold during a movement continues along a particular path, even if temporarily disrupted or occluded by a perturbation. In practice, Bayes' Rule is a way in which such expectation of continuity can be quantified.

6.3.5 Law of Prägnanz (Good Form, Clarity)

The Law of Prägnanz focuses on simplicity and will prompt visualizations according to the simplest way of grouping items. We perceptually organize shapes to simple forms, as in pithiness. The Law of Prägnanz is the tendency to interpret ambiguous images as simple and complete vs. complex and incomplete. An example is how shapes overlapping each other can cause ambiguity, as shown in Figure 6.2e. A potential resolution is an iterative method such as Mean-shift Clustering, Eq. (6.8), where $N(x)$ is the neighborhood of the set of points, x. Depending on the Gaussian kernel bandwidth, Eq. (6.9), the Mean-shift algorithm iteratively shifts points until there is a convergence of partitioning the clusters into semantically meaningfully groups. This is probable to work well with proprioceptive afferents as it may account for the noise in signals which is expected, and necessary for physiological function.

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)} \tag{6.8}$$

$$K(x_i - x) = e^{-c\|x_i - x\|^2} \tag{6.9}$$

Dimensionality reduction is probably the most commonly applied (and potentially misinterpreted) analysis of high-dimensional motor signals [102]. It is simply a way to quantify whether a high-dimensional ensemble of signals evolves (i.e., has variance) along all dimensions equally, or inhabits a lower-dimensional subspace. Conceptually, it is just the singular value decomposition of a covariance matrix, where the number of ‘large’ singular values (principal components) quantifies the rank of the covariance (the effective ‘dimensionality’ of the data), and the left singular vectors (principal vectors) form a basis for those dominant variances (the basis for the effective subspace the data inhabit). Independent Component Analysis and Nonnegative Matrix Factorization is a variations on this idea that do not require orthogonality of the basis vectors, and the latter also imposes a non-negative constraint on the elements of the basis (as neural signals are conceptualized as intensities or spiking frequencies that are > 0). It is good to see that some work is beginning to be done on dimensionality reduction in tactile afferents, which are famously difficult to record from even in animal preparations [152]. Our current work is beginning to apply dimensionality reduction to higher-dimensional simulated proprioceptive signals [6].

6.3.6 Supplementary Laws

Other Gestalt grouping laws that can be applied to sensory stimuli integration include the Laws of Focal Point, Symmetry, Common Fate, Common Region, Synchrony, Convexity, Isomorphism, Parallelism, Unity, Element Connectedness, and Figure vs. Ground.

6.4 Functional Utility of the Sensory-Motor Gestalt

Figure 6.4 describes our working hypothesis of the Sensory-Motor Gestalt in operation. At any time point, proprioceptive (and other sensory) information define a state of the body that lies within a particular manifold of like inputs (Laws of Proximity and Similarity) and feasible next states (Laws of Continuity and Closure). Such body state allows feasible transitions to ‘next’ proprioceptive states via feasible motor actions that will lead to a, usually predicted, new body state (Laws of Continuity and Closure).

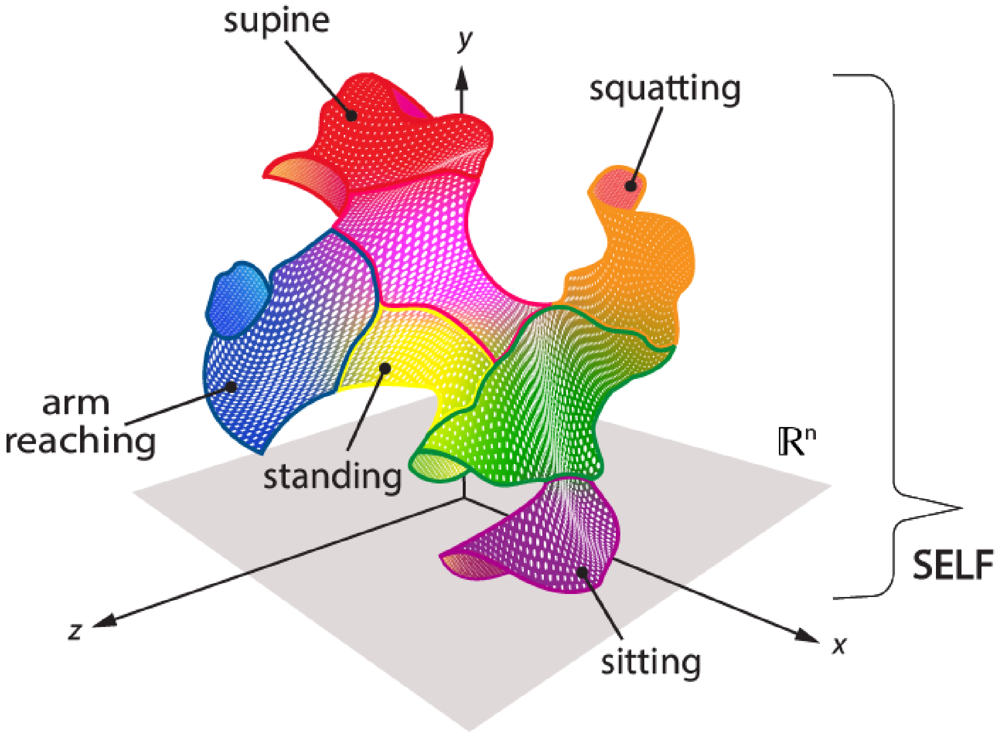


Figure 6.4: We envision the representation of minimal self as a collection of categorized states in $\mathbb{R}^N$ space formed from sensory and motor maps, and made useful by the agency they provide. Our data-driven projection method categorizes the set of feasible inputs from muscle spindles for each specific task performed (i.e., arm reach, sit, squat, standing) as a manifold. Transitions from one state to another occur through point-to-point transitions along the manifold. The high-dimensional space of afferent modalities has an underlying structure given by the anatomy of the body and the physical transitions it can undergo such as changing postures via self movement.

6.5 Abstracting Self from Sensorimotor Experiences for Neuromuscular Systems

Now let's examine how the foundations of neuromuscular systems can provide context to constructing the minimal operative self via proprioceptive signals. In Nature, proprioception provides animals with awareness of the state of their body and of their relation to the environment.

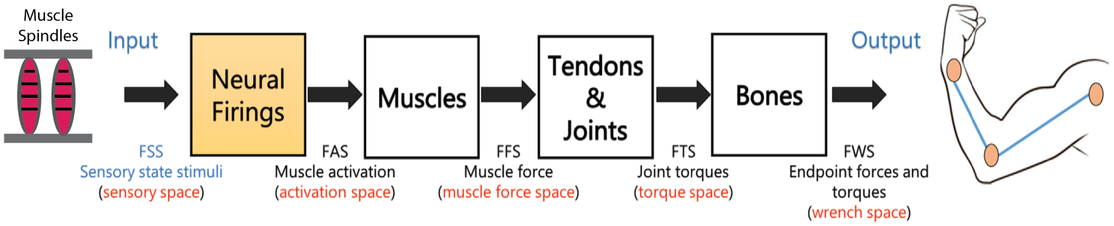


Figure 6.5: The neuromechanical perspective of how sensory inputs are transformed to motor outputs (adapted from [178]). A Feasible Sensory Set (FSS) defines the afferent stimuli that are plausibly detectable for a given state of the body (i.e., joint posture, force production, and kinematic task). By incorporating the influence of proprioceptive space via neural spike firings, an under-constrained mapping of transformations can be reinforced from neural motor commands in the Feasible Activation Sets (FAS) to mechanical outputs (limb movements).

Proprioceptive signals arise from mechanoreceptors that reflect the state of tissues, which are driven by muscle forces, joint and body postures, and skin deformations. When integrated with other sensory modalities, this reflection of body state at any given moment in time and space provides the nervous system with an overall representation of bodily position, actions, and task experiences. Neuroscientists have long been intrigued with how the brain represents the body and forms models of bodily states through proprioception [69]. However, there is still no consensus regarding how these representations, facilitated by multi-muscle control, compartmentalize and process high-dimensional afferent information as continuous feedback for ongoing tasks.

The fundamental formulation of a control law for a linear system (without loss of generality) is

$$\dot{x} = Ax + Bu \tag{6.10}$$

where the outputs y (and therefore sensory and proprioceptive signals) are a function of the state x and the control signals u

$$y = Cx + Du \tag{6.11}$$

By definition, the equations of motion (i.e., $\dot{x} = Ax$) are an important determinant of the feasible transitions away from any given state. Moreover, changes in sensory and proprioceptive signals

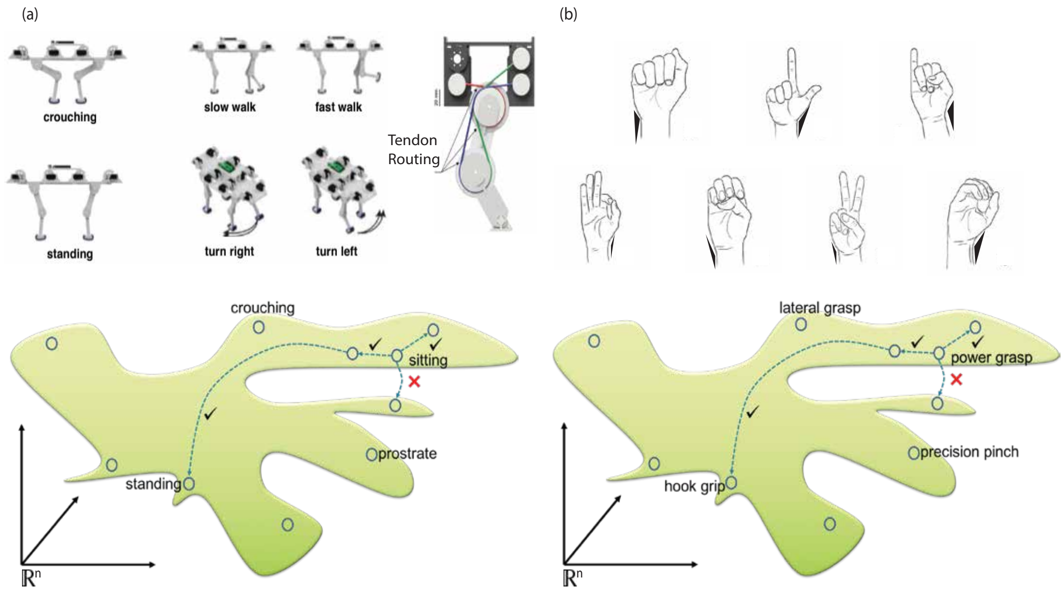


Figure 6.6: Test-bed applications for Sensory-Motor Gestalt implementation in tendon-driven systems. The Gestalt can provide state cases and their feasible transitions along the manifold for agents. (**a**) Proposed neuromorphic cat-like robot and limb in hardware. Image adapted from [116]. (**b**) Various hand and finger states depicted using the American Sign Language. The nervous system registers the proprioceptive feedback generated with each hand shape is unique to each letter signed with gestures.

are driven by changes in state (i.e., $y = Cx$). This is a formal way to conceptually anchor some aspects of the Sensory-Motor Gestalt. *Please note we do not claim or endorse that the engineering concept of 'state' applies to biology. But the Sensory-Motor Gestalt is a formal way to describe how the stream of sensory and proprioceptive signals is useful to biological behavior in a way that is agnostic to how those signals are processed.*

We can conjecture how the nervous system processes incoming afferents (e.g., proprioceptive signals) by observing how neural activation commands mathematically map to mechanical outputs, as shown in Figure 6.5. Neural commands simply refer to the nervous system's distribution of excitatory impulse signals to activate muscle tissue. For tendon-driven limbs, [178] emphasizes that the nervous system's primary function is to use (i.e., learn, explore, and exploit) the set of feasible neural commands from the optimized activation space with dimensionality of vector

$\boldsymbol{a} \in \mathbb{R}^N$, where N is the number of independently controlled muscles. From activation space, vector spaces are successively mapped to muscle force space, to joint torque space, then lastly to output wrench space to produce a set of feasible mechanical outputs (forces and movements).

In prior work [8], we have extrapolated this perspective of muscle redundancy to feasible sets of proprioceptive signals, called Feasible Sensory Sets (FSS). These are defined by a body's anatomical structure and the mechanical tasks being performed. Here, we first introduced the concept of trajectory-specific proprioceptive manifolds, which are the unique multidimensional and time-varying combinations of afferent signals that obligatorily emerge during a limb movement. We demonstrated that a given movement gives rise to a distinct sensory manifold embedded in the 12-D space of spindle information that is largely independent of the choice of muscle coordination strategy. These are referred to as manifolds because they are a systematic collection of points (i.e., spindle neural spikes) that provide information for its control.

Following this work it remained unknown whether spindle signals suffice to discriminate limb movements. We used a 4-musculotendon, 2-joint cat hindlimb model to simulate muscle spindle length and velocity signals (II and Ia, respectively) during repeated cycles of five distinct end-point movements, similar to the manifolds in Figure 6.3. In [6], we concluded that proprioceptive information can usefully discriminate limb states—but only after conducting minimal pre-processing of high-dimensional multi-muscle ensembles to low-dimensional subspace components. This finding may this explain the documented subcortical pre-processing of afferent signals of various mammals [152]. It is this resulting set of constrained sensory signals that we believe could suffice as a minimal representation of the artificial self and should be incorporated into the Sensory-Motor Gestalt paradigm. We project the usefulness of Sensory-Motor Gestalt to be a suitable core to execute on different applications that utilize neuromuscular dynamics, incorporate neuromorphic and bio-inspired architectures, and classification of human bodily states (Figure 6.6).

6.6 Role of Self and Identity in Autonomous Robotics and Synthetic Biological Agents

A semblance of selfhood, identity, and agency should be expected outcomes for constructing a dynamic sensorimotor representation [171]. For robots, concepts of identity are typically viewed as a necessity for interactions in social environments [43]. For humans, the self and identity combination are purposed for storing the traits, stereotypes, characteristics, and roles they play in social settings. What features constitute a person's self? How do disparate sensory perceptions cohesively fuse together to form a singular experience of self? Although these questions are typically addressed within the human scope, we can also apply these inquires to autonomous robots that are bioinspired and create their own experiences with action.

It is our opinion that sensorimotor contingencies (discussed in Related Work) do not achieve their full potential if solely used to estimate error signals in closed-loop controllers. We believe one can ask the extent to which these contingencies facilitate and embody a self, reflect an identity, and activate agency needs to be thoroughly explored. Self and identity are often used interchangeably to encapsulate the entirety of individual's behavior, character, and the restricted contextual constraints in which they operate within. However, it is important to clearly know the distinctions of these terms if we're determined to adequately construct models that emulate their functions.

Agency is known as the control of intentional actions and volition; leading to the ability to plan and action ownership. For the purpose of our study we distinguish the self and identity according to Oyserman's [136] conceptualization. It is thought best to consider self, self-concept, and identity as nested elements: self is the top-tier construct, self-concepts reside within the self, and identities reside within self-concepts. Oyserman defines self as the ability to consider oneself as an object. The self maintains reflexive capacity that is able to direct an agent to what is "me"; it is the focal point of personal account and a reference for anchoring temporal sequences of events (e.g., memory recall). Identities are “content and readiness to act and employ mindsets to make meaning.” Personal identities are the traits, characteristics, values and goals belonging to the agent. Altogether self and identity are mental concepts, social products, and forces of action. As Oyserman states, what makes this nested unit interesting is that they appear to predict behavior over time. What is not fully understood by many in literature is *how* this happens.

6.7 Related Work

Further research into the topic of dynamic sensorimotor representations led us to original work on the sensorimotor contingency theory [134], which has motivated an assortment of studies in the area of human perception as it relates to understanding the nature of actions and their sensory effects. Sensorimotor contingencies derived from the notion that vision should be treated as an environmental exploratory activity. According to [77], sensorimotor contingencies spawned multidisciplinary projects that investigated how to model the action-sensory relationship of robotic systems, which spanned the manipulation, classification, and categorization of external objects. The researchers view the goal for most of these studies as autonomous robots learning skilled behaviors via learning the structure of complex sensorimotor spaces and how actions affect the environment. Despite these contributions, [20] believed there have been few attempts to formally define sensorimotor contingencies, which they view as a prerequisite for testing this approach via models and empirical study. The sensorimotor contingencies view on perceptual awareness have also been criticized for lacking a suitable foundation in the biology of autonomous agency. Prior work on building computational approaches of body representations, self, identity and Gestalt have been attempted with a grounding in minimal embodied, psychological and cognitive aspects [56]. One drawback of past implementations is that they’re unencumbered with understanding the manifold of feasible transitions, which therefore leads to the incorrect perspective that any action is permissible. Our approach addresses how the sensorimotor self should constrain one’s agency and perceptual space to feasible tasks. Attempts to create sensory-to-motor maps as a body representation (i.e., body schema) have been accomplished by achieving robotic self-recognition using a dynamic Bayesian network [65], online learning of arm reaching motor maps for humanoid robots using open and closed loop control [62], body representations as cross-modal map learning of invariance in multi modal sensory data [190], and estimation of a kinematic model for serial robots [117]. To our knowledge, our approach is the first that considers the inherent link among the feasible capabilities of the body, the feasible sensory information that will emerge, and the physics of the world as the manifold defining agency, and therefore delineating the concept of self.

6.8 Conclusion

The emergence of self and its role in biological and artificial agents continues to be a subject of debate across many disciplines; leading to the perception that there is a lack of congruence among perspectives. The absence of a unified concept of self presented us with an opportunity to propose a sensorimotor mechanism by which the self can emerge, via the Gestalt laws of perceptual organization, in the context of artificial systems operating in the physical world (i.e., robots). This enables us to investigate the foundation of self, identity, learning, and agency as the multifaceted interplay of proprioception and action while exploring their implications to autonomy. The emergence of self through sensorimotor interactions has applications ranging from a self-other distinction to ‘social’ systems for robot-human and robot-robot interactions. Traditionally, self and identity are considered to be theoretical concepts, social constructs, and therefore enablers of agency. We visit these concepts in reverse order to propose that it is sensorimotor agency that can enable the emergence of self and identity—which is an evolutionarily plausible order of events [188]. The Sensory-Motor Gestalt provides a solid foundation to enable such cross-fertilization to move towards the creation of truly autonomous and versatile robots, and promote advances in artificial intelligence.

Chapter 7

Conclusion

This work's long-term goal is to equip a robot with an emulated nervous system (i.e., a NeuRoBot) that forms a repertoire of physiologically-inspired sensory and motor couplings to explore and exploit physical actions and transitions among them. For this thesis, we have focused on building the software framework of the sensorimotor couplings, which can then be merged with the hardware plant of the model. From our initial hypothesis, we anticipated that the physiological sensory signals that result directly from a range of immature to skilled motor actions would suffice to create a self-generated body representation that prompts learning of useful, functional actions and potentially evolving behaviors. We believe the work successfully met the criteria for forming useful body representations for tendon-driven systems such as the NeuRobot.

Using various computational methods, we demonstrated the plausibility that an individual body representation of the self (i.e., body schema) for neuromuscular-driven robotic systems can be self-generated from:

- Defining feasible behaviors and muscle activation patterns that produce task-oriented limb mechanics.

- Using computational integration of multimodal, spatiotemporal sensorimotor contingencies of physiological realistic (artificial) proprioceptive afferents.

- Quantifying somatosensory modalities: Muscle spindle for posture, GTO for muscle force.

- Encoding multisensory mathematical formalisms for system-specific state space and actions.

In this thesis, we addressed the body senses that arise (e.g., muscle spindle afferents and visual cues) to enable future studies investigating how those same sensory signals affect the representation of our physical self and the actions we make. First, I demonstrated that a given movement gives rise to a distinct sensory manifold embedded in the 12-D space of muscle spindle information that is largely independent of the choice of muscle coordination strategy. Given that musculotendon lengths and velocities are fully determined by joint kinematics, such manifolds provide a rich set of information to use in its control. Secondly, I show that high-dimensional multi-muscle proprioceptive ensembles can usefully discriminate limb states and be utilized as a suitable classifier for inter-trajectory comparisons—but only after minimal pre-processing. Lastly, we present the Sensory-Motor Gestalt, which through computational approaches, demonstrates how the syntactical relationship between motor actions and bodily sensory states can lead to an epistemological continuum that streamlines the body state into useful behavior for constructing the foundations of agency, identity, and sense of self in hybrid robots and synthetic biological agents.

To conclude, let's discuss some of the challenges that were undertaken and how we intend to map out contingency plans for future research directions. We sought to address an underlying question: how sensory signals shape the motor choices we make throughout daily life activities or in simple task actions? From our past research experiments, we have learned that every limb movement is associated with a unique set of specific sensory states, which we call the Feasible Sensory Set (FSS). We must consider how the nervous system obtains and processes sensory data to create a body sense that interacts with explicit or implicit internal models of the body and external influences on the body. Distinguishing oneself from the environment and having an introspective representation of self is a fundamental, biological challenge that the CNS of animals and humans must encounter and solve daily. Within that self-representation is a sense of ownership, that our actions, behaviors, and bodies belong to a certain individual, which is key for survival and performance. Although there were once only relatively few studies on the awareness of one's body [44], increasing volumes of research over the years has allowed us to make a commitment to this thesis and fathom the realistic plausibility of creating self-aware devices.

The novelty in our work focused on addressing the remaining important challenges and limitations in self-modeling based on the spatial constraints of states, the context of multisensory integration, and control for exploration. Also, the distinction in our work is that we are explicitly analyzing the unavoidable, physiologically basic, high-dimensional FSSs that accompany each movement. This can be made possible by the use of interactive parallel coordinates and graphical animation tools. Another key feature is the methodical design for autonomic self-knowledge that will equip the system with the ability to express the causative relationships between sensory inputs and the efferent motor events that are activated. Prediction of expected afferentation outcomes will be a consequence of this relationship. From the perspective of neuromuscular systems, the mathematical and geometric perspective we applied served to ground the arguments of muscle abundance and the feasibility of task constraints. Our methods for obtaining these results can be employed towards robotics, and brain-machine interfaces (BMIs) optimized on the limits of simulated neural drive obtained from sensory inputs. Machines that possess self-aware computing will relieve engineers, such as ourselves, of the need to assess system constraints and resource availability at design-time. This need has the potential to be either be significantly reduced or eliminated. The goal with our proposed framework was to not directly install a full-body representation or model into a suitable hardware host but to instead provide it with the rules, principles, and fortitude to form a personal schema on its own merit...from the basis of proprioceptive signals. We believe our software implementation will provide valuable insight for engaging a new generation of systems featured with computational self-modeling that will showcase a level of robustness and flexibility that tendon-driven systems in the robotics field. At this time, the framework design is tentative and open for modification as the project develops in the future. Our past research in the 1) Mirror Neuron System for reaching arm tasks, 2) Agency according to self-other dichotomy processing, 3) translating psychological self-awareness to machine subjective experience by way of neuro-physiology, and 4) the forming of feasible sensory sets has primed us for making progress in the phase of implementing the actualization of self-aware computing.

This thesis helped us make to make an effective contribution to the field of self-model computing and assist in providing firm answers to the following inquiries:

- How do we understand and interact with the world around us (solely via afferentation)?
- How do we display that sensory understanding with motor control?
- What processes occur to allow us to make sense of the world via our senses?
- What are the sensorimotor dynamics that lead to artificial self-modeling and self-awareness?

A self-aware system, which we discussed in the Background section, should be able to comprehend and facilitate the overheads and tradeoffs associated with the act of learning information about itself. Subsequently, we can make fundamental progress in interpolating sensorimotor schemas of self-impressions and experiences and their effects on the acquisition of knowledge. If developed correctly, the system's ability to self-generate its states can contribute to the goal of transforming autonomous systems into devices that have the means to form self-knowledge of experiences (discover conditions and patterns that occur during operation). Through useful body representations, systems can obtain the ability to evaluate their effectiveness while improving their action planning and decision-making capabilities. Overall, we hope to have contributed to the aim of robot capabilities to be more easily extended and adapted to novel situations and lifelong learning.

Bibliography

[1] Eyal Amir, Michael L Anderson, and Vinay K Chaudhi. Report on darpa workshop on self-aware computer systems. Technical report, 2007.

[2] Beulah Amsterdam. Mirror self-image reactions before age two. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 5(4):297–305, 1972.

[3] James R Anderson and Gordon G Gallup Jr. Which primates recognize themselves in mirrors? *PLoS Biology*, 9(3), 2011.

[4] AF Barto, S Sutton, and Charles W Anderson. Neuronlike adaptive elements that can solve difficult systems. *IEEE Trans. Sys. Man Cyber*, 13(5):834–846, 1983.

[5] Nicolò F Bernardi, Mohammad Darainy, and David J Ostry. Somatosensory contribution to the initial stages of human motor learning. *Journal of neuroscience*, 35(42):14316–14326, 2015.

[6] Jasmine Berry, Ali Marjaninejad, and Francisco J Valero-Cuevas. Minimal pre-processing of multi-muscle ensembles of spindle signals improves discriminability of limb movements, 2020. In Preparation.

[7] Jasmine A Berry and Alice C Parker. The elephant in the mirror: Bridging the brain's explanatory gap of consciousness. *Frontiers in systems neuroscience*, 10:108, 2017.

[8] Jasmine A Berry, Robert Ritter, Akira Nagamori, and Francisco J Valero-Cuevas. The neural control of movement must contend with trajectory-specific and nonlinearly distorted manifolds of afferent muscle spindle activity. In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pages 1188–1194. IEEE, 2017.

[9] Bennett I Bertenthal and Kurt W Fischer. Development of self-recognition in the infant. *Developmental psychology*, 14(1):44, 1978.

[10] Lynda IA Birke and Ruth Hubbard. *Reinventing biology: Respect for life and the creation of knowledge.* Georgetown University Press, 1995.

[11] Kwabena Boahen. A neuromorph's prospectus. *Computing in Science & Engineering*, 19(2):14–28, 2017.

[12] Tamara Bonaci, Jeffrey Herron, Charlie Matlack, and Howard Jay Chizeck. Securing the exocortex: A twenty-first century cybernetics challenge. In *2014 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*, pages 1–8. IEEE, 2014.

[13] James Bonaiuto, Edina Rosta, and Michael Arbib. Extending the mirror neuron system model, i. *Biological cybernetics*, 96(1):9–38, 2007.

[14] Josh Bongard and Hod Lipson. Evolved machines shed light on robustness and resilience. *Proceedings of the IEEE*, 102(5):899–914, 2014.

[15] Josh Bongard, Victor Zykov, and Hod Lipson. Resilient machines through continuous self-modeling. *Science*, 314(5802):1118–1121, 2006.

[16] Luca Bonini, Monica Maranesi, Alessandro Livi, Leonardo Fogassi, and Giacomo Rizzolatti. Space-dependent representation of objects and other's action in monkey ventral premotor grasping neurons. *Journal of Neuroscience*, 34(11):4108–4119, 2014.

[17] Luca Bonini, Monica Maranesi, Alessandro Livi, Leonardo Fogassi, and Giacomo Rizzolatti. Ventral premotor neurons encoding representations of action during self and others' inaction. *Current Biology*, 24(14):1611–1614, 2014.

[18] Doug M Boyer, Yaron Lipman, Elizabeth St Clair, Jesus Puente, Biren A Patel, Thomas Funkhouser, Jukka Jernvall, and Ingrid Daubechies. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. *Proceedings of the National Academy of Sciences*, 108(45):18221–18226, 2011.

[19] Marcel Brass and Patrick Haggard. The hidden side of intentional action: the role of the anterior insular cortex. *Brain Structure and Function*, 214(5-6):603–610, 2010.

[20] Thomas Buhrmann, Ezequiel Alejandro Di Paolo, and Xabier Barandiaran. A dynamical systems account of sensorimotor contingencies. *Frontiers in psychology*, 4:285, 2013.

[21] Vittorio Caggiano, Leonardo Fogassi, Giacomo Rizzolatti, Peter Thier, and Antonino Casile. Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. *science*, 324(5925):403–406, 2009.

[22] Thomas A Carlson, George Alvarez, Daw-an Wu, and Frans AJ Verstraten. Rapid assimilation of external objects into the body schema. *Psychological Science*, 21(7):1000–1005, 2010.

[23] Laurie Carr, Marco Iacoboni, Marie-Charlotte Dubeau, John C Mazziotta, and Gian Luigi Lenzi. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the national Academy of Sciences*, 100(9):5497–5502, 2003.

[24] Rita Carter. *Mapping the mind*. Univ of California Press, 1999.

[25] Emilie A Caspar, Andrea Desantis, Zoltan Dienes, Axel Cleeremans, and Patrick Haggard. The sense of agency as tracking control. *PloS one*, 11(10):e0163892, 2016.

[26] David Chalmers. The hard problem of consciousness. *The Blackwell companion to consciousness*, pages 225–235, 2007.

[27] Dempsey Hsiu-Ju Chang. The gestalt principles of similarity and proximity apply to both the haptic and visual grouping. In *Eighth Australasian User Interface Conference (AUIC2007)*, pages 79–86. Flinders Press, 2007.

[28] Hillel J Chiel, Lena H Ting, Örjan Ekeberg, and Mitra JZ Hartmann. The brain in its body: motor control and sensing in a biomechanical context. *Journal of Neuroscience*, 29(41):12807–12814, 2009.

[29] Patricia Smith Churchland. Reduction and the neurobiological basis of consciousness. 1988.

[30] Philip Clayton. Mind and emergence: From quantum to consciousness. 2004.

[31] Arthur D Craig. Human feelings: why are some more aware than others? *Trends in cognitive sciences*, 8(6):239–241, 2004.

[32] Arthur D Craig and AD Craig. How do you feel–now? the anterior insula and human awareness. *Nature reviews neuroscience*, 10(1), 2009.

[33] Francis Crick and Christof Koch. A framework for consciousness. *Nature neuroscience*, 6(2):119–126, 2003.

[34] Hugo D Critchley and Neil A Harrison. Visceral influences on brain and behavior. *Neuron*, 77(4):624–638, 2013.

[35] Hugo D Critchley, Stefan Wiens, Pia Rotshtein, Arne Öhman, and Raymond J Dolan. Neural systems supporting interoceptive awareness. *Nature neuroscience*, 7(2):189–195, 2004.

[36] Gislin Dagnelie. Retinal implants: emergence of a multidisciplinary field. *Current opinion in neurology*, 25(1):67–75, 2012.

[37] Antonio Damasio and Raymond J Dolan. The feeling of what happens. *Nature*, 401(6756):847–847, 1999.

[38] Avinash De Sousa. Towards an integrative theory of consciousness: part 1 (neurobiological and cognitive models). *Mens sana monographs*, 11(1):100, 2013.

[39] Monique W De Veer and Ruud Van den Bos. A critical review of methodology and interpretation of mirror self-recognition research in nonhuman primates. *Animal Behaviour*, 58(3):459–468, 1999.

[40] Fabienne Delfour and Ken Marten. Mirror image processing in three marine mammal species: killer whales (orcinus orca), false killer whales (pseudorca crassidens) and california sea lions (zalophus californianus). *Behavioural processes*, 53(3):181–190, 2001.

[41] Daniel C Dennett. *Consciousness explained.* Penguin uk, 1993.

[42] Yan Duan, Marcin Andrychowicz, Bradly Stadie, OpenAI Jonathan Ho, Jonas Schneider, Ilya Sutskever, Pieter Abbeel, and Wojciech Zaremba. One-shot imitation learning. In *Advances in neural information processing systems*, pages 1087–1098, 2017.

[43] Brian Duffy. Robots social embodiment in autonomous mobile robotics. *International Journal of Advanced Robotic Systems*, 1(3):17, 2004.

[44] H Henrik Ehrsson, Charles Spence, and Richard E Passingham. That's my hand! activity in premotor cortex reflects feeling of ownership of a limb. *Science*, 305(5685):875–877, 2004.

[45] Nathan Faivre, Anat Arzi, Claudia Lunghi, and Roy Salomon. Consciousness is more than meets the eye: a call for a multisensory study of subjective experience. *Neuroscience of consciousness*, 2017(1):nix003, 2017.

[46] MN Fargeau, N Jaafari, S Ragot, JL Houeto, C Pluchon, and R Gil. Alzheimer's disease and impairment of the self. *Consciousness and Cognition*, 19(4):969–976, 2010.

[47] Ali Farshchiansadegh, Alejandro Melendez-Calderon, Rajiv Ranganathan, Todd D Murphey, and Ferdinando A Mussa-Ivaldi. Sensory agreement guides kinetic energy optimization of arm movements during object manipulation. *PLoS computational biology*, 12(4):e1004861, 2016.

[48] Todd E Feinberg. Neuroontology, neurobiological naturalism, and consciousness: a challenge to scientific reduction and a solution. *Physics of Life Reviews*, 9(1):13–34, 2012.

[49] WR Ferrell, SC Gandevia, and DI McCloskey. The role of joint receptors in human kinaesthesia when intramuscular receptors cannot contribute. *The Journal of physiology*, 386(1):63–71, 1987.

[50] Andrew A Fingelkurts, Alexander A Fingelkurts, and Tarja Kallio-Tamminen. Long-term meditation training induced changes in the operational synchrony of default mode network modules during a resting state. *Cognitive Processing*, 17(1):27–37, 2016.

[51] Andrew A Fingelkurts, Alexander A Fingelkurts, and Tarja Kallio-Tamminen. Trait lasting alteration of the brain default mode network in experienced meditators and the experiential selfhood. *Self and Identity*, 15(4):381–393, 2016.

[52] Andrew A Fingelkurts, Alexander A Fingelkurts, and Carlos FH Neves. "machine" consciousness and "artificial" thought: An operational architectonics model guided approach. *Brain research*, 1428:80–92, 2012.

[53] American Association for Research into Nervous, Mental Diseases, and JR Searle. How to study consciousness scientifically. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1377):1935–1942, 1998.

[54] Christina T Fuentes and Amy J Bastian. Where is your arm? variations in proprioception across space and tasks. *Journal of neurophysiology*, 2009.

[55] Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1):14–21, 2000.

[56] Shaun Gallagher. A pattern theory of self. *Frontiers in human neuroscience*, 7:443, 2013.

[57] Vittorio Gallese et al. How the body in action shapes the self. *Journal of Consciousness Studies*, 18(7-8):117–143, 2011.

[58] Vittorio Gallese, Luciano Fadiga, Leonardo Fogassi, and Giacomo Rizzolatti. Action recognition in the premotor cortex. *Brain*, 119(2):593–609, 1996.

[59] GG Gallup and SHILLITO DJ ANDERSON JR. The cognitive animal: Empirical and theoretical perspectives on animal cognition, 2002.

[60] Gordon G Gallup. Chimpanzees: self-recognition. *Science*, 167(3914):86–87, 1970.

[61] Bruce J Gantz, Christopher Turner, Kate E Gfeller, and Mary W Lowder. Preservation of hearing in cochlear implant surgery: advantages of combined electrical and acoustical speech processing. *The Laryngoscope*, 115(5):796–802, 2005.

[62] Chris Gaskett and Gordon Cheng. Online learning of a motor map for humanoid robot reaching. *Proc. 2nd Int. Conf. Computat. Intell., Robot. Autonom. Syst. (CIRAS 2003)*, 2003.

[63] James Jerome Gibson. The senses considered as perceptual systems. 1966.

[64] Kevin Gold and Brian Scassellati. A bayesian robot that distinguishes" self" from" other". In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 29, 2007.

[65] Kevin Gold and Brian Scassellati. Using probabilistic reasoning over time to self-recognize. *Robotics and autonomous systems*, 57(4):384–392, 2009.

[66] Melvyn Goodale and David Milner. *Sight unseen: An exploration of conscious and unconscious vision.* OUP Oxford, 2013.

[67] Melvyn A Goodale, A David Milner, et al. Separate visual pathways for perception and action. 1992.

[68] Marcus A Gray, Neil A Harrison, Stefan Wiens, and Hugo D Critchley. Modulation of emotional appraisal by false physiological feedback during fmri. *PLoS one*, 2(6):e546, 2007.

[69] Michael SA Graziano. How the brain represents the body: insights from neurophysiology and psychology. *Common mechanisms in perception and action*, 2000.

[70] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? In *Handbook on ontologies*, pages 1–17. Springer, 2009.

[71] Stuart Hameroff and Roger Penrose. Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and computers in simulation*, 40(3-4):453–480, 1996.

[72] Takashi Hanakawa, Ilka Immisch, Keiichiro Toma, Michael A Dimyan, Peter Van Gelderen, and Mark Hallett. Functional properties of brain areas associated with motor execution and imagery. *Journal of neurophysiology*, 89(2):989–1002, 2003.

[73] Robert M Haralick and Linda G Shapiro. *Computer and robot vision*, volume 1. Addison-wesley Reading, 1992.

[74] Nalin Harischandra and Örjan Ekeberg. System identification of muscle–joint interactions of the cat hind limb during locomotion. *Biological cybernetics*, 99(2):125, 2008.

[75] Justin Wildrick Hart and Brian Scassellati. Mirror perspective-taking with a humanoid robot. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[76] Stephen W Hawking and George Francis Rayner Ellis. *The large scale structure of space-time*, volume 1. Cambridge university press, 1973.

[77] Nicholas Hay, Michael Stark, Alexander Schlegel, Carter Wendelken, Dennis Park, Eric Purdy, Tom Silver, D Scott Phoenix, and Dileep George. Behavior is everything: Towards representing concepts with sensorimotor contingencies. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[78] Markram Henry. The blue brain project. *Nat Rev Neurosci*, 7(2):153â160, 2006.

[79] Louis M Herman. Body and self in dolphins. *Consciousness and cognition*, 21(1):526–545, 2012.

[80] Sayaka Hihara, Tomonori Notoya, Michio Tanaka, Shizuko Ichinose, Hisayuki Ojima, Shigeru Obayashi, Naotaka Fujii, and Atsushi Iriki. Extension of corticocortical afferents into the anterior bank of the intraparietal sulcus by tool-use training in adult monkeys. *Neuropsychologia*, 44(13):2636–2646, 2006.

[81] Thomas T Hills and Stephen Butterfill. From foraging to autonoetic consciousness: The primal self as a consequence of embodied prospective foraging. *Current Zoology*, 61(2):368–381, 2015.

[82] Matej Hoffmann, Hugo Marques, Alejandro Arieta, Hidenobu Sumioka, Max Lungarella, and Rolf Pfeifer. Body schema in robotics: a review. *IEEE Transactions on Autonomous Mental Development*, 2(4):304–324, 2010.

[83] Katherine RS Holzbaur, Wendy M Murray, and Scott L Delp. A model of the upper extremity for simulating musculoskeletal surgery and analyzing neuromuscular control. *Annals of biomedical engineering*, 33(6):829–840, 2005.

[84] Pierre Jacob and Marc Jeannerod. Ways of seeing: The scope and limits of visual cognition. 2003.

[85] Kian Jalaleddini, Chuanxin Minos Niu, Suraj Chakravarthi Raja, Won Joon Sohn, Gerald E Loeb, Terence D Sanger, and Francisco J Valero-Cuevas. Neuromorphic meets neuromechanics, part ii: the role of fusimotor drive. *Journal of neural engineering*, 14(2):025002, 2017.

[86] W James. The principles of psychology, vol. 2. henry holt and company, 1890.

[87] Harry J Jerison. Evolution of the brain and intelligence. *Current Anthropology*, 16(3):403–426, 1975.

[88] Mari Jibu, Kunio Yasue, and K Yasue. *Quantum brain dynamics and consciousness.* John Benjamins C., 1995.

[89] Eric R Kandel, Henry Markram, Paul M Matthews, Rafael Yuste, and Christof Koch. Neuroscience thinks big (and collaboratively). *Nature Reviews Neuroscience*, 14(9):659–664, 2013.

[90] Eric R Kandel, James H Schwartz, Thomas M Jessell, Department of Biochemistry, Molecular Biophysics Thomas Jessell, Steven Siegelbaum, and AJ Hudspeth. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

[91] Julian Paul Keenan, Hanna Oh, and Franco Amati. An overview of self-awareness. *The Oxford Handbook of Social Neuroscience*, page 314, 2011.

[92] James M Kilner, Karl J Friston, and Chris D Frith. Predictive coding: an account of the mirror neuron system. *Cognitive processing*, 8(3):159–166, 2007.

[93] Jaegwon Kim. *Physicalism, or something near enough*, volume 19. Princeton University Press, 2007.

[94] Tilo TJ Kircher, Carl Senior, Mary L Phillips, Sophia Rabe-Hesketh, Philip J Benson, Edward T Bullmore, Mick Brammer, Andrew Simmons, Mathias Bartels, and Anthony S David. Recognizing one's own face. *Cognition*, 78(1):B1–B15, 2001.

[95] Stanley B Klein. The cognitive neuroscience of knowing one's self. 2004.

[96] Christof Koch. The quest for consciousness a neurobiological approach. 2004.

[97] Konrad P Körding and Daniel M Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247, 2004.

[98] Uriah Kriegel. Consciousness as sensory quality and as implicit self-awareness. *Phenomenology and the Cognitive Sciences*, 2(1):1–26, 2003.

[99] Michael Kubovy, Alex O Holcombe, and Johan Wagemans. On the lawfulness of grouping by proximity. *Cognitive psychology*, 35(1):71–98, 1998.

[100] Yasuo Kuniyoshi. Learning from examples: Imitation learning and emerging cognition. *Humanoid Robotics and Neuroscience: Science, Engineering and Society, G. Cheng, Ed.(CRC Press, Boca Raton, FL, 2015)*, pages 223–250, 2015.

[101] AD Kuo and Zajac FE. Human standing posture: multi-joint movement strategies based on biomechanical constraints. *Progress in brain research*, 97:349–358, 1993.

[102] Jason J Kutch and Francisco J Valero-Cuevas. Challenges and new approaches to proving the existence of muscle synergies of neural origin. *PLoS computational biology*, 8(5), 2012.

[103] Christopher M Laine, Akira Nagamori, and Francisco J Valero-Cuevas. The dynamics of voluntary force production in afferented muscle influence involuntary tremor. *Frontiers in computational neuroscience*, 10:86, 2016.

[104] Christopher Landauer and Kirstie L Bellman. Self-modeling systems. In *International Workshop on Self-Adaptive Software*, pages 238–256. Springer, 2001.

[105] Mark R Leary and Nicole R Buttermore. The evolution of the human self: Tracing the natural history of self-awareness. *Journal for the Theory of Social Behaviour*, 33(4):365–404, 2003.

[106] Michel A Lemay and Patrick E Crago. A dynamic model for simulating movements of the elbow, forearm, and wrist. *Journal of biomechanics*, 29(10):1319–1330, 1996.

[107] Joseph Levine. Materialism and qualia: The explanatory gap. *Pacific philosophical quarterly*, 64(4):354–361, 1983.

[108] JP Lewis. Fast normalized cross-correlation, industrial light and magic. *unpublished*, 2005.

[109] Peter R Lewis, Arjun Chandra, Funmilade Faniyi, Kyrre Glette, Tao Chen, Rami Bahsoon, Jim Torresen, and Xin Yao. Architectural aspects of self-aware and self-expressive computing systems: From psychology to engineering. *Computer*, 48(8):62–70, 2015.

[110] Peter R Lewis, Arjun Chandra, Shaun Parsons, Edward Robinson, Kyrre Glette, Rami Bahsoon, Jim Torresen, and Xin Yao. A survey of self-awareness and its application in computing systems. In *2011 Fifth IEEE Conference on Self-Adaptive and Self-Organizing Systems Workshops*, pages 102–107. IEEE, 2011.

[111] Rodolfo Riascos Llinás. *I of the vortex: From neurons to self*, volume 50. MIT press Cambridge, MA, 2001.

[112] GE Loeb, WS Levine, and Jiping He. Understanding sensorimotor feedback through optimal control. In *Cold Spring Harbor symposia on quantitative biology*, volume 55, pages 791–803. Cold Spring Harbor Laboratory Press, 1990.

[113] Gerald E Loeb and Jeremy A Fishel. Bayesian action&perception: Representing the world in the brain. *Frontiers in neuroscience*, 8:341, 2014.

[114] HC Lou, JP Changeux, and A Rosenstand. Towards a cognitive neuroscience of self-awareness. *Neuroscience & Biobehavioral Reviews*, 83:765–773, 2017.

[115] Pete Mandik. Action-oriented representation. *Cognition and the brain: The philosophy and neuroscience movement*, pages 284–305, 2005.

[116] Ali Marjaninejad, Darío Urbina-Meléndez, Brian A Cohn, and Francisco J Valero-Cuevas. Autonomous functional movements in a tendon-driven limb via limited experience. *Nature machine intelligence*, 1(3):144–154, 2019.

[117] Ruben Martinez-Cantin, Manuel Lopes, and Luis Montesano. Body schema acquisition through active learning. In *2010 IEEE international conference on robotics and automation*, pages 1860–1866. IEEE, 2010.

[118] Victoria McGeer. Autistic self-awareness. *Philosophy, Psychiatry, & Psychology*, 11(3):235–251, 2004.

[119] R Christopher Miall. Connecting mirror neurons and forward models. *Neuroreport*, 14(17):2135–2137, 2003.

[120] RC Miall, D Jo Weir, Daniel M Wolpert, and JF Stein. Is the cerebellum a smith predictor? *Journal of motor behavior*, 25(3):203–216, 1993.

[121] Milana P Mileusnic, Ian E Brown, Ning Lan, and Gerald E Loeb. Mathematical models of proprioceptors: I. control and transduction in the muscle spindle. *Journal of neurophysiology*, 2006.

[122] Milana P Mileusnic and Gerald E Loeb. Mathematical models of proprioceptors: Ii. structure and function of the golgi tendon organ. *Journal of Neurophysiology*, 2006.

[123] David Milner and Mel Goodale. *The visual brain in action*, volume 27. OUP Oxford, 2006.

[124] Melanie Mitchell. Self-awareness and control in decentralized systems. In *AAAI Spring Symposium: Metacognition in Computation*, pages 80–85, 2005.

[125] Alain Morin. Levels of consciousness and self-awareness: A comparison and integration of various neurocognitive views. *Consciousness and cognition*, 15(2):358–371, 2006.

[126] Alain Morin. Self-recognition, theory-of-mind, and self-awareness: What side are you on? *Laterality*, 16(3):367–383, 2011.

[127] Thomas Nagel. What is the mind-body problem. *Experimental and theoretical studies of consciousness*, pages 1–7, 1993.

[128] U Neisser. Cognition and reality, 1976.

[129] Ulric Neisser. The roots of self-knowledge: Perceiving self, it, and thou a. *Annals of the New York Academy of Sciences*, 818(1):19–33, 1997.

[130] Chuanxin M Niu, Kian Jalaleddini, Won Joon Sohn, John Rocamora, Terence D Sanger, and Francisco J Valero-Cuevas. Neuromorphic meets neuromechanics, part i: the methodology and implementation. *Journal of neural engineering*, 14(2):025001, 2017.

[131] Alva Noë, Alva Noë, et al. *Action in perception*. MIT press, 2004.

[132] Georg Northoff, Alexander Heinzel, Moritz De Greck, Felix Bermpohl, Henrik Dobrowolny, and Jaak Panksepp. Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, 31(1):440–457, 2006.

[133] Elizaveta V Okorokova, James M Goodman, Nicholas G Hatsopoulos, and Sliman J Bensmaia. Decoding hand kinematics from population responses in sensorimotor cortex during grasping. *arXiv preprint arXiv:1904.03531*, 2019.

[134] J Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5):939–973, 2001.

[135] David J Ostry and Paul L Gribble. Sensory plasticity in human motor learning. *Trends in neurosciences*, 39(2):114–123, 2016.

[136] D Oyserman, K Elmore, and GS Smith. Self, self-concept, and identity-handbook of self and identity, 2012.

[137] David I Perrett, Mark H Harries, Ruth Bevan, S Thomas, PJ Benson, Amanda J Mistlin, Andrew J Chitty, Jari K Hietanen, and JE Ortega. Frameworks of analysis for the neural representation of animate objects and actions. *Journal of experimental Biology*, 146(1):87–113, 1989.

[138] Carissa L Philippi, Justin S Feinstein, Sahib S Khalsa, Antonio Damasio, Daniel Tranel, Gregory Landini, Kenneth Williford, and David Rudrauf. Preserved self-awareness following extensive bilateral brain damage to the insula, anterior cingulate, and medial prefrontal cortices. *PloS one*, 7(8):e38413, 2012.

[139] Jean Piaget. Piaget's theory of cognitive development. *Childhood cognitive development: The essential readings*, 2:33–47, 2000.

[140] Steven M Platek, Jaime W Thomson, and Gordon G Gallup Jr. Cross-modal self-recognition: The role of visual, auditory, and olfactory primes. *Consciousness and cognition*, 13(1):197–210, 2004.

[141] Joshua M Plotnik, Frans BM De Waal, and Diana Reiss. Self-recognition in an asian elephant. *Proceedings of the National Academy of Sciences*, 103(45):17053–17057, 2006.

[142] Helmut Prior, Ariane Schwarz, and Onur Güntürkün. Mirror-induced behavior in the magpie (pica pica): evidence of self-recognition. *PLoS biology*, 6(8), 2008.

[143] Jennifer S Rabin, Asaf Gilboa, Donald T Stuss, Raymond A Mar, and R Shayna Rosenbaum. Common and unique neural correlates of autobiographical memory and theory of mind. *Journal of Cognitive Neuroscience*, 22(6):1095–1111, 2010.

[144] Samik Raychaudhuri. Introduction to monte carlo simulation. In *2008 Winter simulation conference*, pages 91–100. IEEE, 2008.

[145] Diana Reiss and Lori Marino. Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Sciences*, 98(10):5937–5942, 2001.

[146] Magnus JE Richardson and Tamar Flash. Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis. *Journal of neuroscience*, 22(18):8201–8211, 2002.

[147] Bernhard Rinner, Lukas Esterle, Jennifer Simonjan, Georg Nebehay, Roman Pflugfelder, Gustavo Fernandez Dominguez, and Peter R Lewis. Self-aware and self-expressive camera networks. *Computer*, 48(7):21–28, 2015.

[148] Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annu. Rev. Neurosci.*, 27:169–192, 2004.

[149] Giacomo Rizzolatti, Luciano Fadiga, Vittorio Gallese, and Leonardo Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2):131–141, 1996.

[150] Philip Robbins. Consciousness and the social mind. *Cognitive Systems Research*, 9(1-2):15–23, 2008.

[151] Philippe Rochat. Five levels of self-awareness as they unfold early in life. *Consciousness and cognition*, 12(4):717–731, 2003.

[152] Udaya Bhaskar Rongala, Anton Spanne, Alberto Mazzoni, Fredrik Bengtsson, Calogero Oddo, and Henrik Jörntell. Intracellular dynamics in cuneate nucleus neurons support self-stabilizing learning of generalizable tactile representations. *Frontiers in cellular neuroscience*, 12:210, 2018.

[153] Perrine Ruby and Jean Decety. Effect of subjective perspective taking during simulation of action: a pet investigation of agency. *Nature neuroscience*, 4(5):546–550, 2001.

[154] Hannes P Saal, Benoit P Delhaye, Brandon C Rayhaun, and Sliman J Bensmaia. Simulating tactile signals from the whole hand with millisecond precision. *Proceedings of the National Academy of Sciences*, 114(28):E5693–E5702, 2017.

[155] Louis A Sass and Josef Parnas. Schizophrenia, consciousness, and the self. *Schizophrenia bulletin*, 29(3):427–444, 2003.

[156] R Keith Sawyer. Emergence in psychology: Lessons from the history of non-reductionist science. *Human development*, 45(1):2–28, 2002.

[157] Daniel L Schacter, Donna Rose Addis, and Randy L Buckner. Remembering the past to imagine the future: the prospective brain. *Nature reviews neuroscience*, 8(9):657–661, 2007.

[158] Gerald E Schneider. Two visual systems. *Science*, 1969.

[159] Lorenzo Sciavicco and Bruno Siciliano. *Modelling and control of robot manipulators*. Springer Science & Business Media, 2012.

[160] William Seager. *Theories of consciousness: An introduction and assessment*. Routledge, 2016.

[161] Anil K Seth, Keisuke Suzuki, and Hugo D Critchley. An interoceptive predictive coding model of conscious presence. *Frontiers in psychology*, 2:395, 2012.

[162] Reza Shadmehr, Maurice A Smith, and John W Krakauer. Error correction, sensory prediction, and adaptation in motor control. *Annual review of neuroscience*, 33:89–108, 2010.

[163] Dan Song, N Lan, GE Loeb, and J Gordon. Model-based sensorimotor integration for multi-joint control: development of a virtual arm model. *Annals of biomedical engineering*, 36(6):1033–1048, 2008.

[164] Henry P Stapp. Quantum theory and the role of mind in nature. *Foundations of Physics*, 31(10):1465–1499, 2001.

[165] Luc Steels and Michael Spranger. The robot in the mirror. *Connection Science*, 20(4):337–358, 2008.

[166] Fabio Sterpetti. Models, brains, and scientific realism. In *Model-Based Reasoning in Science and Technology*, pages 639–661. Springer, 2016.

[167] Donald T Stuss. Self, awareness, and the frontal lobes: A neuropsychological perspective. In *The self: Interdisciplinary approaches*, pages 255–278. Springer, 1991.

[168] Susan D Suárez and Gordon G Gallup Jr. Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of human evolution*, 10(2):175–188, 1981.

[169] Keisuke Suzuki, Sarah N Garfinkel, Hugo D Critchley, and Anil K Seth. Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia*, 51(13):2909–2917, 2013.

[170] Gregg A Tabot, John F Dammann, Joshua A Berg, Francesco V Tenore, Jessica L Boback, R Jacob Vogelstein, and Sliman J Bensmaia. Restoring the sense of touch with a prosthetic hand through a brain interface. *Proceedings of the National Academy of Sciences*, 110(45):18279–18284, 2013.

[171] Evan Thompson. Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and the cognitive sciences*, 4(4):407–427, 2005.

[172] Neil C Thompson, Kristjan Greenewald, Keeheon Lee, and Gabriel F Manso. The computational limits of deep learning. *arXiv preprint arXiv:2007.05558*, 2020.

[173] Sebastian Thrun. -a lifelong learning perspective for mobile robot control. In *Intelligent Robots and Systems*, pages 201–214. Elsevier, 1995.

[174] Giulio Tononi. Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, 215(3):216–242, 2008.

[175] Colwyn B Trevarthen. Two mechanisms of vision in primates. *Psychologische Forschung*, 31(4):299–337, 1968.

[176] Francisco J Valero-Cuevas. An integrative approach to the biomechanical function and neuromuscular control of the fingers. *Journal of biomechanics*, 38(4):673–684, 2005.

[177] Francisco J Valero-Cuevas. A mathematical approach to the mechanical capabilities of limbs and fingers. In *Progress in motor control*, pages 619–633. Springer, 2009.

[178] Francisco J Valero-Cuevas. *Fundamentals of neuromechanics*. Springer, 2016.

[179] Francisco J Valero-Cuevas, BA Cohn, HF Yngvason, and Emily L Lawrence. Exploring the high-dimensional structure of muscle redundancy via subject-specific and generic musculoskeletal models. *Journal of biomechanics*, 48(11):2887–2896, 2015.

[180] Francisco J Valero-Cuevas, Heiko Hoffmann, Manish U Kurse, Jason J Kutch, and Evangelos A Theodorou. Computational models for neuromuscular function. *IEEE reviews in biomedical engineering*, 2:110, 2009.

[181] Francisco J Valero-Cuevas and Marco Santello. On neuromechanical approaches for the study of biological and robotic grasp and manipulation. *Journal of neuroengineering and rehabilitation*, 14(1):101, 2017.

[182] Susanne J van Veluw and Steven A Chance. Differentiating between self and others: an ale meta-analysis of fmri studies of self-recognition and theory of mind. *Brain imaging and behavior*, 8(1):24–38, 2014.

[183] HEJ Veeger, FCT Van Der Helm, LHV Van Der Woude, GM Pronk, and RH Rozendal. Inertia and muscle contraction parameters for musculoskeletal modelling of the shoulder mechanism. *Journal of biomechanics*, 24(7):615–629, 1991.

[184] Kai Vogeley, Patrick Bussfeld, Albert Newen, Sylvie Herrmann, Francesca Happé, Peter Falkai, Wolfgang Maier, Nadim J Shah, Gereon R Fink, and Karl Zilles. Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage*, 14(1):170–181, 2001.

[185] Stephen Warshall. A theorem on boolean matrices. *Journal of the ACM (JACM)*, 9(1):11–12, 1962.

[186] Max Wertheimer. Laws of organization in perceptual forms. *A source book of Gestalt Psychology*, 1923.

[187] Justin HG Williams, Andrew Whiten, Thomas Suddendorf, and David I Perrett. Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews*, 25(4):287–295, 2001.

[188] Frank R Wilson. *The hand: How its use shapes the brain, language, and human culture.* Vintage, 1999.

[189] Daniel M Wolpert, R Chris Miall, and Mitsuo Kawato. Internal models in the cerebellum. *Trends in cognitive sciences*, 2(9):338–347, 1998.

[190] Yuichiro Yoshikawa, Koh Hosoda, and Minoru Asada. Does the invariance in multi-modalities represent the body scheme?-a case study with vision and proprioception. In *2nd Intelligent Symposium on Adaptive Motion of Animals and Machines*. Citeseer, 2003.

[191] Dan Zahavi and Andreas Roepstorff. Faces and ascriptions: Mapping measures of the self. *Consciousness and cognition*, 20(1):141–148, 2011.

[192] Felix E Zajac. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Critical reviews in biomedical engineering*, 17(4):359–411, 1989.

[193] Shagufta Zia, Frederick WJ Cody, and Donald J O'Boyle. Identification of unilateral elbow-joint position is impaired by parkinson's disease. *Clinical Anatomy*, 15(1):23–31, 2002.