2-2010

A Behavioral Approach to Human-Robot Communication

Shichao Ou
University of Massachusetts Amherst, chao@cs.umass.edu

A BEHAVIORAL APPROACH TO HUMAN-ROBOT COMMUNICATION

A Dissertation Presented

by

SHICHAO OU

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2010

Computer Science

A BEHAVIORAL APPROACH TO HUMAN-ROBOT COMMUNICATION

A Dissertation Presented

by

SHICHAO OU

Approved as to style and content by:

Roderic Grupen, Chair

Andrew Barto, Member

Allen Hanson, Member

Rachel Keen, Member

Andrew Barto, Department Chair
Computer Science

To my grandfather, Yuanren Ou.

ACKNOWLEDGMENTS

I owe my gratitude to many people as this dissertation would not have been possible without their generous help. First and foremost, I am heartily thankful to my advisor, Professor Roderic Grupen, whose encouragement and guidance from the inception of this thesis to the final stage enabled me to develop a in-depth understanding of the subject. Without his support and vision, the journey would have been much more arduous and most likely not as fruitful as it turned out to be. Rod's warm personality, academic excellence attracted many people who shared similar interest as I, and created an encouraging atmosphere for us to work collaboratively. Without this atmosphere and the collaboration, it would be impossible to attempt a dissertation of such breadth and gravity. For this, I feel extremely blessed because in the end it has led to something unique that I can be proud of. More importantly, based on the work done in this dissertation, we can foresee many new directions in which this topic can be further explored.

I am very grateful to the members of my thesis committee—Professor Andy Barto, Allen Hanson, and Rachel Keen. I could not have asked for a better committee. They each graciously lent me their expertise and wisdom in their specific areas—Andy in intrinsically motivated learning, Al in computer vision and Rachel in developmental psychology. It is this perfect combination that enabled me to see the problem in new ways that are previously not explored. The many encouraging and inspirational discussions significantly contributed to the formation of this thesis.

I would like to thank the many friends and colleagues who helped making this project a reality. A special thanks to Stephen Hart, whose work became the foundation of my research. Working with Steve has saved me a great deal of time and

effort as his dedication to excellence has produced highly robust behavioral programs for robots that made the learning process in my experiments relatively a breeze. I would also like to thank Shiraj Sen for his selflessness as countless times he set aside his own work to lent a hand helping both Steve and I to advance our research. Many thanks go to the members of the Laboratory for Perceptual Robotics for the discussions and collaborations—specifically, Rob Platt, Patrick Deegan, John Sweeney, Emily Horrell, Bryan Thibodeau, Dirk Ruiken, Dan Xie, Scott Kuindersma, Yun Lin, and Grant Sherrick. I would like to thank Marwan Mattar of the computer vision lab for helping me understand the latest statistical vision techniques, and also Dr. Shaowu Peng for introducing me to a useful mathematical framework for representing objects hierarchically in terms of smaller components. My sincere gratitude goes to the anonymous subjects who took part in the human-robot interaction studies described in this thesis. Without their patience and active participation, I would not be able to collect the data needed to complete this dissertation. Leeanne Leclerc and Priscilla Scott deserve special acclimation because they ensured all of the mundane administrative details got taken care of such that I can simply focus on getting my research done.

I would like to thank my wife, Sherry Li, for her love, her kindness, and her patience—that she had to put up with my constant departures from home for the past three years as I worked towards completing this degree in a different state. I would also like to thank my parents, Yinggang and Liancheng for their love, support and advice that helped me get through some of the low points of my graduate career. Finally, with great pride I dedicate this thesis to my late grandfather, whose wisdom inspired us, his children and grandchildren, to develop an unquenchable thirst for knowledge. With this thesis, I hope I have made you, my grandfather proud. I wish you were here to share my joy and happiness.

ABSTRACT

A BEHAVIORAL APPROACH TO HUMAN-ROBOT COMMUNICATION

FEBRUARY 2010

SHICHAO OU

B.S., SOUTH CHINA UNIVERSITY OF TECHNOLOGY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Roderic Grupen

Robots are increasingly capable of co-existing with human beings in the places where we live and work. I believe, however, for robots to collaborate and assist human beings in their daily lives, new methods are required for enhancing human-robot communication. In this dissertation, I focus on how a robot can acquire and refine expressive and receptive communication skills with human beings. I hypothesize that communication has its roots in motor behavior and present an approach that is unique in the following aspects: (1) representations of humans and the skills for interacting with them are learned in the same way as the robot learns to interact with other "objects," (2) expressive behavior naturally emerges as the result of the robot discovering new utility in existing manual behavior in a social context, and (3) symmetry in communicative behavior can be exploited to bootstrap the learning of receptive behavior.

Experiments have been designed to evaluate the approach: (1) as a computational framework for learning increasingly comprehensive models and behavior for communicating with human beings and, (2) from a human-robot interaction perspective that can adapt to a variety of human behavior. Results from these studies illustrate that the robot successfully acquired a variety of expressive pointing gestures using multiple limbs and eye gaze, and the perceptual skills with which to recognize and respond to similar gestures from humans. Due to variations in human reactions over the training subjects, the robot developed a preference for certain gestures over others. These results support the experimental hypotheses and offer insights for extensions of the computation framework and experimental designs for future studies.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Motivation

In recent years, there has been an increased demand for personal robots in aerospace, elder-care and medical applications (Figure 1.1). For robots to operate and assist humans in such a variety of environments, they must possess the ability to convey their intentions to human partners and infer human intentions from their actions as well.

Figure 1.1. Personal robots are being considered for new roles in aerospace, elder-care and medical applications. Effective communication is essential for achieving successful collaboration in these scenarios.

This thesis addresses how a robot can acquire and refine communication skills through daily interactions with humans. The main focus is the development of behavioral communication skills—gestures—rather than verbal ones. This distinction is relevant and critical since it is my hypothesis that communicative skills convey intentions and that intentions derive from behavior. It therefore follows that all forms of communication have their roots in sensorimotor behavior. The hope is that through studying the simpler problem of non-verbal communication, a grounded and scalable approach can be developed that may extend to more expressive verbal communication, or at least shed some light on how it can be tackled. The psychology literature [35] suggests that gesture and language are highly related, since in the human brain, regions that handle these functions share common neurological pathways.

Communication has both expressive and receptive dimensions. On the expressive side, current state-of-the-art approaches [8, 76, 23] often advocate for pre-programmed communicative behavior emulating or mimicking important human social behavior such as gaze direction, pointing, nodding, and beckoning. On the receptive side, independent sensory modules are often proposed for the detection and recognition of humans and human behavior. Impressive human-robot interaction dialogs using these behaviors have been demonstrated [113, 77, 37]. However, these approaches do not speak to the origins of such behavior, nor do they carry "meaning." This thesis opts for an approach that studies the origins of communicative behavior and how some commonly understood gestures can arise naturally from interactions with humans in the environment, without explicit third-party programming.

I advocate a learning approach because human gestures are dynamic. Even the simplest gesture can take on many forms and the same motion can possess a variety of meanings under different contexts or cultures. For instance, the hand waving movement can mean "hello," "good bye," or even "no," depending on the time and the context under which the movement is performed. While in some cultures pointing

with the index finger is generally acceptable behavior, in others this action is often considered offensive [75]. For robots, due to the different physical appearance and morphology (as seen in Figure 1.1), simple mimicry of human gestures may not be the most effective way for a robot to communicate. For instance, it is reasonable to assume that different robots will employ different means of indicating directions and target positions. Sometimes, several versions of the gesture may be needed to convey a given intention effectively in different contexts. A learning approach is useful in this case because it enables the robot to acquire new communicative actions as the need arises and to adapt communicative behavior to meet the context and the changing needs of a communicative partner. Furthermore, a learning approach also has the potential for specializing gestures to different tasks and populations.

For robots to develop expressive and receptive communicative skills autonomously in the course of natural interactions with humans, a number of important questions need to be addressed.

1. Under what conditions will communicative behavior naturally arise and how can these conditions be maintained?

2. What are the action primitives? What states and actions represent communicative and behavior, respectively?

3. How are expressive and receptive behavior related and how do they interact?

The next section presents an overview of the approach to these issues.

1.2 Approach

For robots to develop communicative behavior effectively in the course of natural interactions with humans, the conditions underlying "stable" human-robot dyads must first be established and maintained. I hypothesize that stable dyads are formed

when the individual agents are underactuated and mutually rewarded. *Underactuation* specifies that there exist tasks that are achievable by a human-robot team that neither agent alone can achieve. Objects can be too heavy, or objects can be unreachable by some agents and reachable by others. *Mutual reward* conditions require that each agent in a human-robot team be rewarded for participating constructively in a dyadic relationship, although the rewards can be different for different agents. For example, in the case where the object is too heavy for either the robot or the human to lift alone, when the robot conveys the intention to lift the object and the human chooses to help, the robot is rewarded for lifting the object and the human receives a sympathetic reward for successfully helping the robot to achieve its goal.

Predicated on these conditions for fostering communication, this thesis presents a learning framework (Chapter 3) for developing expressive communicative behavior for engaging a human's assistance, as well as recognizing the intentions of the human partner and acting to reciprocate the gesture. There are three distinctions between the learning approach adopted here and related work in human-robot interaction (HRI): (1) expressive communicative actions are learned in conjunction with manual skills using the same framework, (2) models of humans and skills for interacting with them are learned in the same way as the robot models and learns to interact with other "objects," and (3) expressive and receptive communicative behavior share knowledge structures and therefore expressive behavior can be used for interpreting intentions of others and thus the receptive learning process benefits as a result.

This work argues for an approach to develop communicative behavior in conjunction with manual skills because I believe gesture has its roots in manipulation behavior. It has been suggested in the psychology and neuroscience literature that for humans, the development of manual, gestural and language skills are highly interrelated [38, 7, 35, 28]. However, in the field of robotics and AI, most research considers these problems separately. This work argues that manipulation and communicative

behavior share many important learning issues. The development of communicative behavior benefits from a learning framework with support for hierarchy, generalization, and knowledge transfer, as do other forms of sensorimotor behavior. I will use the control basis framework developed in the Laboratory for Perceptual Robotics at UMass over the past several years. Using this framework, I believe it is possible for robots to learn to reuse manipulation behavior for the purpose of communication and knowledge supporting manipulation behavior can likewise be applied to convey information. These properties make this approach efficient for learning and therefore well-suited for human-robot interaction where training occurs in real-time.

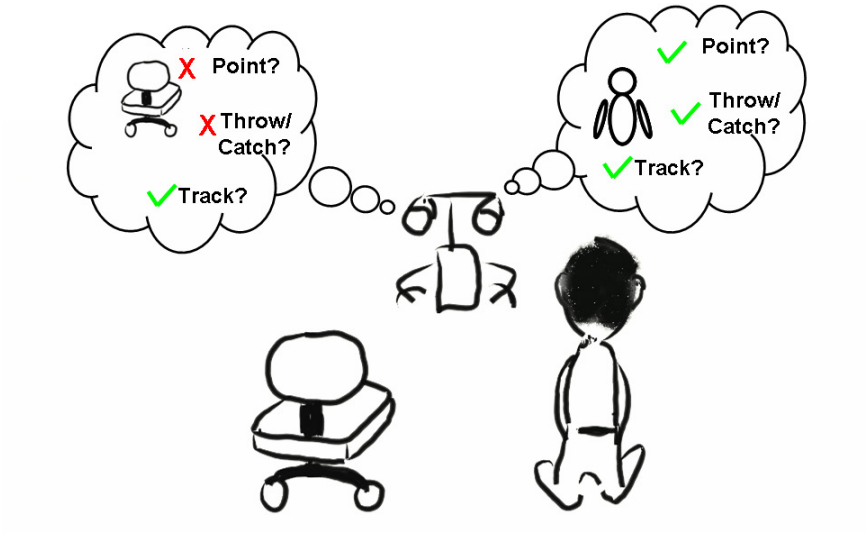


Figure 1.2. Humans and objects are modeled in this work as behavioral affordances. At run-time, the robot differentiates objects using not only visual features, but also known behavioral responses. For instance, humans afford tracking, respond to pointing, and are likely to play "throw and catch" with the robot. Comparatively, a chair is much less responsive and affords only tracking.

As a natural outcome of combining manual and communicative behavior learning, the affordances of human beings are learned in the same way as affordances of objects in the environment (Figure 1.2). In essence, the proposed learning framework is an affordance-based modeling approach that subscribes to the Gibsonian view [34] that

our perception and understanding of the world is stored and applied in terms of the behavior that the environment affords. Similarly, exploratory interactions with humans provide information regarding how behavior is afforded by humans and therefore are stored as a collection of affordances rather than conventional visual appearance. The models of the affordances of human beings are learned and enriched over time as the robot's means of interaction grows. This is in contrast to research that uses hand-coded perception and social behavior. This approach observes sensory invariants of the human social partners to support recognition and inform strategies for interaction. Humans are special objects with complicated kinematic structure, independent motion, whose appearance changes day to day. In this work however, I hypothesize that human behavior, though dynamic and varied, under the social context of *underactuation* and *mutual reward* is relatively more predictable than human visual appearance and therefore can lead to informative models of social behavior.

Receptive behavior, on its surface, seems to require insight into the state of mind and goals of the expressive communicate partner [96], while expressive behavior can be viewed as a direct extension of goal-oriented manual behavior. This has lead to challenges regarding uniform methods for learning. The approach explored in this dissertation takes a decidedly different tack. I take the position that there exists symmetry between expressive and receptive behavior, and therefore receptive social behavior can benefit from the knowledge gained from the expressive gesture learning process. This approach is consistent with recent observations of *mirror neurons* [89] from the psychology and neuroscience literature where it is found that the same neurological pathways responsible for generating actions also participate in recognizing intentions from another agent. Similarly, empirical studies in recent years have also led developmental psychologists [29, 118] to conclude that infants' perception of others' actions is influenced by their own goal-directed action capabilities. In my work, existing behavioral programs are used for robots to parse events and interpret actions

performed by humans. This is made possible by the shared knowledge structure due to the use of a consistent behavioral learning framework.

Prevailing approaches in the field [50, 68, 11] generally treat gesture recognition as a motion capture recognition problem where human motion observations are matched against motor templates derived from demonstration. These techniques rely on high dimensional motion capture data to achieve reasonable matching performance. As a result, the computational complexity is high and therefore matching is generally performed as an offline process. When much noisier and sparser vision data are used, the performance also degrades dramatically. Furthermore, under a constraint context where the human employs alternative gestures to convey the same information, the motion trajectory may be significantly different. In these approaches, all behavior is represented in terms of time-series of Cartesian postural data. This is at best a geometric simulation of human motor activity and does not reflect insight into shared meaning between the human and the robot. It has been suggested that research is lacking on extracting abstract conceptual/intentional motives from observed demonstrations [97]. This limits the generality of approaches to date.

This work is takes a simpler approach that is inspired by the *teleological stance* from Gergely [33], who suggested that one-year old infants extract "goals", "means" and "constraints" to interpret the behavior of others. In my work, I propose a mechanism for exploiting the symmetry between expressive and receptive behavior. This is achieved by allowing the robot to create off-policy monitors and attach them to existing behavior to find auxiliary models that correlate with rewarding events. I believe these models can be used as cues for interpreting human motion. As we will see from examples in later chapters, this approach eases computational overhead because it allows the robot to abstract sequential events rather than raw motion trajectory data.

Figure 1.3 provides a road map for communicative behavior learning: under social conditions of underactuation and mutual reward expressive behavior emerges as a continuation of manual skill learning. Robots can thus discover that human can be recruited as external resources if the right action is performed. Receptive behavior learning benefits from expressive programs learned since they are used as blueprints for recognizing the same gesture coming from the human and from this a reciprocal assistive behavior can be explored and learned. During the course of communicative learning, interactions with humans also provide the robot with opportunities to build increasingly refined models of humans in the form of control circuits for behavioral affordances of human beings. Although beyond of the scope of this dissertation, it is conceivable that effective communication can lead to guidance from humans for the robots to learn more complicated manual skills, thus completing the cycle.

Experiments have been designed to demonstrate the feasibility of the approach. Each experiment consists of a number of stages that involves a bi-manual humanoid robot and a human partner. Subjects of convenience participate in each stage of the study and interact with the robot, some for training and some for evaluation, one person at a time.

Experiment 1 demonstrates how the framework enables the robot to reuse existing manual behavior for establishing an increasingly complex affordance model of humans, in a series of learning stages. From a simple initial concept that a human is a motion segment of a certain size that affords visual tracking, it evolves into a more complex model that a human-scale segment also contains multiple kinematically connected parts that afford simultaneous tracking.

Experiment 2 places the robot in a situation where a desired object is out of reach and the robot must recruit a nearby human and direct him to the object to help accomplish its goal. Experimental design establishes plausible conditions of underactuation and mutual reward and seeks to evaluate how well the robot can solicit

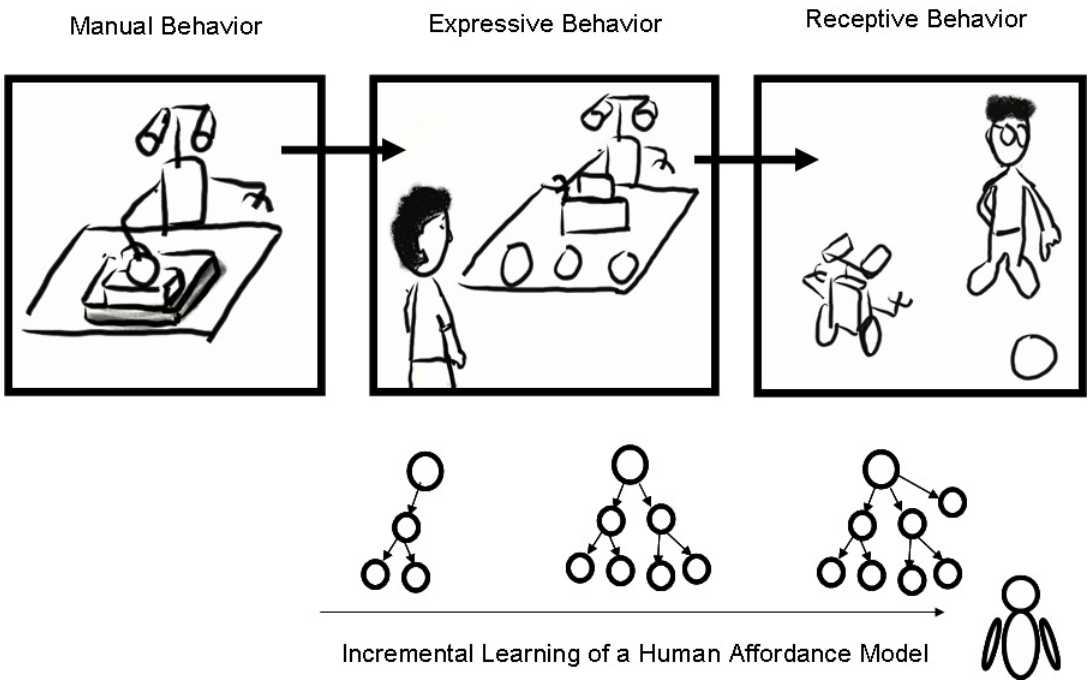


Figure 1.3. A road map for learning communication skills. To learn expressive behavior, the robot can first acquire manual skills through intrinsically motivated learning. These skills under the appropriate context naturally give rise to expressive behavior. By exploiting the symmetry in communicative behavior, receptive skill can also be learned. The bottom of the figure illustrates using the same framework, the robot can also incrementally build knowledge structure of human beings the interaction continues.

appropriate human assistance. Subjects are not familiar with the goal of the project and are not instructed as to which object the robot wishes to obtain. Results are promising as the interactions produced two different communicative behaviors, both of which clearly exhibited significantly better performance than a baseline where the human has to make a random guess. results we can extrapolate how other gestures, such as size-hinting, beckoning and rejection, can all arise naturally using the same approach. Also in this stage, as the robot learns more behaviors while interacting with humans, these behaviors can also be incorporated into the expanding knowledge tree as part of a hierarchical affordance model of humans.

Experiment 3 places the robot in a reciprocal setup to experiment 2, where the objects are now placed out of reach of the human and are instead reachable by the robot. With the added ability to track movements of different parts of the human body, including the arms, I hypothesize the robot can reuse knowledge gained from expressive behavior and use them as templates for recognizing the same behavior when performed by the human. Results from interactions with subjects of convenience confirm that the robot is able to recognize various pointing gestures exhibited by different people and learn the appropriate behavior for assistance.

1.3 Contributions

The contributions of this dissertation are the following:

1. A unique approach for robots to learn about humans, where a human is modeled as a collection of behavioral affordances the robot discovers from its experience. It builds conditions under which a robot can make progress toward increasingly sophisticated models of humans over time. Experiments demonstrate a series of stages where the robot learns a kinematic model of the human body that affords reliable tracking and later learns to include affordances for collaborative behavior as the human-robot team negotiates strategies for collaborating to achieve a common goal.

2. The extension of a behavioral learning framework intended for developing manual skills to the domain of learning communicative behavior. An algorithm is presented to enhance the framework's ability to adapt to new contexts while maintaining much of the previous acquired knowledge structure. It is applied to enable robots learn expressive behavior. In addition, an approach is proposed to allow robots to exploit the symmetry in communicative behavior for the purpose of learning receptive behavior.

3. The proposal of a developmental trajectory for robots to acquire communication skills to interact with humans. This begins with the robot first learning various manual skills through intrinsically motivated exploration. Next, by subjecting the robot to conditions of mutual reward and underactuation, expressive communicative behavior emerges naturally as the robot discovers the utility of manual behavior under the new context. Finally, receptive behavior is learned by reusing existing manual skills and knowledge structure gathered during the expressive behavior learning process.

1.4 Chapter Organization

Chapter 2 offers a review of the psychology and the computer science literature to provide the theoretical background of the approach taken in this thesis. Chapter 3 describes the *control basis* framework for constructing multi-objective control circuits that will be useful for learning communicative behavior. Examples are presented to show how increasingly comprehensive behavior is learned as a humanoid robot explores control configurations that employ different sensorimotor resources. The remainder of the document focuses on individual components of the overall approach for robots to learn communicative behavior with humans, and elaborates on each separately.

Chapter 4 presents the affordance-based approach for modeling humans and demonstrates how a robot can build an increasingly comprehensive model of the affordances of humans from natural interactions. Chapter 5 presents a general algorithm for robots to "repair" existing learned programs by generating sub-goals and learning a new repair policy, and shows how this algorithm can be applied to enable a humanoid bimanual robot to learn expressive communicative behavior by using existing manual as the basis of learning. At the end of this chapter, the human affordance model is further extended as the robot discovers that humans respond to pointing gestures.

With a sufficiently comprehensive model of humans, Chapter 6 demonstrates how the robot can use previously learned behavioral programs for parsing and recognizing the same gesture from a human. Chapter 7 provides a discussion and conclusions of the work presented in this document.

CHAPTER 2

LITERATURE REVIEW

Research in several disciplines have influenced the proposed approach reported in this dissertation. Section 2.1 shows supporting evidence from the psychology and neuroscience literature for the inextricable connection between manual and communicative behavior during the development of a human infant. In Section 2.2, the theory of *mirror neurons* and the associative memory in the neocortex is reviewed as it motivates the computational model of memory advanced in this dissertation. My goal here is to form a unified model capable of both expressing behavior with explicit intention and recognizing intention in the behavior of others. Generalization and transfer are the key ideas proposed for transforming sensorimotor behavior into a gestural lexicon. My inspiration on this front comes once again from developmental psychologists. In Section 2.3, the *teleological stance* of György Gergely and its implications for identifying intention and the object of intentional actions are discussed. This thesis draws inspiration from the developmental processes of a human infant, and observes that infants learn in stages and through constant interaction with the world. Along these lines, Section 2.4 reviews research in developmental programming for robots and Section 2.5 discusses the Gibsonian notion of affordance and its application to knowledge organization and world modeling. Finally, Section 2.6 summarizes the important issues in human-robot interaction and the current approaches for tackling these problems. In particular, I focus on the prevailing methods with which robots build models for the detection and tracking of humans, and compare my work with these methods.

2.1 Manual Behavior and Communicative Gestures

Psychologists acknowledge a tight connection between communicative gesture and manual behavior. In the 1930s, Lev Vygotsky noted that "...initially, pointing is nothing more than an unsuccessful attempt to grasp something..." [116]. In this case, a manipulation behavior is described as the origin of the communicative pointing action. As infants attempt to reach for out-of-reach objects, even though they inevitably fail, in the presence of a caregiver, the action is recognized and interpreted as the "intention" to acquire the object and thus the action becomes a gesture. When infants become older, more sophisticated abstract gestural actions begin to emerge as an infant's manipulation skills continue to improve. For instance, it is common for infants to pretend to drink from an empty cup to indicate the desire for a drink. Later this often evolves to pantomiming without a cup as the infant's understanding of semantic meanings of actions improve [4].

Greenfield [38] hypothesized links between the origins of tool use and language, and also suggested that manipulation behavior for tool use may have played a causal role in the evolution of gestural communication. In both Bradshaw's [7] and Gibson's [35] books, it is noted that patients with apraxia who have difficulty in executing purposeful movements of the arm and hand and thus learning the use of tools, also have trouble performing pantomiming gestures to convey their intention. These studies provide evidence for the connection between communicative gesture and manipulation behavior.

Similar evidence also exists in the neuroscience literature. It is found that in most right-handed individuals, both the dominant hand and communication (including language) are controlled by the same neural circuitary in the left hemisphere, and vice-versa for the left-handed population [57]. Kimura concludes that the hemispherical co-location of language and the dominant hand strongly suggests a commonality of neural control for manipulative and communicative behavior. More recently, through

the use of functional magnetic resonance imaging (fMRI), Frey [28] observed activities in the same brain regions both when the subject performs manipulative tool-use actions and when the subject performs a related communicative gesture.

In a comparative study [35] of chimpanzee and human infant development, Gibson noted that despite the fact that both human and chimps possess potential tool-using and symbolic capabilities, the behavior of infant chimps and infant humans differs greatly in manipulative and communicative domains. From a very young age, human infants begin to engage in repetitive object manipulation behaviors such as grasping, shaking and kicking to recreate interesting "spectacles," while the chimpanzees did not. More importantly, by the second year, infants become more interested in object-object relationships while chimpanzees are only interested in single objects. From this evidence, it is suggested that the human infants' capacity to learn complex sequential actions involved in manipulation tasks and subsequent interest in object-object relationships allows humans to eventually develop complex systems of communication, including language, since sequencing behavior (utterances) to form more complicated ones, and associating the causal outcome of manual actions are the key to developing effective communication skills.

For this work, this insight is applied to robotics to show that it can lead a general-purpose computational framework to enable robots to learn gesture in a grounded manner. Importantly, I contend that these forms of communicative actions can be built into social behavior without first constructing a complex mental model of the human social partner—it relies only on discovering the causal relationships between "gesturer" and "gesturee." The "gesture" begins as a motor-artifact associated with a sensorimotor function and is recognized as a reliable means of causation. Ultimately, it is adapted for use as an effective means of communicating one's intentions, and is initiated and perhaps stylized to that purpose.

2.2 Mirror Neurons, Action Generation and Recognition

Mirror neurons [89] has been suggested as a possible neural basis underlying both action generation and predictions of other's behaviors and mental states [10]. These neurons show similar activity when a monkey observes the goal-directed action of another agent and when it carries out that action itself. This observation has led researchers to hypothesize that there exists a common coding between perceived and generated actions [86, 83, 37, 29, 118]. Therefore, these neurons may play an important role in processes used by humans and other animals to relate their own actions to actions of others.

Several research groups have attempted to create a computational account of the mirror neuron to enable robotic systems to learn from humans. Jenkins and Mataric [50, 68] implement an on-line encoding process that maps observed joint angles onto movement primitives. Thus a simulated upper-body humanoid can learn to recognize and imitate a sequence of arm trajectories. Others (Demiris and Hayes [20], Atkeson and Schaal [3]) have adapted the notion of mirror neurons to predictive forward models that can be used to classify the observed trajectories. However, Jenkins' approach relies on motion capture data and Atkeson demonstrates behavior by moving the robot directly. Neither method is suitable in the context of face-to-face human robot interaction. Breazeal's imitation learning work [10] on the other hand uses vision. In this case, through an imitation game where the robot randomly generates facial configurations through motor babbling, and the human imitates the robot's facial expressions, the robot gathers samples to train a neural network that maps between perceived human facial features to its own facial joint space. Thus, the robot Leonardo learns a generative model for facial expression recognition and generation.

Along the same lines, research on task oriented human-robot interaction has been attempted where a robot engages in a lengthy dialog with a human, playing games such "hide and seek" [113], "push the right button"[8] or "find object in boxes"

[9]. In these studies, the focus is on higher level issues such as perspective taking abilities or understanding visual occlusions. As a result, the idea of mirror neurons, although mentioned, is de-emphasized in implementation. For instance, although the robot recognizes the intention of the user by parsing observatins using templates or schemas, these schemas and templates are not the result of the action generation process, and are instead hand-crafted. Similarly, social behavior employed by the robot is also the result of programming.

Although the approach taken by this thesis also finds support in the theory of mirror neurons and treats the problems of generation and recognition of communicative behavior as a whole, it differs from the methods mentioned above in several important ways:

- While most work treats the action generation problem as a low-level motion trajectory mapping problem [68, 3] or a joint space motor control problem [67], this thesis advocates learning communicative behavior from a higher level using motor-primitives acquired during manual skill learning. The level of abstraction requires a framework that supports hierarchical learning and knowledge transfer and has the benefit of allowing the interplay between the robot and the human to be considered as part of the learning process.

- Rather than using mirror neuron analogies to focus on imitating human motion, a process that ultimately does not lend insight into the origins of communicative actions, this thesis takes the position that "purposeful action=communicative behavior" and attempts to ground communicative behavior (the exchange of useful information) using the same control primitives (actions) that support other motor skills.

- The recognition process proposed in this work differs from previous work as it does not attempt to identify intentions by matching entire motion trajectories,

but instead focuses on extracting and matching simpler cues as those proposed by Gergely in his *teleological stance*, which is the subject of the next discussion.

2.3 Teleological Stance for Recognition of Goal-directed Behavior

In the paper "What Should a Robot Learn From an Infant?" [33], Gergely argues that psychological research reveals that one-year old infants are able to attribute goals to actions and to evaluate the rationality of actions.

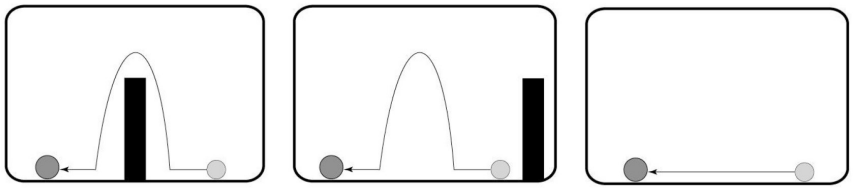


Figure 2.1. Gergely's animated apparatus for establishing that one-year-old infant infers goals and evaluates the rationality of actions [33].

To illustrate this, Gergely presented an example in which the infants were shown a computer-animated goal-directed action (as shown in the left figure in Figure 2.1). After the infants became familiar with this action (when their gazes began to shift away), they were shown two other animated situations illustrated in the middle, and the right panels of Figure 2.1. The results show that the infant's attention focused on the animation in the middle for longer period of time. A possible explanation is that the infant can infer that this is not a rational action as the obstacle was no longer in the way. In contrast, even though the action shown in the right is perceptually novel, it was an expected rational action for the case.

To explain these remarkable inferential feats for one-year-olds, Gergely proposes a non-mentalistic (reality-based) teleological action interpretational strategy called the "*teleological stance*." The *teleological stance* hypothesizes that infants perform

inference based on a teleological explanatory relation among 3 aspects of reality: the future state of reality in relation to the behavior (goal), the observed behavior (means), and relevant physical contexts that constrain possible actions (constraints).

Figure 2.2. The "magic box" study where an adult demonstrates how to actuate a light switch using an unusual head-bumping action, even though the hands are not occupied and see if the infant would imitate the action [33].

To illustrate this, an experiment called the "magic box" study (Figure 2.2) was conducted. An adult demonstrated how to actuate a light switch using an unusual head bumping action in front of the infant. For one group of infants, a constrained context was presented in which the adult's hands were occupied with a blanket (Figure 2.2 left), and for another group, her hands were clearly free (Figure 2.2 right). Results show that most infants in the first group (constrained context where the demonstrator's hands were occupied) did not imitate the action, because the condition that prohibited the demonstrator using her hands did not apply in their case. Therefore, they chose to use their hand to turn on the light instead. However, for the second group, most imitated the action because the hand constraints in this case did not exist for the demonstrator and therefore, Gergely contends, the infants concluded the choice of the demonstrator's head to actuate the switch must be the result of a rational decision. A possible interpretation of these experiments is that one-year-olds understand the goals, and were able to determine the rationality of the action based on the physical constraints of the current context. Moreover, these results also indi-

cate that the infants' recognition process relies on the end-state of the behavior as an important cue for recognizing the intention of the others. The trajectory of motion and in this case, even whether the same part of the body is used for bringing about the end state matters little.

Compared to the traditional views where a complex mental model of others are required, the teleological stance provides a simplier interpretation for one-year-old infant's ability to imitate the behavior observed in others. When this is applied to AI and robotics, simplicity translates to computational efficiency. In this thesis, this principle is applied to interpret gestures from a human and ultimately determine how to help. First, the robot learns an array of skills/programs in its own terms and masters these skills in a variety of run-time contexts. Then the robot can interpret events in the world through the prism of these skills, even those that it observes passively. To classify the behavior of a human, or any other agent, rather than matching the entire skill program state-by-state, transition-by-transition, I propose an approach for the robot to extract important cues for inferring intentions of others based on its own prefernces for action in operating context. Chapter 6 demonstrates the feasibility of this approach on a bimanual robot.

2.4 Developmental Learning

One of the key elements of the proposed learning approach for robots to develop communicative behavior is the use of developmental stages—structured learning episodes, where the robot learns behavior incrementally through tasks of increasing level of difficulty. The approach incorporates mechanisms for learning general strategies and subsequently assimilating additional run-time contexts to control the incremental complexity of learning in an "open" environment. Developmental staging can be observed in infant development, engaging processes of growth and maturation and supported by external constraints that parents put on the environment. Some

constraints are there to guarantee the infant's safety, while some are intentionally introduced to allow infants to play with toys of different levels of complexity at the frontier of the infants developing world model. This approach to acquiring skills is the key concept behind *developmental robotics*. It aims to explore theories of epigenetic development to build adaptable and more capable robotic systems [1, 71, 84].

Developmental staging has been successfully demonstrated for robots to learn useful behavior. Gomez [36] and Lee [63] both provide time-varying developmental constraints to guide robot exploration. Constraints are relaxed incrementally as the robot gains more competency. Staged learning provides a means for an agent to build knowledge incrementally and learn increasingly complex skills [2, 55, 15]. Edsinger and Kemp showed how a humanoid can develop knowledge about its appendages (i.e., its hands and fingers) and held tools in a coarse-to-fine, proximal-to-distal, multi-stage experiments [24, 23].

In contrast to traditional approaches, where complete and deterministic knowledge of the world is needed to ensure success, developmental roboticists advocate *situated learning* where the system uses its sensorimotor resources to explore the environment. A number of researchers, e.g., Sandini [93], Grupen *et al.* [85, 14, 47], and Asada *et al.* [1], have proposed computational methods for robotic systems to learn in situ by exploring interactions with the environment using combinatorics of their sensorimotor resources. They have argued that this approach can lead to adaptive complex behavior suitable for acting in unstructured "open" environments.

Situated learning also implies that physical embodiment is required for a learning agent—another key distinction between developmental robotics methods and traditional methods in artificial intelligence. From the rule-based approach [81, 73], to the formal representation of commonsense knowledge [43, 45, 18], traditional artificial systems learn using symbolic abstractions of the world, rather than grounded sensorimotor signals. For instance, large-scale knowledge collection projects such as

CYC [65] and OpenMind [105] gather knowledge in the form of logical assertions through textual analysis. However, these systems have yet to achieve real-world competence in any behavioral task. A possible explanation, argued by the developmental robotists, is that knowledge acquired through symbolic and textual analysis lacks the sensorimotor grounding necessary for such knowledge to be applied to real-world situations. Learning through physically embodied robots ensures knowledge is acquired in a grounded manner.

Grounded situated learning has been applied for robots to learn about the visual appearance of its own limbs [80]), or what things its hand and fingers can actively control [24]. In the domain of language and communication, a number of compelling recent studies in developmental robotics illustrate that robots can ground language [108, 92, 82] by learning the association between words (as sound utterances [90, 82] or textual tokens [110]) and actions.

2.5 Affordance Modeling and Behavioral Knowledge Representation

In terms of knowledge representation, this work adheres to the theory of "affordance learning." J.J. Gibson states that "... the affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. [34]" He argued that our perception and understanding of the world is stored and applied in terms of behavior that the environment affords. Therefore, embodiment and interaction with the world are necessary for building grounded knowledge, a crucial part of cognitive development. This work advocates a unified approach for acquiring and representing knowledge and treats robot interactions with humans in precisely the same way as it acquires skills for interacting with inanimate objects. I hypothesize that not only are learning processes and representations shared between human beings and inanimate objects, but knowledge regarding how to interact with inanimate

objects also informs social interactions driven by social incentives (mutual reward and underactuation). When directed toward humans, these behaviors automatically become communicative in nature. In this section, a background review of the existing applications of Gibson's "theory of affordances" in the computer science literature is provided.

Gibson's theory of affordances has a significant impact on the field of Human Computer Interaction (HCI). The theory of affordances is widely cited as the underlying guiding principle for software interfaces [30, 106] and high-degree-of-freedom input device designs [120]. Affordance is interpreted to be an objective property of the environment that is associated with specific capabilities of the actor, and can be learned from experience. Conversely, prior experience can influence how a person predicts affordances. If an affordance of an object does not match the expectation of the actor, this often leads to confusion and the affordance may not be discoverd. For instance, in Figure 2.3, doors with wide horizontal bars naturally suggest pushing on the bar from a human's previous experience with manipulation of objects in general. However, if the design of the horizontal bar actually affords pulling and rotating instead of just pushing, it can easily lead a human to believe the door is locked and cannot be opened. This is an example that shows how a relatively small visual change can have a dramatic impact on our policies for interacting with the larger concept of "door." The application of the affordance theory helps designer to ensure that user interfaces are built to highlight the affordances of the devices, rather than obscuring them [30].

In the field of robotics and AI, affordance theory is applied to learn generalizable properties of objects that are more robust than visual appearance models. Several recent robot learning techniques have been proposed and applied to demonstrate the extraction of environmental affordances. Chamero defined affordances as a relationship between an agent and an object in terms of the potential for action [12].

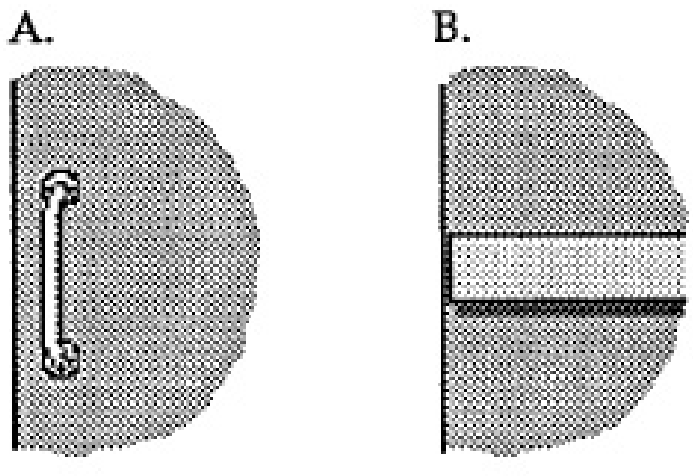


Figure 2.3. Different door handles suggest different affordances. Narrow vertical handles suggest grabbing and pulling, while wide horizontal bars suggest pushing.

Fitzpatrick *et al.* illustrated how a robot can learn "pushing" and "grasping" affordances by interacting with objects [26]. Stoytchev's robot has learned to use tools by exploring object-object affordances between tools and other objects [111]. More recently Sinapov has shown how the sounds derived from interacting with objects are strongly correlated with other affordances and that by association, inform policies for action [104].

Most work in the affordance modeling focuses on grounding knowledge by learning affordances in terms of low-level primitives or hand-coded behavior. In contrast, the majority of work in traditional AI focuses on high-level symbolic planning without low-level grounding in the robot behavior. To bridge the gap, a formalism of affordances called "Object-Action Complexes" (or OACs) has been created to both ground representations of the world in the robot's interactions with objects and to use them for higher-level planning tasks [32].

According to [70], there are two properties of affordances that Gibson implies but never directly states. The first is that affordances can be nested so that the potential for action can incorporate one or several action possibilities. For example,

an apple affords eating, but eating is composed of biting, chewing, and swallowing. Secondly, Gibson implies that affordances are binary: they either exist or they do not. For example, an object is either graspable or it isn't. However, in the real-world, an action possibility exists probabilistically, conditioned on other properties of the run-time environment. For instance, a stair is climb-able but the difficulty level associated with this affordance depends on the number and size of the steps. In robotics, OACs and the framework proposed in this work both allow modeling of hierarchies of affordances. However, regarding the second property, OACs [32] uses binary assertions, while the framework in this thesis supports encoding affordances in terms of probabilities.

2.6 Modeling Humans for Human-Robot Interaction

The HRI research community is focused on problems regarding collaboration, i.e., how activities of a human and a robot can be coordinated to produce an adaptive policy for cooperation [97]. Schaal pointed out that for the collaboration between humans and robots to be successful, there exist a number of significant challenges: the detection of humans in the environment, the recognition of human gestures and intentions, and the conveyance of intentions from the robot to the human using motor actions. In Section 2.2, some of the seminal work in HRI relating to motor action generation and recognition has already been reviewed.

Many studies in HRI [10, 23, 76] rely on existing methods from the computer vision literature for the purpose of detection and tracking of human motion, or the recognition of human gestural cues. In this section, an overview of these techniques is provided.

2.6.1 Computer Vision Techniques for Modeling Humans

Finding humans, tracking human motions and identifying them in natural settings are difficult problems for computer vision. A great deal of research focues on different aspects of these problems for decades. The literature on this subject is too exhaustive to review completely. In this section, a few examples are described to illustrate the current prevailing approaches.

Human detection is difficult because human appearance changes daily and body motions are non-rigid. Occlusion, variations in pose, clothing, and articulated motion all contribute to the challenge. Currently, the most effective approaches for whole-body human detection and tracking are part-based methods, where human subjects are modeled as assemblages of parts with kinematic relationships between features. Earlier work in this line of research used 3D kinematic models [46, 31, 64]. However, for these methods, stereo correspondence is an issue and also 3D models have many parameters and degrees of freedom that introduce computational complexity.

As a simpler alternative, there have been approaches where the human body is modeled as a tree of 2D parts [102, 48, 87] where a generative probabilistic model of humans is learned using labeled training data. Inference (using non-parametric belief propagation, or NBP) is performed on the graph structure to detect humans and estimate their poses. For simplicity, some researchers do not rely on a complex graphical model. Instead they define a number of constraints using prior knowledge about the human body and then either search [74] or use dynamic programming to assign labels to body segments [88]. However, these methods are computationally intensive and so far do not satisfy the real-time performance requirements of human-robot interaction. More importantly, none of these approaches have been proven to work robustly in general, dynamic environments.

For many scenarios in human-robot interaction studies, robots interact with humans who are standing close to them, and directly facing their cameras. Therefore,

for many, the use of human part detectors such as faces and hands are sufficient. Faces and hands are two of the most broadly studied human features. For detection, many methods have been proposed, focusing on a great variety of features such as color, shape, texture [16, 44]. Texture-based approaches currently yield the best results, since texture is relatively robust under different illumination conditions. The Viola & Jones face detection algorithm [115] is an example of this type of approach, where a cascade of rectangular texture-based classifiers (Figure 2.4) is trained to achieve efficient and robust detection of faces. However, the down-side of using simple rectangular detectors is that they are much less descriptive than other types of features, e.g. features derived from steerable Gaussian filters. As a result, these approaches are prone to failure when the face is slightly rotated or partially occluded.

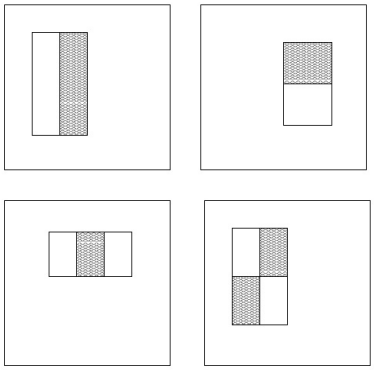


Figure 2.4. The rectangular features used in the Viola Jones face detection algorithm. These features are simple to compute and their effectiveness has been demonstrated in the face detection domain.

A great deal of work focuses on the problem of hand detection and tracking individually. Most attempts are made in the context of in-place video sequences (the cameras remained fixed). Under such constrained conditions, many found skin-color-based detection sufficient [54, 95]. There are many instances in a natural environment where these methods will fail, for example, situations where other skin-color objects or faces are present. Under constrained postural and viewing angle conditions (e.g. from

the view of a head-mounted display of a wearable computer), robust hand detection against arbitrary backgrounds can be achieved [61]. Here, the detector was trained using the same cascade filter algorithm that Viola & Jones used for face detection. Due to the descriptive limitations of the original rectangular detectors, the authors augment the feature set with customized and more computationally expensive features. As a result, a simpler, *flock of features approach* is devised for tracking once the hand has been detected [60]. Many particle filter methods have been proposed for hand tracking (c.f. literature review [6]), however, they are only effective after the hand has been detected.

Finer features associated with eyes, mouth and eye-brows have been studied under the context of facial expression recognition [21]. Applications that use eye tracking as a natural pointing device to replace a computer mouse have been designed [49]. They are applications of different feature tracking algorithms such as particle filters and are sometimes embellished with domain specific prior knowledge to increase robustness. They either assume a face is visible against a clean background or rely on face detection algorithms to locate a facial region in cluttered environments in order to initialize the tracker. Therefore, these methods fail in the same situations that cause face detection to fail.

2.6.2 A Behavioral Approach to Human Modeling

The approach taken by this thesis is distinguished from prevailing methods in computer vision and human-robot interaction (HRI) in two important aspects: (1) robots can actively take actions to change their perception to make vision problems simpler, (2) robots can take actions with incomplete knowledge of the surrounding environment and make progress toward its goal while exposing new information.

However, this idea has been recently picked up by the robotics community and several researchers demonstrated the utility of using basic behavior and interaction

to improve visual learning. For instance, Steels' Aibo experiments [109] show that social learning through natural interaction contributes significantly to successful new visual category formation and language learning (in the form of "first few words"). Fitzpatrick [27] shows that a robot can use simple probing actions to overcome difficult computer vision problems such as foreground object extraction. Katz and Brock [56] uses these ideas to extract kinematic models of articulated objects using hand-crafted motions. Edsinger's humanoid robot, Domo, has been shown to make self-other distinctions by identifying visual patches that are controllable [24].

However, to the best of my knowledge, this approach has not been applied to learning about humans, where a robot improves its understanding of humans incrementally through interactions. For instance, I hypothesize that it is possible for robots to learn to model humans in a way similar to Edsinger's work [24]—by finding actions capable of "controlling" (albeit, indirectly) the human subject. Humans are indeed independent agents whose actions cannot be completely predicted. However, humans are social beings and therefore respond predictably to social cues and gestures under the appropriate conditions. In this thesis, these conditions are defined by the "under-actuation" and "mutual reward" conditions stated earlier. Given these conditions, with the right learning framework, it is up to the robot to explore its actions and to discover behavior that has a high likelihood of "persuading" the human to help by conveying intentions and, in a sense, "controlling" the human in order to achieve a mutual goal. Vice versa, this allows humans to control the robot as well, by engaging the appropriate gestural actitity.

More specifically, this thesis proposes an affordance modeling approach for robots to learn about humans by defining how humans afford controllable behavior. The central thesis is that compared to visual appearance alone, behavioral patterns are much more informative, predictable, and reliable than ungrounded symbols.

2.7 Summary

In summary, the approach taken by this thesis is an advance in HRI on two fronts: (1) it unifies the way the robot learns motor skills, objects, and human beings and as a result, knowledge and skill transfer can occur due to a common representation and learning structure, and (2) it creates learning mechanisms for robots to acquire models of human beings and how to interact with them at the same time since behavior is now the main focus of the learning processes in both cases. Furthermore I hypothesize that this unification not only simplifies the learning process but also can provide significant improvement in learning efficiency in many cases due to knowledge and skill reuse. The goal of this thesis is to develop mathematical formalisms necessary to realize these insights and to provide experimental results to support these claims.

CHAPTER 3

A FRAMEWORK FOR LEARNING MANUAL BEHAVIOR

The representational foundation for this work is based on the *control basis*, a representation for learning hierarchical control programs given sensory and motor resources. It was originally introduced by Huber and Grupen [47] as a means for robots to autonomously construct controllers and actively explore the combinatoric space of sensory and motor resources. The framework is recently extended by Hart [40, 39] with elements of intrinsic motivation, hierarchy and generalization.

Using this framework, a designer can guide a robot's learning process by controlling the resources and external stimuli made available to the robot at different times, thus creating a series of increasingly challenging learning stages. The robot learns simple programs first and subsequently moves onto more challenging scenarios using programs learned in the previous stages.

Hart's thesis [41] demonstrated the framework focusing on the development of manipulation skills through intrinsically motivated exploration using simple graspable objects. In joint work with Hart, we proposed an affordance-based modeling approach for objects. Hart demonstrated the approach on the constructions of object stacks. To explore complex "objects" such as humans beings, this thesis further expands the framework in several aspects: (1) a multi-modal sensory processing pipeline is integrated with the behavioral learning framework, (2) the formalization of a hierarchical catalog model suitable for representing humans, (3) a prospective learning algorithm for robots to adapt behavior to situations where simple local generalization

fails. These extensions lay the groundwork for modeling articulated objects and objects that possess "agency"—entities with the ability to make independent decisions and goal oriented actions. This is essential for the development of communicative behavior in later chapters.

In this chapter, the hierarchical manual behavior learning framework, as presented by Hart [41], is briefly summarized to provide background for the extensions to be described in later chapters. The sensory processing pipeline is introduced in Section 3.2. In Section 3.4, an example borrowed from Hart's thesis is given to illustrate the behavioral learning process. Finally, Section 3.6 describes several hierarchical control programs, learned using the framework presented in this chapter. These control programs will be used as the behavioral substrate for acquiring communicative behavior in future chapters.

3.1 Control Action and State Estimation

The control basis is designed for robots to autonomously construct control actions to explore the combinatoric space of sensory and motor resources. Primitive actions in the control basis framework are closed-loop feedback controllers constructed by combining a potential function $\phi \in \Omega_\phi$, with a feedback signals, and discrete motor variables $\tau \in \Omega_\tau$.

The potential function ϕ is a scalar *navigation* function defined to satisfy properties that guarantee asymptotic stability. Motor variables are discrete, actuatable degrees of freedom with continuous motor inputs u_τ. Feedback for control circuits consists of a variety of features extracted from a discrete set of feedback signals, $f_\sigma \in (\Omega_o \times \Omega_\sigma)$, where o denotes a convolution operator in a set of possible filters Ω_o and $\sigma \in \Omega_\sigma$ represents a physical sensor that publishes a raw signal, g_σ. Patterns of individual responses define vectors $\mathbf{f}$ that can encode relational properties among feedback channels and can likewise be used as feedback in hierarchical control circuits.

Details regarding how **f** is computed will be given in Section 3.2. A specific instance of a control circuit is denoted $c(\phi, f_\sigma, \tau)$ and the number of possible primitive actions is thus bounded by $|((\Omega_o \times \Omega_\sigma) \times \Omega_\phi \times \Omega_\tau)|$.

Currently, the set of potential functions Ω_ϕ in the control basis includes:

- **Quadratic potential function**—a convex quadratic function of the feedback errors. An example is Hooke's law, defined as:

$$\phi_s(f_{\sigma_{ref}}, f_{\sigma_{act}}) = \frac{1}{2}(f_{\sigma_{ref}} - f_{\sigma_{act}})^T(f_{\sigma_{ref}} - f_{\sigma_{act}}) \tag{3.1}$$

  where the difference between the actual and the reference feedback signals, $\sigma_{act}, \sigma_{ref} \subseteq \Omega_\sigma$, captures virtual errors between two features of the same type. This potential function can be employed for configuration control, spatial position control or force control.

- **Harmonic function**—an artificial potential function that satisfies Laplace's equation. It has the property of no local minima or maxima and therefore is used to compute collision-free motion paths.

- **Kinematic conditioning functions**—conditioning fields are used to provide a natural way for the robot to optimize its kinodynamic configuration. Several fields have been implemented to keep a manipulator away from joint range limits (*rang limits field*), to optimize "manipulability" and avoid singularities (*manipulability field*), and to maximize stereo triangulation quality (*localizability field*).

For convenience of discussion in an example later in the chapter, the mathematical definition of quadratic potential function has been given in equation 3.1. Detailed definitions of other potential functions can be found in Hart's thesis [41].

The sensitivity of the potential to changes in the value of motor variables (u_τ) is captured in the task Jacobian, $J = \partial\phi(f_\sigma)/\partial u_\tau$. Reference inputs to lower-level motor units are computed such that

$$\triangle u_\tau = -J^{\#}\phi(f_\sigma) = -(\frac{\phi(f_\sigma)}{\partial u_\tau})^{\#}\phi(f_\sigma), \tag{3.2}$$

where $J^{\#}$ is the Moore-Penrose pseudoinverse of J [78]. With no motor variables attached, the controller becomes a *monitor* $C_M(f_\sigma, \phi)$ that simply observes the feedback signals (off-policy) passively for the purpose of event detection. The use of monitors will be discussed further in Chapter 6.

Multi-objective control actions are achieved in the control basis by combining control primitives using nullspace composition [78]. Nullspace composition allows control primitives be combined in a prioritized manner, ensuring the lower priority controller does not interfere with the objective of the higher priority controller. For instance, given a higher priority controller, $c(\phi_1, f_{\sigma_1}, \tau_1)$ and a lower priority controller $c(\phi_2, f_{\sigma_2}, \tau_2)$, a multi-objective controller can thus be defined as

$$c(\phi_2, f_{\sigma_2}, \tau_2) \triangleleft c(\phi_1, f_{\sigma_1}, \tau_1).$$

The operator "$\triangleleft$"—read as "subject-to"—is used to represent the prioritized combination between any two control actions [47]. A concrete example of this controller construction process is illustrated in Section 3.4.

The state of a control process, denoted as a predicate $p(\phi, \dot{\phi})$, is created to describe the status of the corresponding controller $c(\phi, f_\sigma, \tau)$ when its interacts with the task domain. To support a natural discrete abstraction of the underlying continuous state space, a simple discrete state summary of the dynamics based on *quiescence events* was proposed in [40]. Quiescence events occur when a controller reaches an attractor state in potential ϕ. For state description, Huber [47] proposed $p(\phi, \dot{\phi}) \in \{0, 1\}$

for the control basis, while Coelho [13] adopted a set membership approach that built empirical models of first order control dynamics. Hart defined the state of a controller as $p(\phi, \dot{\phi}) \in \{X, -, 0, 1\}$ [41]. In this dissertation, we will adopt Hart's state description from [41], more formally defined as:

$$p(\phi, \dot{\phi}) \quad = \quad \begin{cases} X & : \quad \phi(f_\sigma) \text{ controller is not activated} \\ - & : \quad \phi(f_\sigma) \text{ has undefined reference} \\ 0 & : \quad |\dot{\phi}| > \epsilon_\phi, \text{ transient response} \\ 1 & : \quad |\dot{\phi}| \leq \epsilon_\phi, \text{ quiescence,} \end{cases} \tag{3.3}$$

where ϵ is a small positive constant. The "X" condition specifies that we either *don't know* or *don't care* what the status of c_i is. Typically, this is the initial state of all controllers immediately after being engaged and before predicate p_i is evaluated. The "$-$" condition means that no target stimuli is present in the feedback signal σ, and the environment does not afford that control action at that time. The "0" occurs during the transient response of c_i as it descends the gradient of its potential, and "1" represents quiescence. Given a collection of n distinct primitive control actions, a discrete state-space $\mathcal{S} \equiv (p_1 \cdots p_n)$ is automatically formulated. Next, the processing pipeline for extracting features and the available sensor signal set Ω_σ are discussed.

3.2 Signal Processing Pipeline

This section provides a description of the signal processing pipeline for extracting features f_σ from sensory channels (Ω_σ). The resulting features form the perceptual basis for robots to generate control actions using the control basis. The robot perceives the world through a broad range of features extracted from visual, proprioceptive, and force signals. This work employs two robotic platforms, Dexter and the uBot (as shown in Figure 3.1). Both robots have a stereo camera pair mounted on a pan/tilt head, two arms and two hands. The difference is that Dexter has two 7-DOF Whole-

Arm Manipulators (WAMs) and two 3-finger 4-DOF hands, while the uBot has only 4-DOF arms and two 2-finger hands. However, the uBot is a dynamically balancing mobile robot with two wheels and Dexter is fixed to the ground.

Figure 3.1. This work employs two robotic platforms, Dexter on the left and the uBot on the right.

For these robots, signals from the following channels can be extracted:

- **Visual**: information is captured from the cameras mounted the robot. This channel of information is sub-divided into subchannels with pre-processing and filtering. A typical color camera image can be decomposed into RGB, YUV, or hue, saturation and intensity (HSI) color spaces. Intensity/gray-scale images are used to compute texture or motion segments.
  - **color** - the hue, saturation and intensity color space is discretized into 18 channels of hue, 10 channels of saturation and 10 channels of intensity.
  - **texture** - multi-scale Gaussian derivative operators according to the Koenderink scale-space theory [59]. Gaussian derivatives can be used to describe various texture features such as scale space corners, ridges, and blobs.

– **motion segments** - a channel of motion segments in the scene. For this work, this channel is implemented as a union of all color channel features in motion. Other alternatives such as dynamic background subtraction or persistent backgrounding [22] have also been implemented.

This visual sensor resource set Ω_γ is thus defined as:

$$\Omega_\gamma \quad = \quad \left\{\gamma_{motion}, \gamma_{hue,i}, \gamma_{sat,j}, \gamma_{int,j} \mid i \in \{1, ..., 18\}, j \in \{1, ..., 10\}\right\}, \quad (3.4)$$

where $\gamma_i \in SO(2)$ is heading toward features on channel i.

- **Force**: a means of measuring when the robot makes contact with objects in its environment, including itself. Forces and torques can be measured from load-cells, strain gauges, capacitive surfaces, or from examining the motor currents of a robot's joints. For this channel a force vector ($\vec{f} \in \mathbb{R}^3$ for Dexter and $\vec{f} \in \mathbb{R}^2$ for uBot) is obtained from finger tip load cells of the robot, and a scalar value f_{net} is computed by normalizing the force vector. This set of signals is

$$\Omega_f \quad = \quad \left\{\vec{f}, f_{net}\right\}, \quad (3.5)$$

where $\vec{f}$ is the force measured on the fingers, and f_{net} the normalized scalar value of f.

- **Proprioceptive**: scalar joint angle values for each joint of the robot. These values determine the robot pose and actuate the robot when modified. This set of sensor resource is:

$$\Omega_\theta \quad = \quad \{\boldsymbol{\theta}_{arm}, \boldsymbol{\theta}_{hand}, \boldsymbol{\theta}_{head}\}. \quad (3.6)$$

where θ_{arm} are the configuration of the robot's arm, θ_{hand} are the configuration of the robot's hand, and θ_{head} are the configuration of the robot's pan/tilt head. Both robots have two arms, two hands and a pan/tilt head.

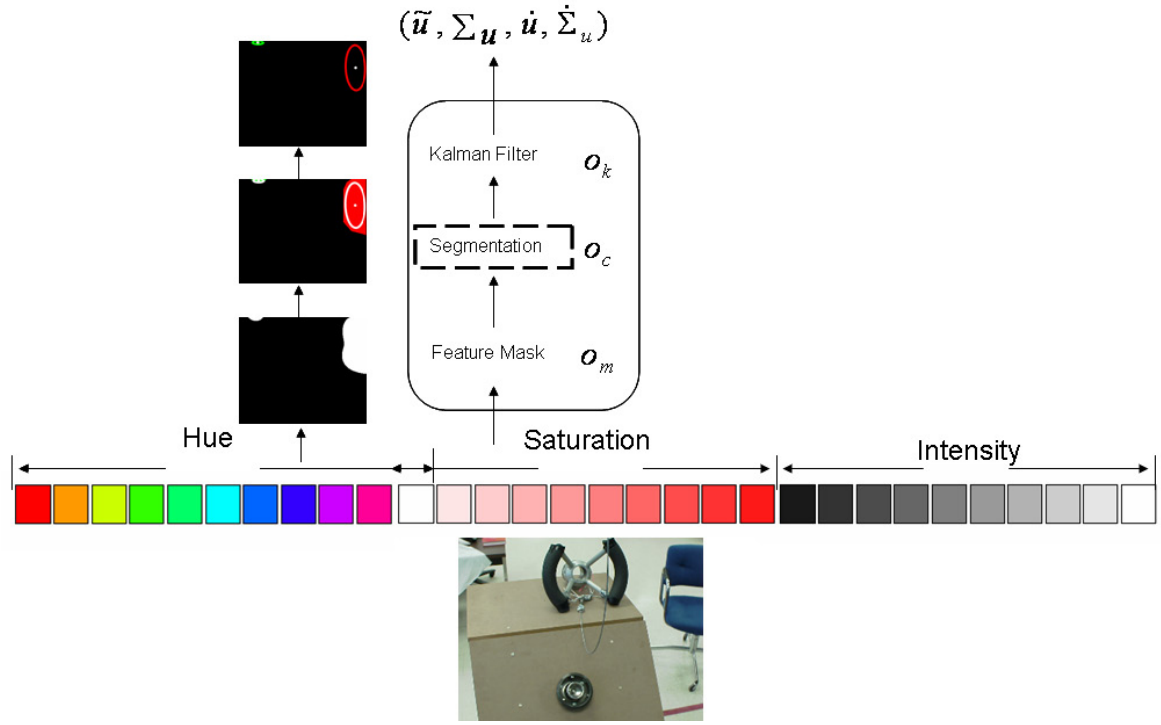


Figure 3.2. The visual signal processing pipeline: raw sensory input from each visual channel is first filtered using a feature mask, then contiguous regions are segmented and finally a Kalman filter is applied to provide a summary of the first order dynamics of each type of feature in space and time.

For this work, only color channels (hue, saturation and saturation) in the visual channel, finger tip forces and scalar joint angle values are processed to extract features. These features are used as potential σ for constructing controllers. The visual channels are processed using a signal processing pipeline (Figure 3.2), through a succession of operators $o \in \Omega_o$. First, each channel of raw sensory input is filtered using a corresponding feature mask operator o_m, e.g. hue values within a certain range. For visual channels, a connected components operator o_c is then applied to segment contiguous regions that share a feature. Finally, a Kalman filter operator o_k is applied

to provide optimal, mean ($\widetilde{u}$) and covariance (Σ_u) estimates of the feature distribution as well as its first order dynamics ($\dot{u}, \dot{\Sigma}_u$) in the presence of noise. Thus, a summary of the first order dynamics of each type of feature in space and time is delivered as a perceptual basis for the subsequent object/human modeling and behavioral learning. It is up to the robot to explore this feature space by constructing controllers using the control basis (discussed next) in search for ones that reliably lead to reward.

The discussion on how related features are modeled and archived is deferred until Chapter 4, while how a robot uses the output from the signal processing pipeline to construct actions for exploring the world is presented next.

3.3 Learning Hierarchical Behavior using Intrinsic Reward

To drive the learning process, this framework defines a simple intrinsic reward function $\mathcal{R}$ that provides reward when a controller state transitions from a non-convergent state to convergence. More formally, Hart defined intrinsic reward as the following:

$$b_i^k = \left((p_i^{k-1} \neq 1) \wedge (p_i^k = 1)\right), \tag{3.7}$$

$$r_i^k = \begin{cases} 1 & : \text{ if } \left(b_i^k \wedge (\sigma_i \subseteq \Omega_{\sigma(env)})\right) \\ 0 & : \text{ otherwise} \end{cases} \tag{3.8}$$

$$r^k = \sum_i r_i^k. \tag{3.9}$$

where p_i^k is the state of a controller $c_i = c(\phi_i, f_{\sigma_i}, \tau_i)$ at step k, and b_i^k is the binary bit indicator for the convergent event for controller i at step k. As a result of Equation 3.9, the intrinsic reward function provides a unit of reward for all controllers that converge at step k, and the reward the robot receives is the sum. The condition that only controllers using feedback signals from the environment ($\sigma \subseteq \Omega_{\sigma(env)}$) can be

rewarded is very important in this formulation since we are interested in learning the effect of actions on the environment.

The state and action spaces $\mathcal{S}$ and $\mathcal{A}$ defined by the set $\{\Omega_\phi \times (\Omega_o \times \Omega_\sigma) \times \Omega_\tau\}$ and reward function $\mathcal{R}$ form a Markov Decision Process (MDP) for control. Value iteration algorithms like Q-learning [112] provide a means of estimating the value, $\Phi(s, a)$, of taking action a in state s using the update-rule:

$$\Phi(s,a) \leftarrow \Phi(s,a) + \alpha\big(r + \gamma \max_{a'} \Phi(s',a') - \Phi(s,a)\big)$$

where $\gamma \in [0, 1]$ is the discount rate, r is the reward received, and $\alpha > 0$ is a step-size. With sufficient experience, this estimate is guaranteed to converge to the optimal value Φ^*. The optimal policy π^* can then be extracted by selecting actions that maximize the expected sum of discounted future reward, such that

$$\pi^*(s) = \text{argmax}_a \Phi^*(s,a).$$

To balance exploration and exploitation, an ϵ-greedy approach is used where the agent with $1 - \epsilon$ probability selects a random exploratory action.

After value iteration converges, the value function, states and actions are packaged in a control schema representing a policy for discovering rewards in a variety of circumstances. *Schemas* can be viewed as a temporally extended abstract actions with three probabilistic outcomes (Figure 3.3) plus the associated knowledge. This abstraction preserves the semantics of primitive controllers and supports the hierarchical invocation of schema.

Representing behavior in terms of a value function provides a natural hierarchical representation for control basis programs where absorbing states in the MDP represent quiescence events in the policy. Therefore, the state of a program can be captured using the same state-predicate representation, $\{X, -, 0, 1\}$, as for a primitive action,

even though that program may have complex internal transition dynamics (Figure 3.3). The learned program can be invoked hierarchically as an abstract action, thus enabling the robot to acquire more complex behavior.

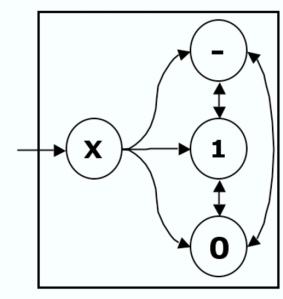


Figure 3.3. An iconic representation of a state transition when a temporally extended sensorimotor program called a schema is invoked hierarchically.

Hierarchy for this work is induced via a developmental staging strategy. Each stage is defined by a set of resources. For example, the robot can be constrained to recruit effector resources consisting of the pan/tilt degrees of freedom of the stereo head, excluding arms and hands. Then, all behavior consists of visual tracking. Later, effectors in the arms can be incorporated to reach to and touch interesting features it tracks visually. Learning in incremental stages allows the programmer to direct the exploratory behavior of the system and thus influence the size of the state and action space. The unified framework on which knowledge is gathered and archived makes reuse and transfer of knowledge feasible. For instance, behavior learned in an early stage can be invoked in all the subsequent stages as a temporally extended action. Examples of these hierarchical programs are given in Section 3.6.

3.4 Example: SearchTrack

In this section, the learning framework is illustrated with an example in which a simple program that has been documented in Hart's dissertation [41]. The description and some of the notations has been updated in this document to improve clarity.

This program is called SEARCHTRACK, it is useful for finding and tracking visual stimuli with Dexter's pan/tilt head is presented. In this stage, Dexter is only allowed to use the effector that direct a pair of cameras, although the behavior depends on sensor feedback, σ, from only one of them. We also assume that the environmental reference that ultimately drives behavior is stimuli that reflect the most highly saturated hues on the image plane of Dexter's left camera, sat_{10}. Given this constraint, two controllers, SEARCH and TRACK are employed.

Both controllers are defined using common resources in the control basis framework. Both control circuits are constructed using feedback signals ($\sigma \subset \Omega_\sigma$) that include the joint angle configuration of Dexter's head, θ_{head}. Moreover, both engage these same degrees of freedom as effector variables ($\tau \subset \Omega_\tau$). SEARCH and TRACK are distinguished solely by the source of their respective control references. Formally, the two controllers are defined using the control basis formulation as follows.

SEARCH - controller $c_{search} = c(\phi, \sigma, \tau)$, where

$$\sigma = \{Pr(\theta_{head}|sat_{10}), (\theta_{head})_{act}\}, \text{ and } \tau = \theta_{head}.$$

To generate the search potential, the error between the reference value sampled from $Pr(\theta_{head}|sat_{10})$ and the feedback $(\theta_{head})_{act}$ is computed

$$\epsilon_{\mathbf{1}} = \boldsymbol{\theta}_{sample} - \boldsymbol{\theta}_{act}, \text{ and}$$

$$\phi_{search} = \boldsymbol{\epsilon_1}^T \boldsymbol{\epsilon_1}.$$

And finally, the error signal that drives the motor unit is computed as

$$\Delta u_\tau \propto -\left(\frac{\partial \phi(\sigma)}{\partial u_\tau}\right)^{\#}_{track} \phi(\sigma)$$

and search action is thus:

$$c_{search} \triangleq c(\phi_{search}, \epsilon_1, \theta_{head}).$$

TRACK - controller $c_{search} = c(\phi, \sigma, \tau)$, where

$$\sigma = \{(\theta_{head})_{obs}, (\theta_{head})_{act}\}, \text{ and } \tau = \theta_{head}.$$

To generate the track potential, the error between the observed sat_{10} image reference and the feedback $(\theta_{head})_{act}$ is computed

$$\epsilon_{\mathbf{2}} = \boldsymbol{\theta}_{obs} - \boldsymbol{\theta}_{act}, \text{ and, once again}$$

$$\phi_{track} = \boldsymbol{\epsilon_2}^T \boldsymbol{\epsilon_2}$$

And finally, the error signal that drives the motor unit is computed as

$$\Delta u_\tau \propto -\left(\frac{\partial \phi(\sigma)}{\partial u_\tau}\right)^{\#}_{track} \phi(\sigma)$$

and the track action is thus:

$$c_{track} \triangleq c(\phi_{track}, \epsilon_2, \theta_{head}).$$

In SEARCH, the control reference is sampled from a probability density function, $Pr(\theta_{head}|sat_{10})$, that summarizes the places where sat_{10} features have been found in

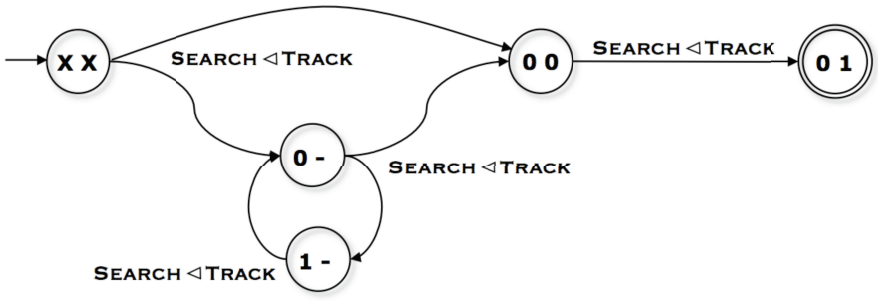


Figure 3.4. The learned policy for searching and tracking features in the environment. The policy begins with a 'XX' state indicating neither the SEARCH nor the TRACK controller has been activated. After executing both SEARCH and TRACK actions concurrently, if the feature stimuli has not yet been discovered (state '0-' or '1-') then the robot continues the search process. On the other hand, if the stimuli is found (state '00') then the robot tracks until its gaze is foveated on the feature (state '01').

the past. This distribution begins with a uniform distribution and is updated as the robot gathers more experience. For TRACK, the visual processing pipeline delivers a stream of coordinates of sat_{10} defined in the heading space (altitude and azimuth angles). A feedback error is computed between the observed heading reference for sat_{10} (θ_{obs}) and the actual heading (θ_{act}) to keep up with the saturation cue. According to the reward model, since the error is directly provided by the stimuli from the enviroment environment, the robot receives reward for the quiescence of c_{track}.

From these two actions, the state space $\mathcal{S}_{st} = (p_{search}\ p_{track})$, and the action set $\mathcal{A}_{st} = \{c_{search}, c_{track}, c_{search} \triangleleft c_{track}, c_{track} \triangleleft c_{search}\}$ were constructed, where $c_{search} \triangleleft c_{track}$ represents concurrent execution of both controllers using nullspace composition, while c_{track} has the higher priority. Given the actions, state space and intrinsic reward, standard Q-learning was used and ϵ-greedy action selection was set to 20% exploration rate. Dexter learned a policy for SEARCHTRACK after 50 episodes of training. Each episode ended when a rewarding event occurred (i.e., the track controller quiesced).

A resulting policy is shown in Figure 3.4. The robot searches until a stimuli is found and begins tracking.

3.5 Generalization

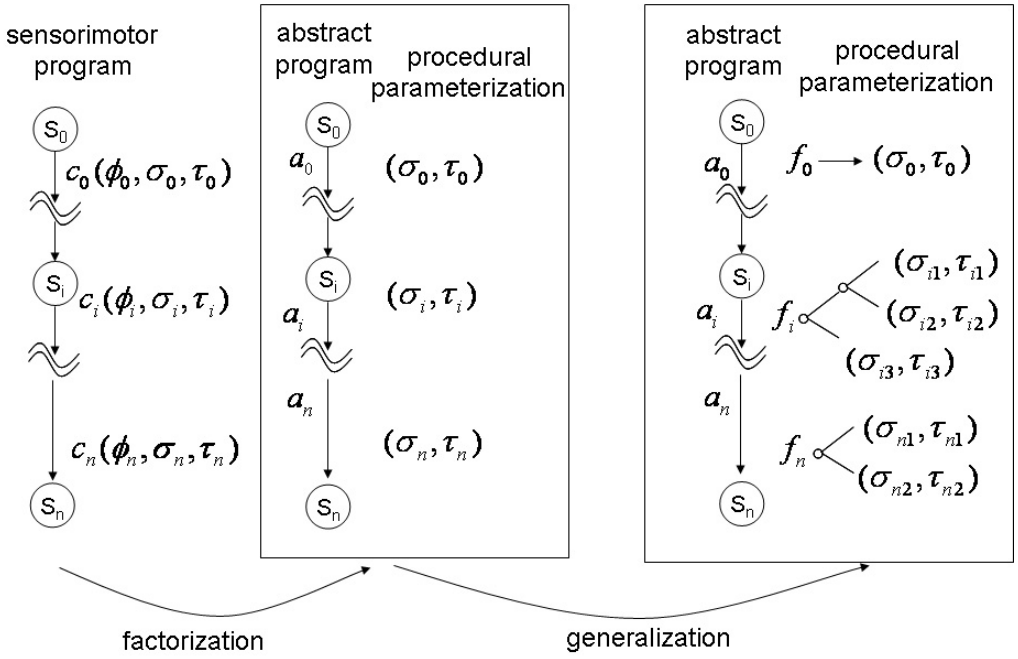


Figure 3.5. Sensorimotor programs are factored into abstract programs and procedural parameterizations such that the structure of the learned program can be re-applied in new environmental contexts defined by $f_i \in \mathcal{F}$ without starting from scratch.

As shown in the SearchTrack example, with constrained context, e.g., limiting the robot's sensor resources to attend only to specific features and effector resources to use only head degrees of freedom, the robot can quickly learn a program for handling the specific context. Once the basic behavior has been learned, more challenging contexts are introduced, such as using objects of various colors or sizes, or placng the object in different regions of the workspace.

To adapt to new contexts, Hart presented a simple generalization strategy [39] where the robot allocates different sensorimotor resources, e.g. if tracking with a

previously learned sensor channel fails, search for another channel. The key to this generalization technique lies in factoring the control program into a *declarative* component and a *procedural* component (Figure 3.5). Factoring allows the robot to quickly generalize to new contexts by observing features f and learning a mapping from f to the appropriate sensorimotor resource (σ, τ) for a given context using a standard decision tree algorithm C4.5.

3.6 Substrate for Learning Communicative Behavior

Using the hierarchical behavior learning framework presented in this chapter, Hart demonstrated on Dexter that a robot can learn a series of increasingly complex manual behavioral programs, with one bootstrapping on behavior learned in previous stages [40]. These behaviors include: ReachTouch for reaching out and touching features it sees, VisualInspect for bringing object closer for inspection, and PickPlace for transporting object to desired locations, each using the previous behavior as an abstract action. A schematic representation of these programs is shown in Figure 3.6.

In the following chapters, these manipulation programs become the behavioral substrate for the learning of communicative behaviors. In Chapter 5, I will show that by using this framework, when social conditions of mutual reward and underactuation are introduced, expressive communicative behavior emerges naturally as the result of intrinsically motivated learning and reusing existing manual behaviors. However, before the robot can learn communicative behavior with humans, it must first acquire some basic concepts about the human object. Hart and I have colloborated on formalizing a technique for building world models using hierarchical manual behaviors acquired in this chapter. Hart demonstrated its use for simple objects [41]. This technique enables robot to represent objects in terms of behavior they afford. In the next chapter, I will introduce this technique under the context of modeling humans, where humans are also learned and represented as a collection of affordances. For organizing

these affordance I will also present a new hierarchical probabilistic representation for this purpose.

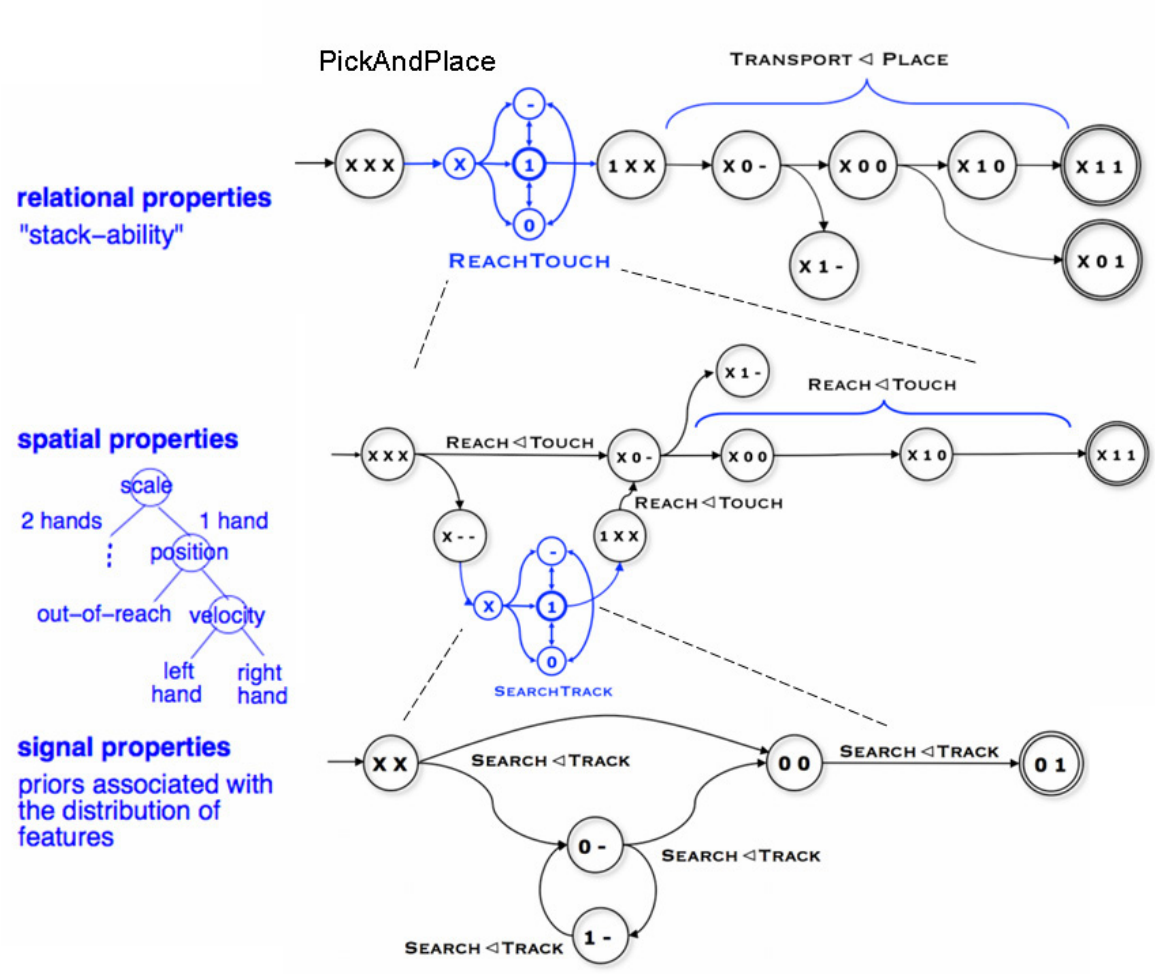


Figure 3.6. A hierarchy of manual behavior emerges as the resulting of intrinsic motivated learning using the framework presented in this chapter. From top to bottom, the control basis formulation enables each behavior to invoke the behavior below as an abstract action (illustrated using an iconic representation from Section 3.3), thus expediting the learning process. The generalization process allows the robot to quickly adapt to new situations and acquire new procedural knowledge in the form of decision trees (as shown on the left of the figure).

CHAPTER 4

BUILDING AN AFFORDANCE MODEL OF HUMANS

We have learned from vision research that finding a robust and invariant representation of humans is a difficult problem because humans are dynamic in appearance and activity. This difficulty is not limited to humans as Figure 4.1 demonstrates. Occlusion, variation in pose, clothing and articulated motion all contribute to the challenge. This work proposes that humans should be represented in terms of their behavior.

In contrast to prevailing techniques where objects or humans are learned passively using statistical machine learning algorithms on large, pre-collected data-sets offline, the psychology literature has revealed that human infants learn by interacting. An infant's concept of an object incoporates the actions they afford. They grab the object, shaking it, putting it into his mouth. Sometimes, new behavior (such as rotating the object) is discovered and enables the infants to extend their understanding of the object. These observations provide us with two important insights for our process of modeling humans: (1) the process is incremental—models are learned and refined over time and improved models lead to the acquisition of complex models and new skills; (2) the process is highly integrated with behavior learning where actions need to be part of the formulation. Neither of these has been demonstrated for a human model.

The following are the hypotheses of this work:

1. Under the appropriate social contexts, human behavior is predictable.

2. Recognizing human behavior does not necessarily require robust tracking of human body parts, or depend on critically on any one feature (like the face), but instead is also related to a holistic relationship between the human subject and the observer and therefore often can rely on simple cues.

3. An incremental and integrated approach for building behavioral models of humans is both possible and desirable as it simplifies the learning process at each step and produces a more robust model for recognition.

Predictability: We argue that in social contexts, human behavior conforms to common standards that are highly predictable and uniquely human. For instance, if you hand someone a book, you can expect that person to take the book, look at the cover, maybe even flip through a few pages and ask some questions about it. If however, the subject is a dog, then totally different behavior would be expected. Social interactions depend on the relationships between the goals and existing behavior of the conversants. Section 1.2 discusses the conditions that define the appropriate social contexts for fostering communicative behavior in a computational framework.

Using Simple Cues: In a social context, the behavior of a human partner is easier to recognize when it can be influenced by the behavior of the robot. In such circumstances, the robot can detect simpler cause-effect aspects of the human without having a complicated model of the human mind. Some researchers, Kruger [94] and Kragic [58] have applied the idea to improve a robot's ability to track human limb motions by focusing on the object the human is interacting with, and using this information to infer the position of the limb when visual tracking alone can produce ambiguous results. This work takes this idea further and argues that humans are defined not only by visual appearance, but more importantly, how we behave and interact with the environment around us, including objects and other agents. This work advocates a developmental approach that uses stages of learning, to build a behavioral model of humans incrementally, beginning with simple cause-effect behav-

ior at first, and then moving on to more complicated and detailed models later as situations require.

An Affordance Model of Humans. The behavior modeling approach taken by this work subscribes to the Gibsonian view that our perception and understanding of the world is stored and applied in terms of the potential behavior that the environment affords [34]. Therefore I argue that as a robot interacts with objects in the world and learns effective manual skills for achieving reward, it can also accumulate a collection of behavioral "affordances" that adequately describe the relationship between proprietary robot actions and the object. For instance, a cup can be described as something that is "grasp-able", "lift-able" and can be used to "contain" other objects. Chairs, on the other hand, all afford "sitting" by the actor and therefore can be considered as "sit-able."

Figure 4.1. Chairs and cups are sometimes difficult to recognize from visual appearance alone.

As shown in the examples shown in Figure 4.1, due to the arbitrary shapes and forms that both cups and chairs can take, the recognition by appearance alone, using passive vision techniques can be extremely challenging. On the other hand, behavioral affordances, such as "grasp-able," "lift-able"[41], "contain-able" [103], or "sit-able" are all defined in terms of functional attributes tested by behavioral programs. The successes (or failures) of which can be easily determined by experiment. Using affordances as defining properties for cups or chairs is robust to variations in physical appearance, or environmental condition changes. Thus, the affordance modeling approach yields "invariant" representations that should perform better than exclusively appearance-based approaches.

We believe by employing the same learning framework that robots use to learn sensory and motor behavior such as the "grasp-able" and "lift-able" properties of objects, a robot can also learn the affordances of humans in the same manner. In Section 4.3, examples are provided to demonstrate how a robot can incrementally accumulate affordances of human using behavioral programs learned in the previous manual skill learning stage. Furthermore, the acquired human model is carried over to the next chapter, where I will show that learning interactive behavior is possible even when the robot only has a coarse and incomplete model of humans. The resulting behavior can be useful as a robust means of confirming human presence, and that perceptual and motor skills also extend the model of objects and humans.

The rest of the chapter is organized as follows. First, we briefly discuss some prior work and ideas that have been adopted for the purpose of modeling affordances of humans. This includes probabilistic frameworks for modeling objects using visual features because I believe they are possible candidate representations for organizing the affordance information. Next, we focus on how we can combine these ideas and adapt them to the control basis for learning multi-modal behavioral affordances of humans through social interactions. Section 4.3 demonstrates how this framework is applied

to enable a bimanual humanoid robot to acquire an affordance model of humans incrementally, using a series of interactions with multiple humans and later, how the learned model is applied for recognition. Finally, in Section 4.5, the implications and potential benefits of the proposed approach are further discussed.

4.1 Related Work in Object Modeling

Researchers in the computer vision community have spent a great deal of effort to model and recognize objects. In recent years, increasing attention has been focused on developing part-based and probabilistic methods. For instance, the feature constellation model by Fergus et al. [25] and the And-Or Graph (AOG) image grammar by Zhu et al. [121, 72] are generative modeling approaches that share the basic idea that objects can be decomposed and represented as a collection of smaller parts. The main difference is that AOG is designed for representing objects in deep hierarchies while the feature constellation approach concentrates on single level decompositions. This work chooses to adopt the graphical model formalism of the And-Or Graph for the purpose of organizing learned affordances in a hierarchical manner. Details will be given in Section 4.2.3.

While many vision techniques focus on modeling the visual appearance of objects in a passive manner, researchers in the robotics community have demonstrated that basic behavior and interaction can be used to improve visual learning [27, 56]. In Chapter 2, we have also covered work by a number of developmental robotists who were inspired by Gibson's theory of affordance and proposed robot learning techniques for modeling objects in terms of the behavior they afford. Using these methods, robots were able to learn affordances of tool-use [111] and affordances in the auditory domain [104].

While most works used hand-coded actions, Hart [41] proposed a computational framework for robots to simultaneously learn new behavior *and* the environmental

conditions that afford its use. The robot is capable of learning and refining behavior by adapting to new contexts as required by the more sophisticated objects it encounters. Hart has demonstrated the use of this framework for the acquisition of affordance models for several objects, two of which are shown in Figure 4.2. Specifically, while both objects afford SEARCHTRACK, REACHTOUCH and BIMANUALTOUCH, only the larger red ball with multi-color patches affords VISUALINSPECT to reveal more multi-color features that are not initially visible when the object is placed on the table.

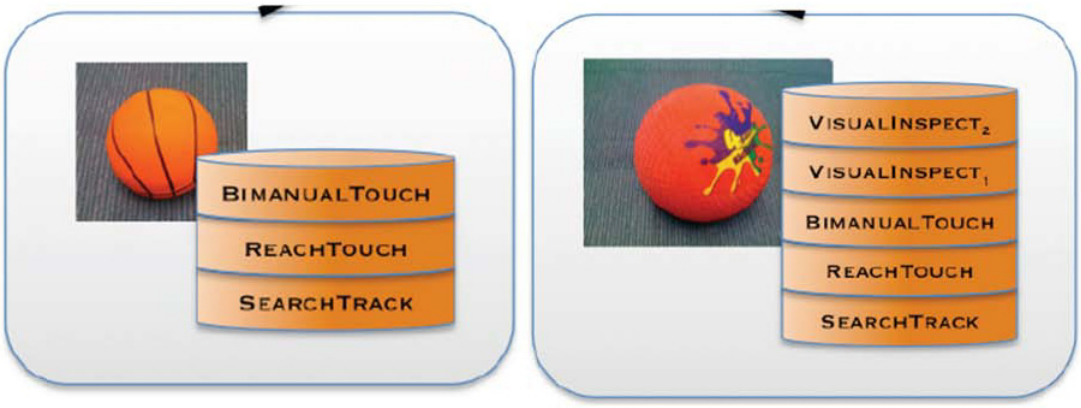


Figure 4.2. Examples of an object "catalog" built using the behavioral affordance modeling approach. Through a series of intrinsically motivated exploratory actions, the robot learns different affordances for the small orange basketball and the larger red ball.

Despite an abundance of attempts to model objects in terms of affordances, there appear to be few applications of this idea to the understanding of human beings. Therefore, this work proposes to model humans as a collection of affordances such that the robot's understanding of humans can be incrementally improved through interactions. Representation-wise, this work shares similar ideas with the part-based methods in the vision literature for probabilistically organizing the data, but adds another dimension to the representation—behavior. My goal is to show how actions can play an important part in the process of modeling the human and how associ-

ated actions are represented in the same model. I also suggest how this behavioral model can be used for both detecting humans in the environment and recognizing the gestures and intentions of the human beings. This work demonstrates this approach with experiments in Section 4.3.

4.2 The Affordance Model Learning Framework

This section presents the affordance model learning framework for robots to actively build knowledge structures of nearby objects in an incremental manner. Section 4.2.1 describes a representation called the affordance catalog for accumulating spatially or behaviorally associated affordances. While most of the behavioral learning framework has already been presented in the previous chapters, the remain pieces relating to affordance learning and organzation are discussed in this chapter.

4.2.1 Catalogs of Affordances

According to Gibson, "affordances" are defined by the behavior that the environment supports (or "affords"). In this framework, objects in the world are represented by a data structure called a *catalog*. Formally, we define a *catalog* $\mathcal{C}$ to be a collection of n plausible affordances (behaviors) and the probability of achieving reward r if the behavior a is executed:

$$\mathcal{C} : \{a, Pr(r|f, a)\}_n$$

where f is the result of some operator applied to signals from sensor resource σ associated with a behavior $a_i \in \mathcal{A}$ (the available sensor resources and signal operators were covered in Section 3.2). It describes a distinctive environmental context and thus allows the robot to build models of contexts that are likely to lead to reward r if a given program $a_i \in \mathcal{A}$ is executed. For this work, we model the probability distributions as Gaussians.

In a developmental learning framework, the robot first learns affordance as behavioral programs in simple contexts. Once the robot has acquired the basic skills, the context is expanded and basic behavior is extended into a more comprehensive set of circumstances. The above representation captures the affordance of an action under the expanded environmental context defined by f. In the following sections, we provide more details on how these affordances are learned and organized to describe interactions with the world.

4.2.2 Affordance Learning

This work explores a unified learning framework based on the control basis to learn behavior and knowledge regarding the world using the same processes. Much of the framework and the resulting manual skills have already been covered in Chapter 3. To learn affordances, the robot can simply apply known behavioral programs in search for ones that lead to reliable reward. Once found, these affordances are added to a data structure called a *catalog*, which was described in Section 4.2.1.

An issue that has not been addressed thus far is related to the focus of attention during the affordance modeling process. In its current form, it is possible for the robot to repeatedly explore the same affordance over and over without any loss of interest. This certainly is undesirable since it will be difficult for the robot to make progress in uncovering new knowledge. The attention-span for an affordance needs to be capped. However, the amount of time required for modeling an affordance is dependent specifically on the features and behavior involved. It is possible to hand-pick a duration for each type of feature-behavior relationships or for simplicity define a single upper-bound fixed duration for all types of feature relationships. However, a single upper-bound would be difficult to define and manually defining a duration requires understanding of the modeling process and even access to the raw sensory information which are difficult for both novice and expert human users alike. Therefore,

it is desirable for an autonomous agent to measure its affordance modeling progress internally. More specifically, when the exploration of an affordance leads to no further knowledge, the robot should direct its attention elsewhere.

This corresponds to *habituation*, a term from psychology for describing the decreasing motivation to attend to a stimulus when it persists over an extended period of time. Computationally, in the control basis, this is defined as an information gain, to capture the fact that there exists a decreasing opportunity to discover new affordance-based facts about a context/object as exploration proceeds. This is similar to work done by Schmidhuber [99, 98, 100] in which the robot seeks to take actions that reduce the "entropy" of the system. Specifically, we define $\mathcal{H}$ by evaluating the information gain of the affordance model $Pr(r|f, a_i)$ for taking action a_i:

$$\mathcal{H} = |\Sigma_i(t) - \Sigma_i(t-1)|$$

where $\Sigma_i(t)$ is the variance of the affordance model at time t. Intuitively, when a feature is first discovered, the model ($Pr(r|f, a_i)$) is inaccurate and uncertain. The sensitivity of model variance to additional experience and exploration is high—the marginal information gain is also high. Additional exploration causes model variance to decrease at a diminishing rate until it stabilizes. When $\mathcal{H}$ decreases below a threshold ($\mathcal{H} < th$), we assume no more information can be gained, and the affordance is habituated. Therefore, the context represented by a catalog is no longer compelling and the agent is well served (cognitively) by attending to other contexts. This metric allows contexts with more variation to be explored more. In Section 4.3, an example is given to demonstrate how this mechanism can be applied to drive the robot's quest to learn an affordance-based kinematic model of humans.

4.2.3 A Hierarchical Affordance Representation for Complex Objects

This section further extends the affordance modeling technique done in conjunction with Hart and describes how the learned affordance catalog can be formulated in terms of a principled probabilistic framework for organizing affordances hierarchically and for facilitating recognition. As the robot learns a collection of affordances, contextual distinctions can be resolved into finer and more discriminative models. For instance, human beings initially (and conspicuously) afford the visual tracking of large scale motion cues. However, over time, and as more nuanced behavior with humans is acquired, the human "context" grows to include the appearance and motions of smaller body parts, sounds and activities. To encode affordance hierarchically, this work adapts a probabilistic framework of the And-Or Graph proposed by Zhu et al. from image scene segmentation [121, 72] for the use of affordance modeling.

In Zhu's approach, objects are modeled as a hierarchy of parts, that forms a parse graph structure. As shown in Figure 4.3, a bicycle can be divided into a frame, wheels a seat, while a wheel can be divided into smaller parts such as spokes and tire. The And-node represents a decomposition of an entity into its parts, while the Or-nodes act as switches for alternative sub-structures. The horizontal connections between nodes encode relations and constraints. To adapt this framework to describe hierarchy of affordances, we define each node in the AOG parse graph as an affordance learned using existing behavioral programs.

More formally, the Affordance And-Or Graph can described as tuple:

$$G =< S, V, R, P >$$

where S is the root node, V are nodes that describe affordances of an operating context, R is a set of observed relations between parts, P is the probability model of the graph. The probability of each node v_i on the graph can be recursively computed

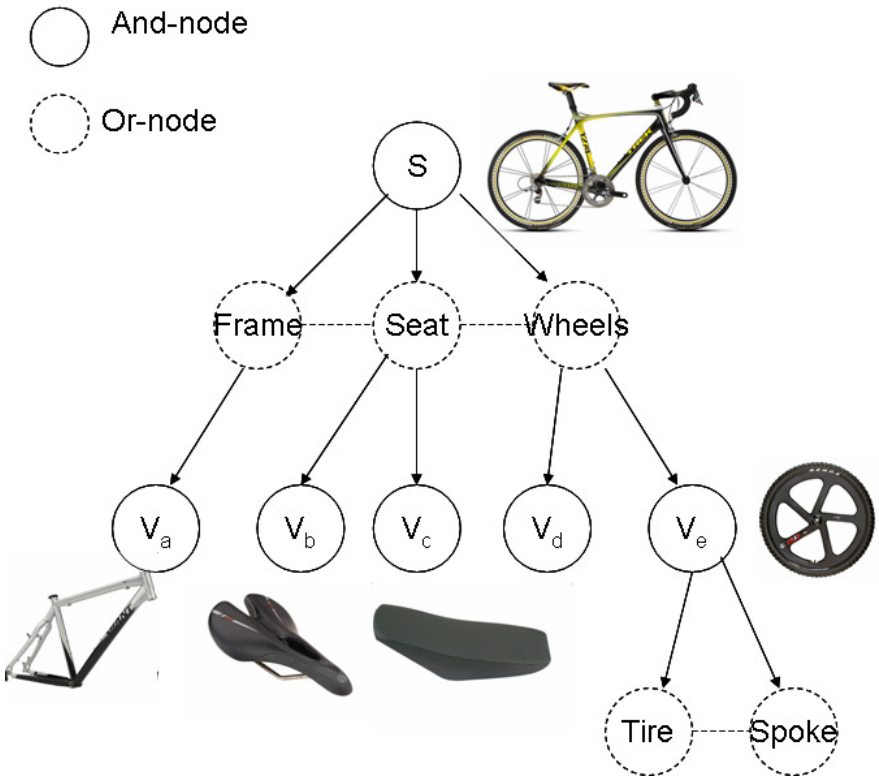


Figure 4.3. A bicycle can be decomposed and represented as a hierarchy of smaller parts using an And-Or Graph image grammar framework. This figure is adapted from [121].

as a product of its N child nodes (v_{ij}, where $j \in \{1, ...N_i\}$) according to the structure of the graph.

$$P(v_i) = \prod_{j=1}^{N_i} p(v_{ij})$$

where each p corresponds to the affordance probability $Pr(r|f, a)$ described in Section 4.2.1. The relational constraints between affordances in the AOG formulation is modeled as a Markov Random Field (MRF) where a probability is computed on the multi-feature affordances discovered by the robot. For instance, the human simultaneously affords tracking the movement of the torso feature and the head feature in a kinematically constrained manner. This kinematic relationship can be encoded in the form of an energy function $E(G)$ in the MRF formulation:

$$\begin{aligned} P(M) &= \frac{1}{Z} e^{-E(G)} \\ &= \frac{1}{Z} e^{-\sum_{<i,j>\in V}(\psi(v_i, v_j))} \end{aligned}$$

where $\psi(v_i, v_j)$ denotes a pairwise relationship of the multi-feature affordance, and Z is the standard Gibbs normalizing partition function. The probability for the entire parse graph is then the product of the two,

$$\begin{aligned} P(G) &= P(V)P(M) \\ &= (\prod_{j=1}^{N(S)} p(v_j)) \frac{1}{Z} e^{-E(G)} \\ &= \frac{1}{Z} (\prod_{j=1}^{N(S)} p(v_j)) e^{-\sum_{<i,j>\in V}(\psi(v_i, v_j))} \end{aligned}$$

After the model is learned, the recognition process consists of finding the collection of features and affordances that maximizes the P(G), s.t.,

$$V^* = argmax_v P(G)$$

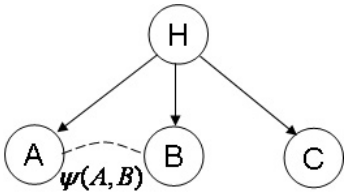


Figure 4.4. This figure shows a hypothetical 2-level hierarchy of a simple human affordance model. The root node H is a random variable that represents a human. The human in this case affords 3 behavior, encoded as random variable A, B and C. The relation constraint between affordance A and B is encoded in the joint distribution $\psi(A, B)$.

The following is an example illustrating how this formulation is applied for the purpose of representing human affordances in a hierarchical manner. Figure 4.4 shows a hypothetical 2-level hierarchy of a simple human affordance model. The root node H is a random variable that represents a human. The visual appearance of a human can be decomposed into 2 parts: an upper-body (encoded as random variable A) and a lower-body (random variable B), both afford tracking. The kinematic constraint of the upper-body and the lower-body is encoded in the joint distribution $\psi(A, B)$. Assuming other than visual tracking, the human also afford another behavior, for instance, a beckoning gesture, encoded as random variable C. Using the affordance modeling technique described earlier in this chapter, the individual affordances $P(A)$, $P(B)$, $P(C)$ and relational affordance $\psi(A, B)$ can be modeled. Thus, the overall distribution for variable H can be computed using:

$$P(H) = P(V)P(M) = \frac{1}{Z}[P(A)P(B)P(C)]e^{-\psi(A,B)}$$

4.3 Case Study: Incremental Modeling of Human Affordances

To verify the proposed approach to incrementally learn affordance models of humans, we employ a bimanual upper-torso humanoid robot, Dexter. We will show that Dexter can learn increasingly complex affordance models of humans by action exploration. The learning occurs in a number of stages, as follows.

4.3.1 Stage 1: Learning Human Motion Affordances

This section describes how the robot acquires a preliminary model of humans in the environment using a sequence of simple staged learning episodes. This is motivated by the development of human infants where maturational constraints dramatically influence the incremental complexity of learning about open interactions with unstructured environments.

In the control basis learning framework, humans represent an operational "context" that is modeled in the same fashion as the many other contexts that exist: by allowing the robot to explore and find actions associated with perceived sensory features that lead to reliable reward. To facilitate learning about humans, we initially bias the robot's visual sensing to be selectively sensitive to certain types of features, e.g. regions of coherent large motion in the environment. This is similar to the maturation process of a human infant where the infant's vision is initially rather primitive and is only responsive to large regions of motion and brightly colored or high-contrast objects.

The regions of coherent motion are computed using a persistent backgrounding technique [22] that enables robots to build background models of the environment. The background model is constructed and updated when no one is in the room, by stitching together snapshots of the scene as the robot randomly scans the surrounding. Using this background model, foreground motion can be segmented and the robot can thus track foreground motion generated by human movement. Although this is

a coarse feature and can be ambiguous if objects such as chairs or tables are moved before the background model is updated. However, the purpose of this work is to demonstrate that robots can learn useful behavior for interacting with humans using such coarse, ambiguous features. Behavior and the associated models can be used to reduce uncertainty as well as to solicit help from humans in the environment. More importantly, models acquired can also be used to bootstrap the learning of more complicated models of humans and interaction.

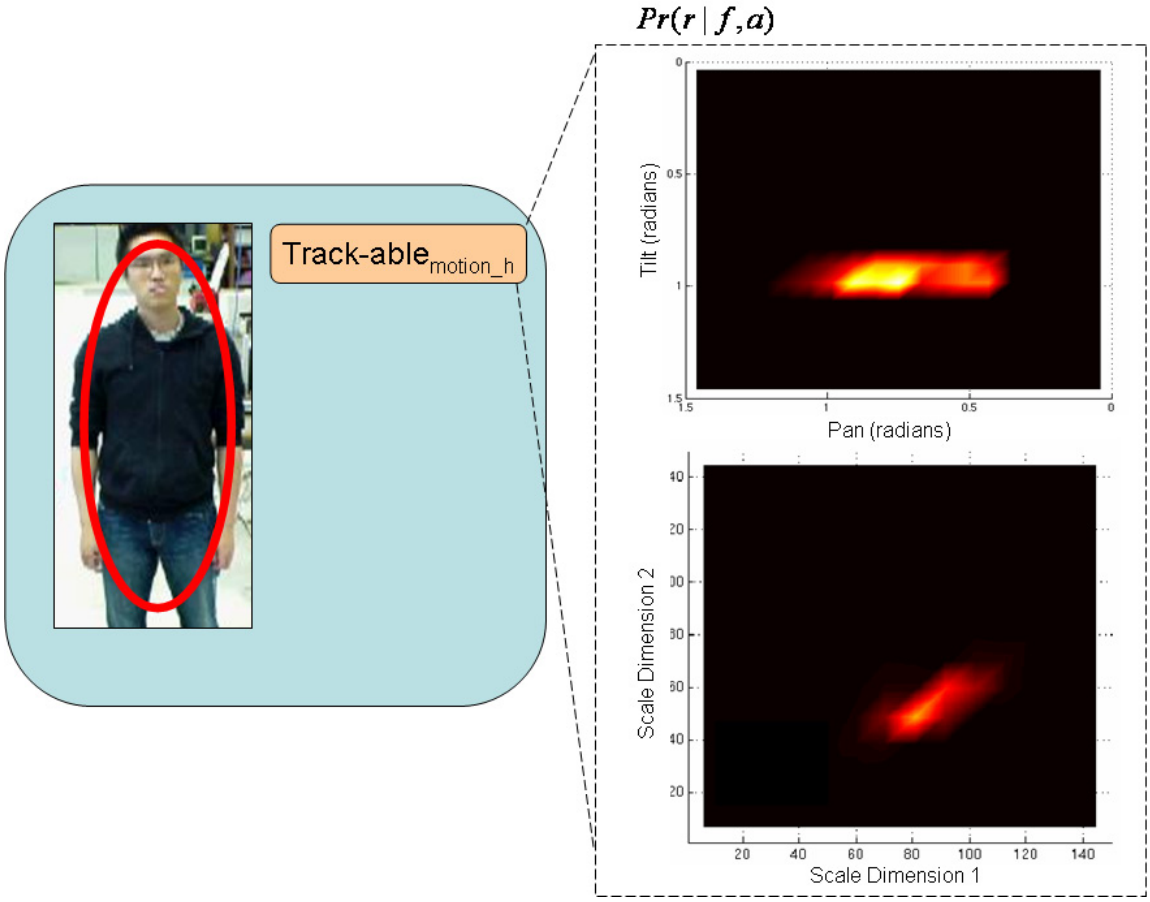


Figure 4.5. The human affordance model after stage 1. The model contains a Trackable affordance, i.e. the probability of reward $Pr(r|f,a)$ given the SEARCHTRACK action. The top distribution shows where motion can be successfully tracked in pan/tilt space. The brighter of the pixel, the higher the probability of reward/success. Similarly, the bottom distribution shows the scale property of the tracked motion feature.

Due to developmental staging, the robot's choice of actions is at first limited—the only affordance it can explore is at this stage is SEARCHTRACK action (Section 3.4), applied to regions of rigid body motion. Dexter is rewarded for successfully tracking these features and such cues are available when humans are active in the environment. Hence, humans are *track-able* and this affordance is added to the catalog (Figure 4.5). Dexter collects samples and uses them to construct the affordance probability distribution $Pr(r|f, a)$ (modeled as Gaussian distribution), i.e. the probability of reward given the action that is configured to search and track f_{motion}. In this case, $f_{motion} = \{f_\theta, f_s\}$, where f_θ is the pan/tilt configuration of the stereo head of when the motion is tracked and f_s denotes the spatial scale of the tracked motion. These probability distributions (shown in Figure 4.5) reflect Dexter's primitive concept of humans: a) the pan/tilt dimensions of the motion feature show that human motions are not likely to be found on the ceiling, nor low on the floor; b) human motion exhibits a distinct distribution in scale space (human-scale motion).

4.3.2 Stage 2: A Kinematic Model

This section describes how the robot continues to refine its model of humans using the same intrinsically motivated behavioral learning framework. In this stage, the robot is allowed to explore color channels for possible kinematic relations. The hue, saturation and intensity (HSI) color spaces are discretized into 18 channels of hue, 10 channels of saturation and 10 channels of intensity (as described in Section 3.2). An example output from the sensory processing pipeline is shown in Figure 4.6. These features are coarse and independently produce an ambiguous summary of the scene. However, this work shows that with a robot capable of configuring controllers to actively attend to features to explore their potential for generating reward, useful structures can be extracted. Moreover, by using knowledge acquired from previous stages, e.g. humans are large motion segments, the robot can now focus its exploration

in regions where more rewards are likely to be found, rather than wasting exploratory actions on background features. In this stage, to facilitate learning, the robot is constrained to only use its head degrees of freedom.

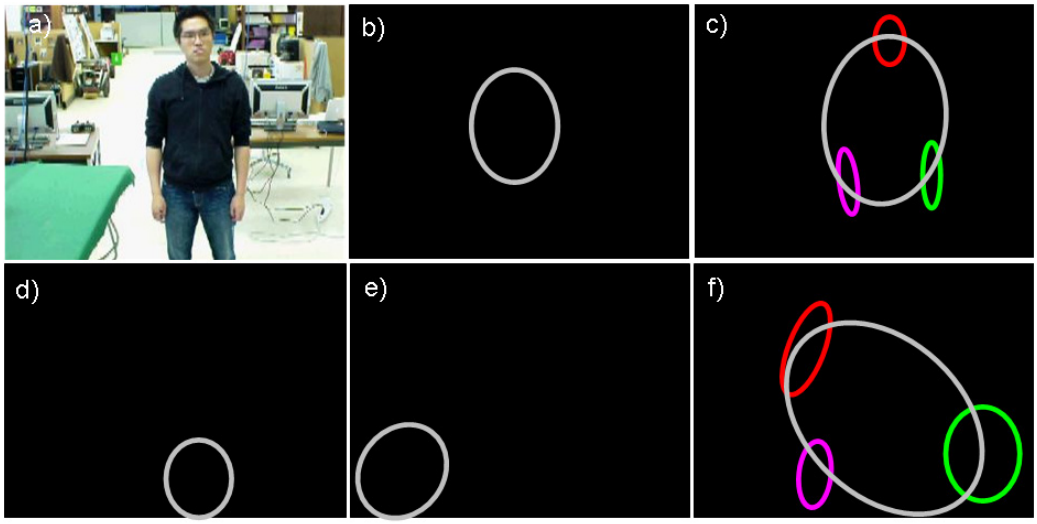


Figure 4.6. Example output of the pipeline (described in Chapter 3). Given a scene from a naturally cluttered lab environment, shown in top left, panels *b*) through *f*) show the output of several channels where segment blobs are tracked. Panels b) and d) correspond to clothing segements the subject is wearing (black jacket and blue pants), c) is a skin color channel where the face and two arms of the human are visible. Channels where the table and the floor show up, are in e) and f) respectively.

Given the enhanced sensor resources, the constrained effector resources, and the reward model described in Section 3.3, the robot is intrinsically motivated to build increasingly deep knowledge structure of complex objects. This is because according to the reward model, the robot receives 1 unit of reward when it discovers a single feature that can be tracked, and an additional unit of reward for finding an additional feature that can be added to its memory structure, the affordance model catalog.

As humans move about in front of the robot, the robot is first attentive to the already familiar human-scale motion (from stage 1). To discover more features that are associated with the motion, the robot samples a feature from the output of the visual processing pipeline (Figure 3.2), and attempts to gather information to ver-

ify the sampled feature's relationship with respect to the motion feature. This is accomplished by using the composite tracking controllers (the only type of tracking controllers valid for handling two or more features at once) to keep both features in view for a period of time, giving the robot the opportunity to gather enough data to both verify their relationship and build probability distributions for describing the relationship. The composite tracking controller is constructed using principles of nullspace composition as described in [40].

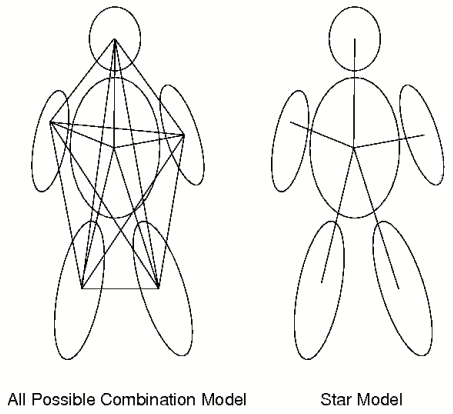


Figure 4.7. A fully connected feature relation graph (left) and a star model (right). Using the star model (right), in which the position and scale distribution of each feature is encoded with respect to a reference feature, in this case, the torso of the human.

To discover features that can be tracked simultaneously and model feature relationships, one possibility is to do so for all pairs of features (Figure 4.7). In this case, for simplicity and computational efficiency, we choose a star-shape model, where we assume that there exists a reference feature with respect to which all other features can be located. The star-shape model also provides a more basic kinematic relations than non-adjacent segments. As a result, the modeling process is simplified such that only the relationship between the reference feature and other features need to be modeled. In this work, the reference feature is the first feature found to be a part

of the original human-scale motion feature, which in this set of experiments, is the feature that corresponds to the torso of the human. The exploration necessary to discover kinematic relations is motivated by the intrinsic reward function, and the duration of the exploration is determined by the habituation process as described in Section 4.2.2. Thus, Dexter uncovers features that are part of the human one by one and the affordance (control configurations) and kinematic models (probability distribution functions) learned are added to the affordance catalog model.

Using the SEARCHTRACK behavior a_{ST}, the robot simultaneously track both the torso feature and features on other parts of the human. Thus data is gathered for the affordance models. Figures 4.8 and 4.9 show two examples of these models. They depict kinematic affordances of the legs and head respectively, relative to the trunk of human. As shown, they remain approximately fixed throughout. Therefore, the change of variance for the model distribution $Pr(r|f_x, a_{ST})$ converges very quickly (see the information gain plot in Figure 4.8 and 4.9). Thus, motive to observe the relation between these visual segments habituates and this behavior is added to the human affordance model. After which, exploration is directed to other features in search of the rest of the human catalog.

The learned kinematic models for the two arms are shown in Figure 4.10. The rate of change of the model variance also drops over time. When the $\mathcal{H}$ value drops below a threshold, the modeling process habituates.

Compared to the head and legs features, the attention to the relative pose of the arms and torso habituates with a significant variance in the relative position. The large variance indicates the arm feature indicate either a non-rigid connection to the reference torso feature, or no connection. For these features, the potential kinematic relationship can be verified with further observations. For instance, the torso and the arm is connected via the shoulder joint. Therefore variance can be reduced if the observation is made between the joint and the torso, or between the arm and the

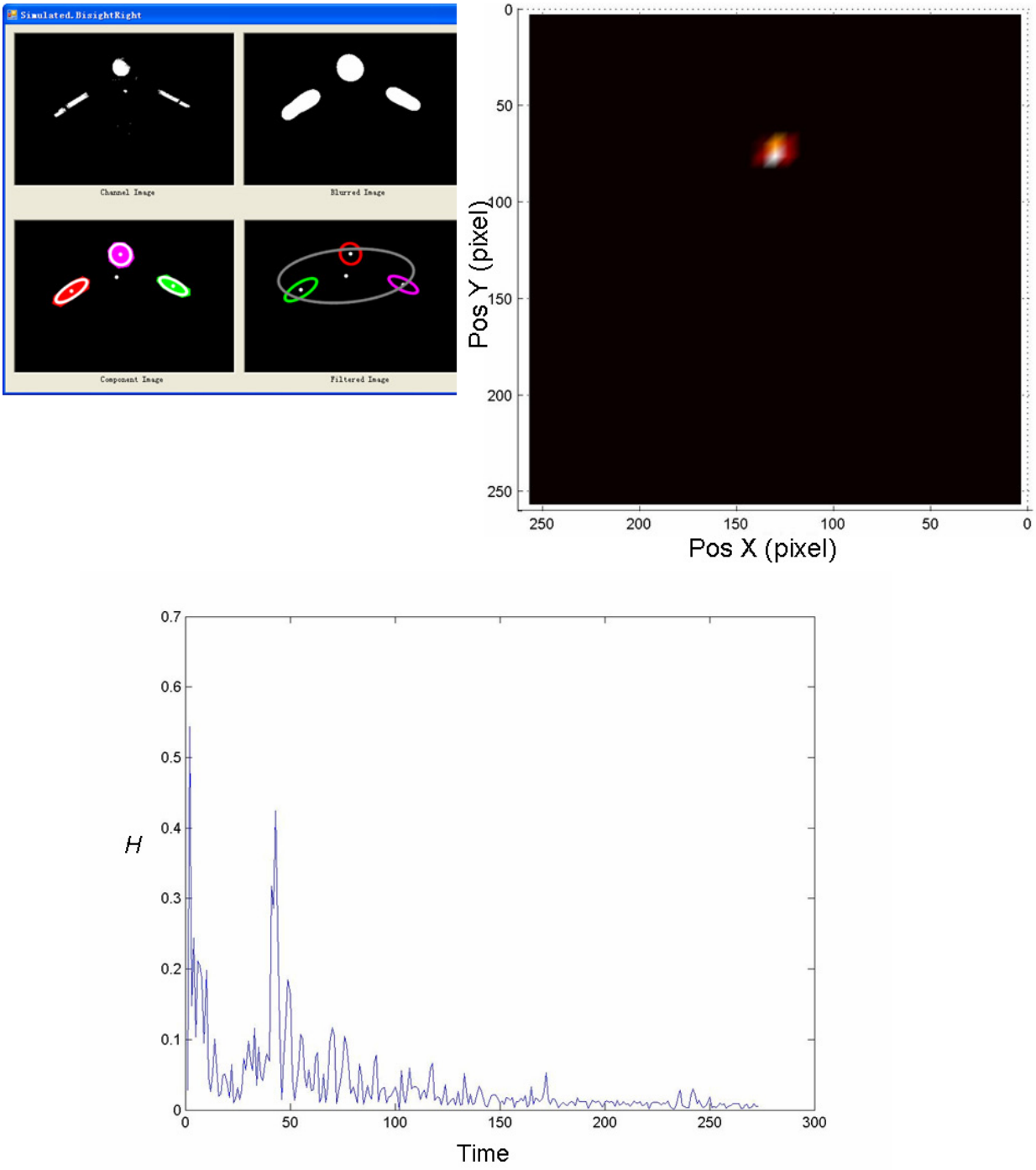


Figure 4.8. The kinematic relations between features associated with legs and torso. The top-left figure shows the pipeline's 4-stage process of a color channel. The top-right figure shows the modeled distribution $Pr(r|f_x, a_{ST})$—the likely relative position where the legs can be found and tracked, given the torso position. The bottom figure shows that as more data are gathered and added to the model, the $\mathcal{H}$ metric gradually decreases and the intrinsic motive to observe this relationship habituates.

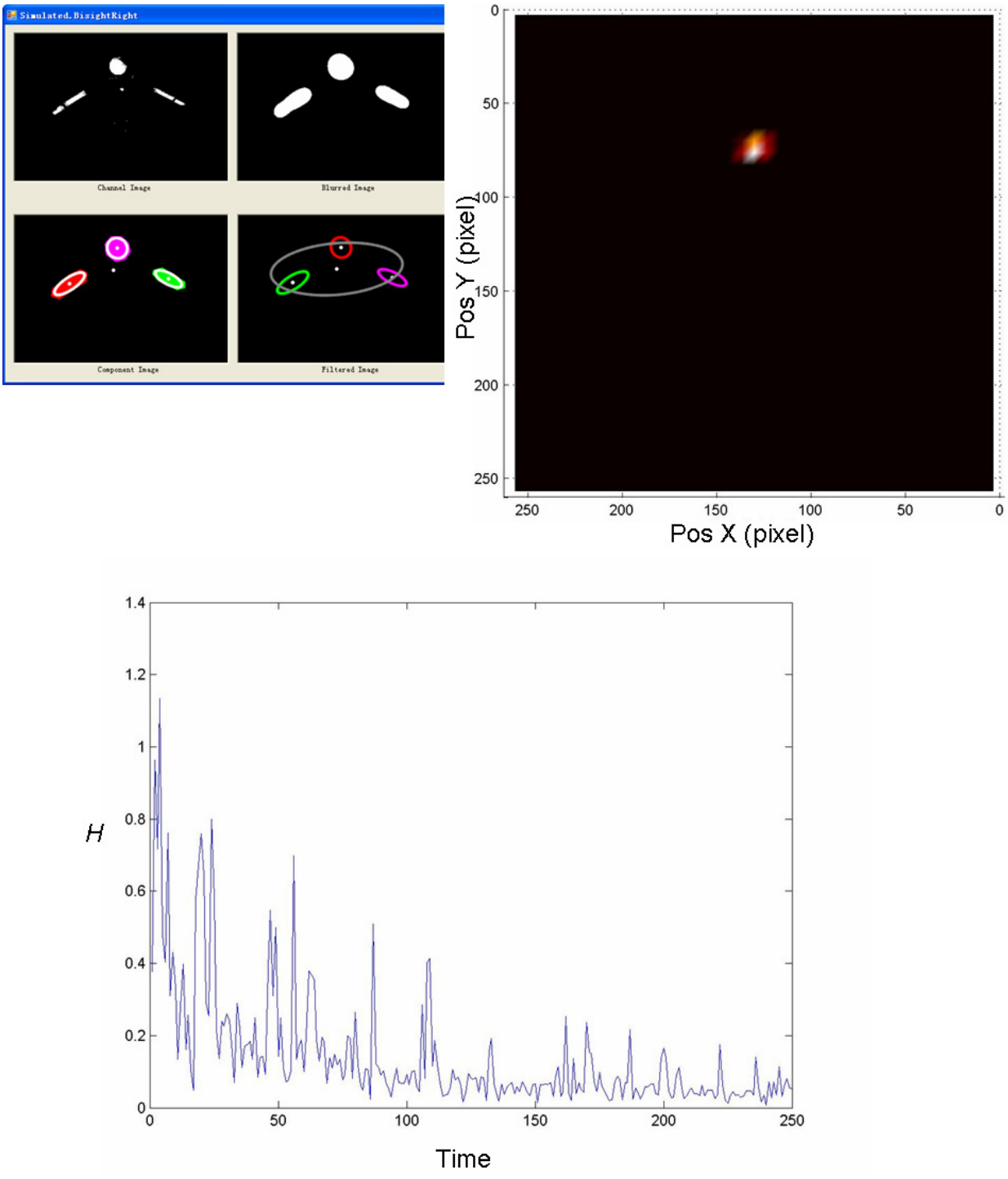


Figure 4.9. The affordance model of kinematic relation between features associated with head and torso. Figure shows the modeled distribution $Pr(r|f_x, a_{ST})$—the likely relative position where the head can be tracked given the torso position.

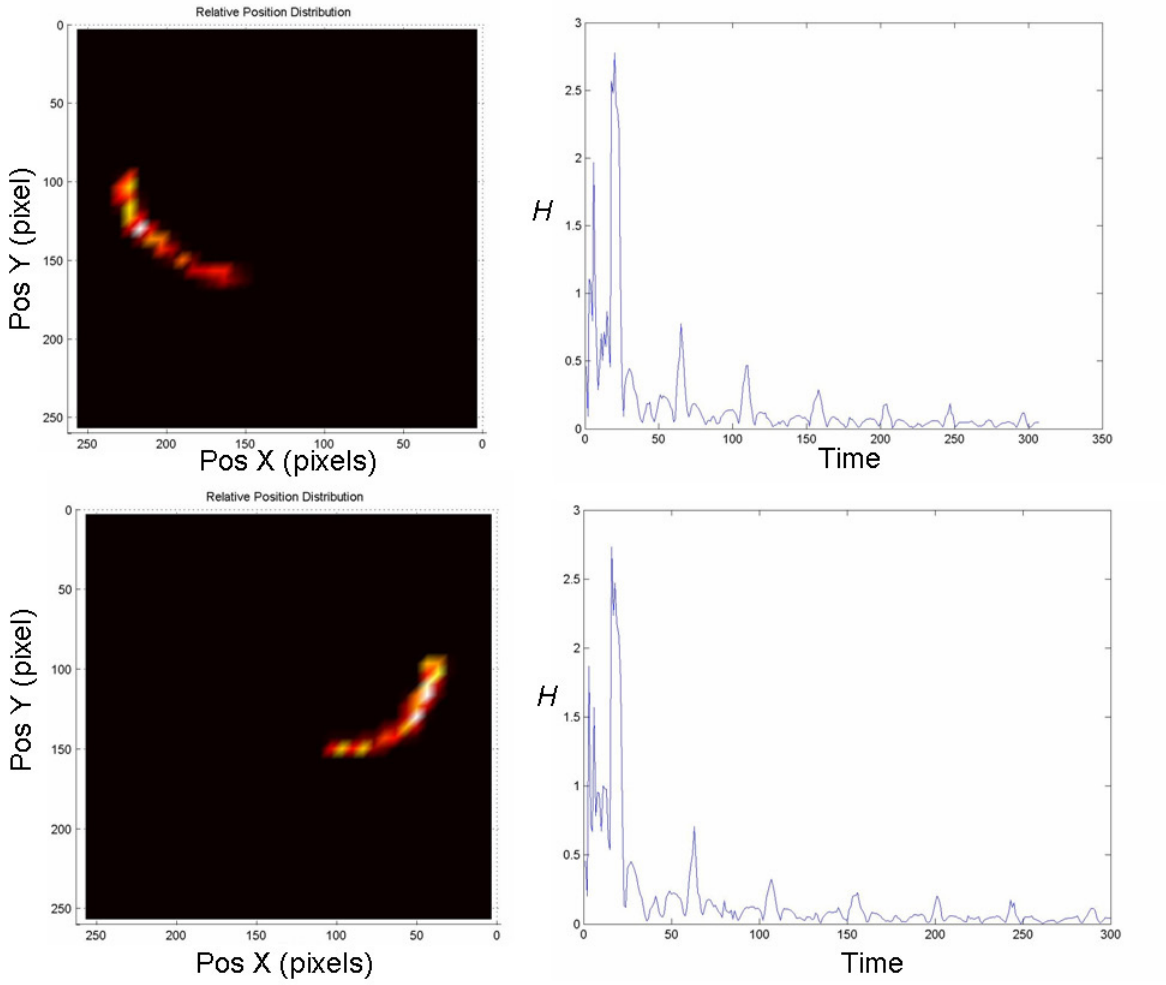


Figure 4.10. The affordance model of the kinematic relation between features associated with arms and torso. Figure shows the modeled distribution $Pr(r|f_x, a_{ST})$—the likely relative position where the arms can be tracked given the torso position.

joint. The position of the shoulder joint can be estimated (Figure 4.11) from the arm motion, using the standard Hough transform voting algorithm [101]. Over time, from a series of estimated shoulder joint positions, the relative position of the shoulder joint position with respect to the reference torso feature can be modeled.

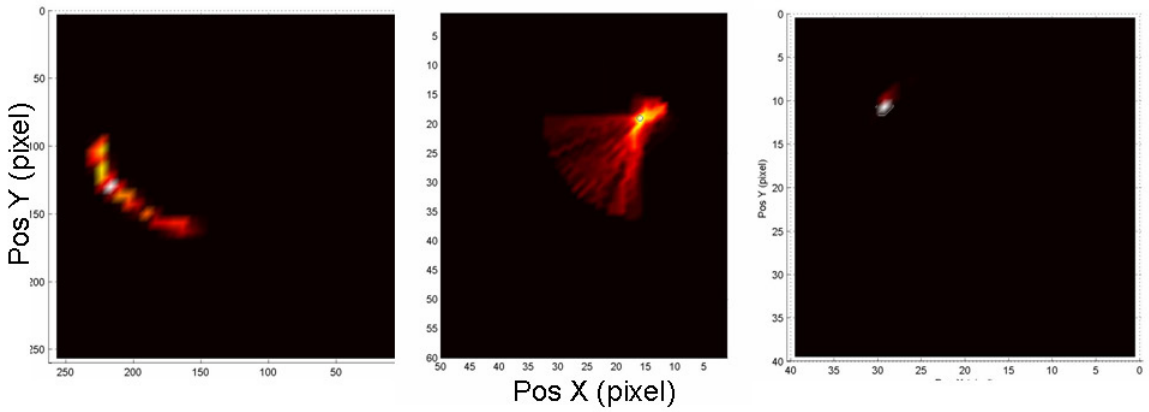


Figure 4.11. Estimating the shoulder joint: a) motion trajectory of the arm feature, b) using a Hough transform voting algorithm, the relative position the shoulder joint can be estimated (the brightest spot), c) a low variance relative position model for the shoulder joint.

Although we have limited our feature sampling to features within the foreground motion feature, this is in fact not necessary. The above mentioned modeling method can even handle features from the background, since the relative position model of these features and any feature on the human will maintain a large variance that cannot be reduced, regardless of how the feature relationship is modeled. For instance, shown in Figure 4.6, the table that is not part of the human can also be tracked along with the human feature. However, when tracking both features, the feature that corresponds to the table moves in a kinematically independent fashion compared the motion of the human. Result shows that for this case, the variance in relative motion is high and cannot be reduced. Therefore, it can be determined that this feature is not a part of the human.

At the end of this stage, the human catalog (Figure 4.12) is further augmented with tracking affordances that forms a basic kinematic description of the human body. The kinematic model includes feature distributions for describing different parts of the human, such as head, torso, arms and legs, in terms of their relative position and scale attributes.

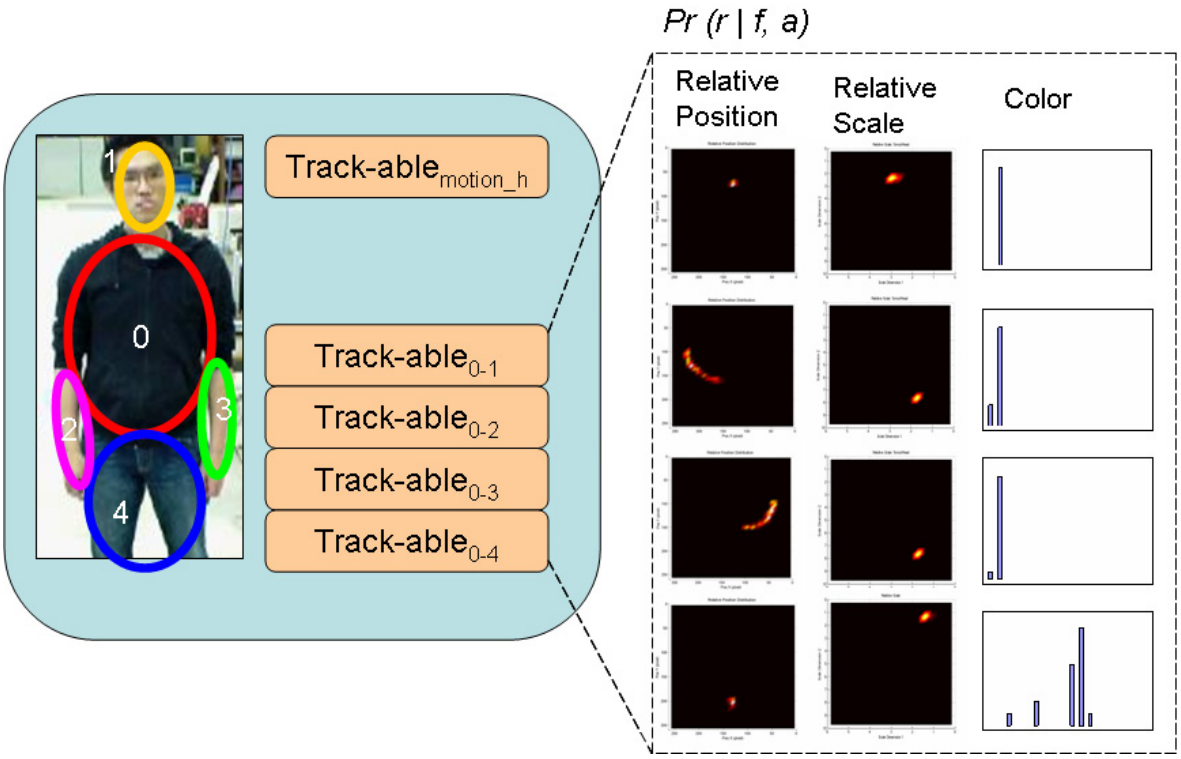


Figure 4.12. The extended human affordance model after stage 2. The robot discovers several new track-able affordances associated with the finer features and estimates distributions that describe the kinematic relationships between different parts of a human body.

4.4 A Hierarchical Behavior Representation using And-Or Graph

Using the Affordance And-Or Graph formulation described in Section 4.2.3, the learned human affordance catalog model can be transformed into a hierarchical tree

structure (Figure 4.13). Using this representation, each recognized affordance can provide probabilistic evidence for finding and tracking humans in a principled manner (Section 4.2.3). This layout clearly shows the coarse-to-fine progression of the modeling process. At the top of the tree is the human root node. One level below, it shows the human can be tracked using whole-body motion. Furthermore, it shows that human object also affords tracking using kinematic structure information where it is decomposed into simultaneous tracking affordances of smaller parts. The horizontal links encode the pairwise affordance relation between the body parts. As we will see in the next chapter, this affordance model will be further enhanced with behavioral affordance that extends beyond visual tracking.

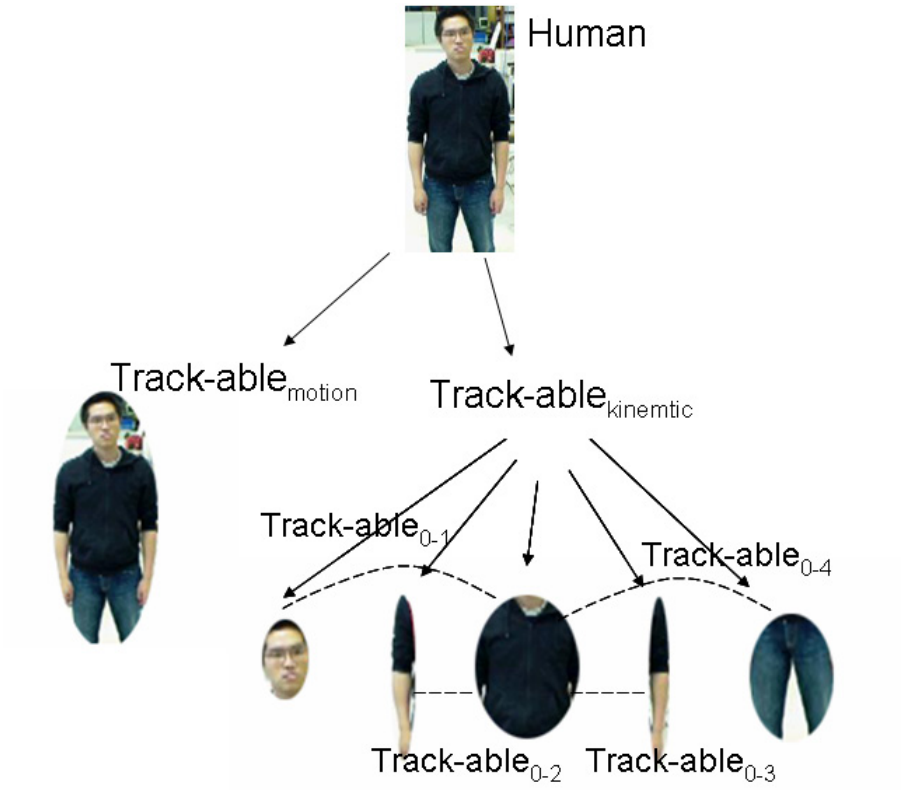


Figure 4.13. The hierarchical And-Or Graph of learned human affordances after the first two stages. The hierarchical will be extended in the next chapter.

This model is indirectly evaluated for human detection in the human-robot interaction study discussed in the next chapter (Chapter 5). Each subject interacted with the robot for about 20 minutes. As shown in Figure 4.14, in a naturally cluttered scene under various lighting conditions, the probabilistic hierarchical formulation enables the robot to detect humans and track their different body parts in an robust manner. Furthermore, the algorithm is able to maintain track of the body parts even though sometimes the part may disappear and they reappear at a later time (Figure 4.14, second picture in the first row).

Figure 4.14. Examples of multi-body tracking of humans using the learned affordance human catalog model. Note that the probabilistic approach enables the algorithm to maintain a robust track of the human under partial occlusion or unstable feature conditions (as shown in the second picture in the top row).

4.5 Discussions

In summary, this chapter presented a novel affordance-based approach of modeling humans, where a robot's understanding of humans is learned and represented in terms of behavior they afford. A framework for learning affordances in an incremental manner is presented. To demonstrate the feasibility of the approach, experiments have been carried out on a bimanual humanoid robot, and showed that the robot can build an increasingly complex behavioral model of the humans it interacts with, in incremental learning stages. In the initial stages, the robot's first learned concept of a human is simply a motion segment of a certain scale that affords tracking. In the subsequent stage, the robot is able to further extend the human affordance model and begin to pay attention to individual body parts and learns their corresponding affordances via visual tracking. Lastly, for organizing the learned affordances, a probabilistic Affordance And-Or Graph representation is presented and preliminary results are shown where the learned affordance model allows humans to be detected and tracked under naturally cluttered environments.

The incremental learning process allows the robot to build models of complexity as required by the context. We believe multi-resolution models of humans are useful because simple models enable robots to learn rewarding behavior that in turn leads to a richer model. In the next chapter, we will discuss how a robot, using hierarchical manual behavior as the basis of learning, can acquire behavior for conveying intentions to a nearby human (detected using models learned in this chapter). Moreover, we will also show that as the robot acquires new behavior for interacting with humans, using the technique and representation discussed in this chapter, how the robot's concept of humans extends beyond simple visual tracking to potential resources that can be "actuated."

CHAPTER 5

EMERGENCE OF EXPRESSIVE BEHAVIOR FROM MANUAL BEHAVIOR

I am interested in how a robot can learn communicative behavior from direct interactions. In this work, interactions with humans via communicative actions is learned in the same manner as the robot learns to interact with other objects in the environment. Furthermore, I am also interested in the variety of gestures that emerge from natural interactions with human partners, as different people may respond to gestures differently—a learning framework may produce some unexpected results.

Many advanced machine learning algorithms are best suited for offline processing of large datasets or simulation runs that generally require tens of thousands of training episodes [112]. For the domain of human-robot interaction (HRI), this is particularly problematic since in order to acquire training data, a human needs to be present. Tens of thousands of training episodes is out of the question. There has been a great deal of work devoted to reducing the training time by teaching with demonstration. Optimizing low-level motion trajectories to achieve tasks such as performing pole-balancing[3], batting a table-tennis ball, or catching table-tennis ball in a cup have been demonstrated. Similar work has focused on teaching robots to produce gestures, but again they either treat gesture learning as a low-level motion trajectory problem [11] or a joint space motor control problem [67]. None of these approaches consider the interplay between the robot and human as part of the gesture learning process: how environmental changes may affect the meaning of gestures, and how the robot can learn to adapt accordingly.

This chapter presents an extension to the control basis framework and attempts to address issues including the origin, adaptivity, and learning efficiency in the development of communicative behavior for robots. The approach formulates adaptive human-robot interaction in the same framework designed to acquire skills for manual (i.e., robot-object) interaction. As a result, gesture learning directly benefits from the ideas of developmental staging, hierarchical learning and skill generalization that already exist in the control literature [40]. We look specifically to communicative behavior that reuses motor skills to convey goals and intentions between a human and a robot partner.

5.1 Adapting Behavior to the Presence of Human Beings

The strategy taken in this work for fostering the emergence of expressive communicative behavior is to introduce more difficult contexts as in the case of developing manipulation behavior. However, for the purpose of developing communicative gesture, we introduce contexts where the robot is underactuated and humans are present. To adapt to the new contexts, our robot Dexter relied solely on local behavior generalization techniques proposed by Hart [39]—if a known schema fails under a new context, attempt to allocate different sensorimotor resources to the failed schema, until the appropriate resources are found.

However, there exist many situations where allocating new resources is not sufficient, for instance, the classic "pick-and-place" task often studied in the robotics literature [53]. Consider a general purpose pick-and-place schema that acquires an object (the "pick" goal) and delivers it to a desired position and orientation (the "place" goal). A successful grasp of the object can depend on characteristics of the place goal. For instance, if the object is a cylindrical peg that is to be placed at the bottom of a cylindrical hole, then the mating surfaces between the peg and the hole must be left unobstructed for the insertion to succeed. The decision about how to

grasp the peg must respect this constraint. Now consider a robot with lots of prior experience with pick-and-place tasks, but none directly focused on the constraints surrounding peg-in-hole insertions. An arbitrary grasp on the peg will likely fail during the place sub-task and the reason for this failure is likely inexplicable in the existing pick-and-place framework. As we will see later (Section 5.3), similar problems exist when the robot attempts to learn gestures to communicate intentions to the human partner. In these cases, both the declarative structure and procedural knowledge must be extended simultaneously for the behavior to be adapted to the dyadic context.

In general, the repair of a schema in response to a new situation can require a larger temporal scope than indicated solely by the actions that fail. The error can be associated with events that are not monitored by the schema and that occurred at some indefinite time in the past. Prospective behavior is an important component of computational approaches to transfer and generalization. It is a term, coined in the psychology literature, to describe a process in which a human infant learns to predict how a strategy might fail in the future and generates alternative strategies to accommodate the new situation.

McCarty *et al.* studied the initial reach to a spoon laden with applesauce and presented to infants in left and right orientations [69]. The developmental trajectory observed is summarized in Figure 5.1. Initial policies are biased toward dominant hand strategies that work well when the spoon is oriented with its handle to the dominant side. However, when it is not, the dominant hand strategy fails. Variations in the applesauce reward distinguish important categories in this process—dominant-side and non-dominant-side presentations of the spoon. One hypothesis holds that this process involves a search for perceptual features that distinguish classes of behavioral utility. When this happens, new perceptual features have been learned that were not present in the original representation. They have been selected from a possibly

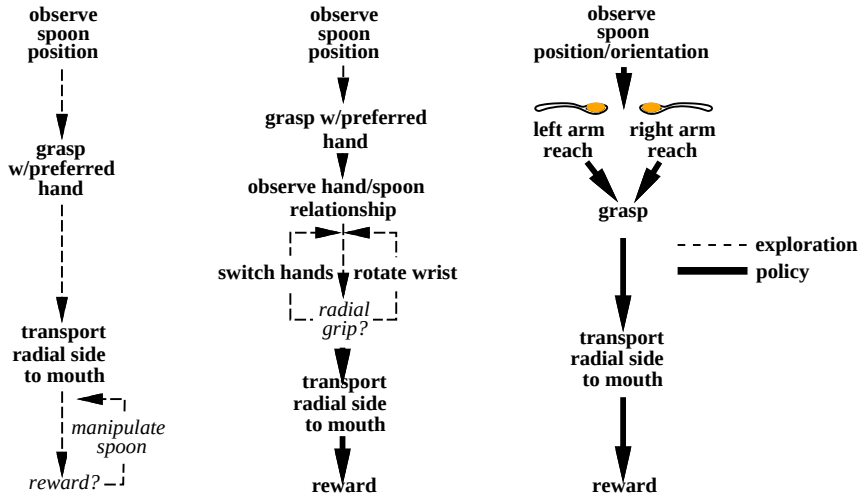


Figure 5.1. Prospective Behavior revealed in the Applesauce Experiment.

infinite set of alternatives because they form a valuable distinction in the stream of percepts—valued for their ability to increase the reward derived from the infant's interaction with the task. One may view this process as one in which properties and constraints imposed by the task are incorporated into a policy incrementally starting with the latter (distal) actions and gradually propagating back through the action sequence to early (proximal) actions.

The applesauce problem and the "pick-and-place" problem share many similarities. However, traditionally in robotics and AI, the "pick-and-place" task is formulated as a planning problem. In [66, 53], a back-chaining algorithm is used that searches backward in time from the desired final state until the initial state is found. This approach requires complete knowledge of the task to begin but does not speak to where that knowledge came from. It is subject to uncertainty introduced by seemingly small inaccuracies in backward chaining predictions compounded over multi-step sequences. Moreover, depending on how task knowledge is represented, this strategy may not share common background (pick-and-place) knowledge with other related tasks.

This is in stark contrast to how the human child would approach this problem. Extrapolating from the spoon and applesauce experiment, we expect that the infant will employ a general-purpose strategy and demonstrate biases that apply generally to the entire class of such tasks. Upon failing with this approach, and only upon failing, will the child search for an explanation for the failure, starting at the peg insertion and backing up to the transport phase, to the grasp, and ultimately to the visual inspection of the peg and hole. Somewhere in this sequence is the reason that the general-purpose strategy doesn't work in this context. Once found, the infant will begin experimenting with corrective actions. Throughout this process, the infant's search for a solution revolves around modifying existing behavior rather than attempting to learn a new strategy from scratch.

The work described herein extends the control basis and presents a prospective behavior repair algorithm for autonomous agents to rapidly accommodate a novel task by applying existing behavior. The main idea of the algorithm is the following: upon failure due to a new context, the robot attempts to fix the problem via local adjustments whose scope expands until a compensatory subtask is learned to handle the exception. Now, the general-purpose schema is extended with a call for the compensatory subtask when the triggering percept is present. The result is a new, integrated, and more comprehensive schema that incorporates prospective behavior for accommodating the new context.

For the rest of this chapter, we will introduce the algorithm for discovering prospective behavior with a simple navigation task with multiple "door" contexts that produce prospective errors. We show that a general-purpose navigation policy in the grid world can be extended with auxiliary percepts and compensatory actions to solve the problem efficiently. We evaluate the proposed algorithm by comparing its performance to that of a "flat" learning problem in which all the required state information is provided *a priori*. Next, we provide a formal description on how the prospective

repair algorithm is adopted in the control basis framework and present a case study where the algorithm is applied to enable our robot, Dexter, to learn communicative behavior in the presence of humans.

5.2 Example: A 2D Navigation Domain Problem

We introduce the prospective repair algorithm by way of a robot navigation task. Figure 5.2 shows a grid world in which a simulated robot navigates through hallways, rooms, doors, and uses buttons to actuate the doors. The circle is the robot's starting position and the triangle represents the goal. The robot's task is to learn a path to the goal, given that a random subset of the doors can be closed at the beginning of each training episode. The buttons for opening doors are scattered in different rooms of the map. The robot has to visit the appropriate buttons to open doors that blocks a known path to the goal.

The robot can move left, right, up, or down. At each grid location, the robot can observe its (x, y) location and three door status indicator bits that represent the status of three, randomly chosen doors out of the six in the map. However, the correspondence between the doors and the indicator bits are not directly observable. The initial status of the doors is randomly assigned at the beginning of each trial. We will evaluate two solutions to this problem. The first is a flat learning approach informed by the full state description, and the second is the proposed prospective repair approach using a sequence of reusable policies in the (x, y) state space with prospective error suppression triggered by the door status indicators.

5.2.1 A Flat Learning Approach

A flat learning approach to the problem is formulated where all the required state information is provided *a priori* and the task is presented to the robot in a single learning stage. This is in contrast to the multi-stage learning approach that is

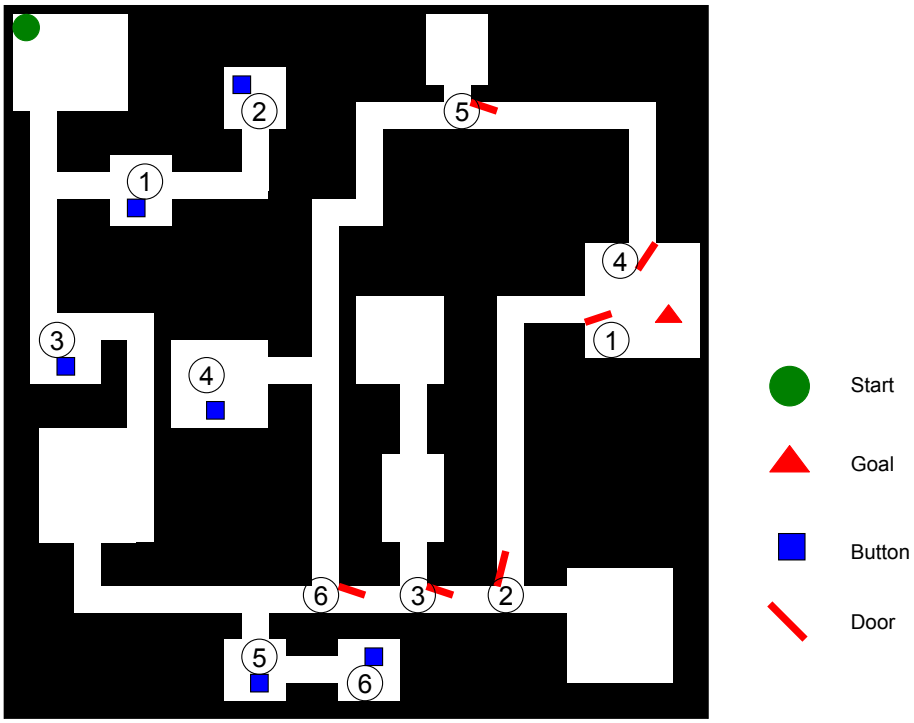


Figure 5.2. A 30×30 grid-world navigation problem. The status of a door is toggled when the robot visits the grid location where the corresponding button is located.

presented next. This grid world navigation task is formulated as a standard reinforcement learning problem using the ϵ-greedy Q-learning algorithm [112] where the robot is rewarded for finding an optimal path to the goal. The state, s, for this formulation includes the (x, y) location of the robot and the 3 observable door status indicator bits. The 4 actions: move up, down, left and right, form the robot's the action set $\mathcal{A}$. A simple reward model is applied: the robot receives 1 unit of reward for achieving the goal and -0.01 units of reward for every step it takes.

In this formulation, the robot receives maximum cumulative reward by taking the fewest number of steps to reach the goal. For every state s the robot encounters and every action a the robot can take from that state, an expected future reward value, or Q-value is estimated. In the beginning, this value is initialized randomly for

every state-action pair $< s, a >$. Through trial-and-error exploration, the Q-learning algorithm enables the robot to incrementally update the Q-value for every $< s, a >$ it encounters. With sufficient exploration, the Q-value for all $< s, a >$ is expected to converge, thus allowing the robot to extract optimal policies for navigating to the goal under all contexts. For these experiments, we define an episode to be one complete traversal by the robot from start position to goal position. Early on, it may take several thousand actions to get to the goal. A trial is defined as one complete learning experiment (until asymptotic performance). Depending on the problem design, it may take several thousand or tens of thousands of episodes before a trial concludes.

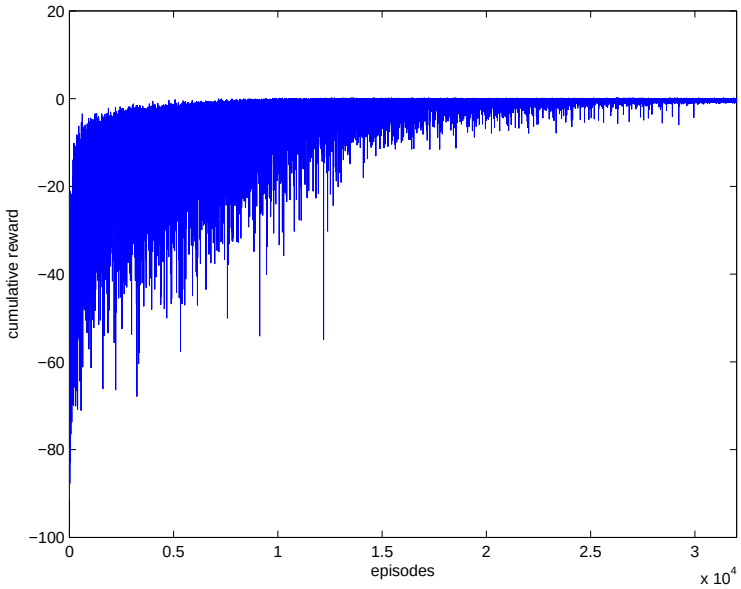


Figure 5.3. Average cumulative reward over 100 trials for using a flat learning approach

The result from the flat learning experiment is presented in Figure 5.3. In the early episodes, the cumulative rewards are large negative numbers because the robot starts out with no prior knowledge about the world, and randomly explores the map with many extraneous steps, building up large negative reward before finally reach-

ing the goal. Slowly, as expected future reward estimates for each state-action pair improve, the number of steps it takes for the robot to reach the goal decreases. As a result, the cumulative reward rises, until it converges at around $30,000$ episodes. This experiment used a discount factor, $\gamma = 1.0$, learning rate $\alpha = 0.1$, and the ϵ-greedy parameter is set to $\epsilon = 0.1$.

The flat learning approach learns to solve this problem in $30,000$ episodes to learn a policy with contingencies for random door configurations. This is a lot of training for an on-line learner, but further reflection on the experiment yields insights that can be used to reformulate the problem. State s includes the (x, y) location and 3 randomly selected door status bits at each cell in the map. However, in many states, the part of s concerning door status is uninformative and optimal decisions can be determined from (x, y) alone. Therefore, performance in the flat learning problem is often compromised by too much state that is encoded inefficiently. In these states, a more general strategy can be applied and much less training is required. To overcome this problem, the hierarchical prospective repair approach is proposed.

5.2.2 A Prospective Learning Approach

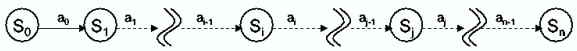
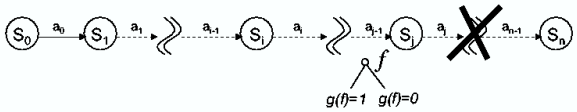
In this section, the proposed prospective repair approach is presented in the context of the multi-door navigation problem. In contrast to the flat-learning approach, the original task is decomposed into a series of problems that can be presented to the robot in an incremental manner. Initially, the robot is presented with the simplest task. Later, it is challenged with more difficult contexts. In the navigation problem, the simplest task is to find the optimal path for reaching the goal when all doors are open. After this policy is acquired, the robot is challenged by closing a specific door until the robot has acquired a policy for handling this case. These skills are reused to construct contingencies for arbitrary door configurations.

The proposed prospective repair algorithm is presented in Algorithm 1. It is divided into 3 main components: (1) a general-purpose strategy is first learned in the simplest context, (2) the robot is challenged with a new context and an auxiliary perceptual feature is learned to differentiate the new context, and (3) a search is conducted for local repairs whose scope expands until a policy is acquired to handle the exception. Algorithm 1 also depicts the schemas created and/or modified after each of these steps. The proposed approach assumes that a general-purpose strategy exists that applies approximately to the different variations in the task. Subtasks are represented as separate policies to preserve the general-purpose policy to remain unaltered.

As shown in Algorithm 1, human guidance also plays an important role in the prospective repair algorithm, in the form of structured tasks of increasing level of difficulty. The simpler task ensures the robot can quickly learn a basic general-purpose strategy while later tasks allow the robot to extend existing policies and learn to handle more complicated contexts. More importantly, such structured tasks can be created by simple adjustments of environmental constraints at an opportune time in the learning process. For instance, opening or closing doors in the robot navigation domain, or offering correctly oriented spoons in the apple sauce experiments. This form of guidance is intuitive to a human teacher as similar strategies can often be observed in human parent/child interactions [69].

Multi-stage training sequences provide for behavior reuse, but they are not sufficient for causing an improvement in learning performance. The appropriate state representation and provisions for re-use are required. This is the key difference between this algorithm and previous approaches to prospective behavior using flat learning algorithms[117]. The global state of the robot, in this case, is represented using only its (x, y) coordinates. The basic policy relies principally on this information and auxiliary state, i.e. door status indicators, are stored separately and only in places where

Algorithm 1 A Prospective Repair Algorithm

TEACHER		ROBOT
• construct a simple initial training context	→ all doors open →	Given a set of percepts: $\mathbf{f} = \{f_1, ..., f_i, f_j, ..., f_n\}$, and actions $\mathcal{A} = \{a_1, ...a_m\}$: 1: Apply factorization technique to define state $s = \{f_1, ..., f_i\}$ where $s \in \mathcal{S}$ contains features that are frequently used for decision making and auxiliary percepts $\mathcal{F} = \{f_j, ..., f_n\}$. 2: Use Q-learning on MDP defined by $< \mathcal{S}, \mathcal{A}, \mathcal{R} >$ to learn a general-purpose policy π, where $\mathcal{R}$ is the predefined reward function for task T. 
• challenge the frontier of existing behavior	→ close single doors →	**recognize the perceptual associations of the sub-task** 3: Execute policy π until it leads to repeated failure and accumulate experience data set, $\mathcal{D}$, recording features $\mathbf{f} \in \mathcal{F}$ and the success or failure of π in that context. 4: Apply a generic discriminative learning algorithm (e.g. $C4.5$) on $\mathcal{D}$ to identify a decision boundary $g(\mathbf{f})$ that differentiates success and failure under policy π. Function g is said to *accept* $\mathbf{f}$ if it predicts success under policy π.  **accommodate the new context** 5: Create a new MDP defined by $< \mathcal{S}, \mathcal{A}, \mathcal{R}' >$, where $\mathcal{R}'$ is a reward for restoring $\mathbf{f}$ to the condition where g accepts $\mathbf{f}$. 6: **for all** states $s \in \mathcal{S}$ in which g does not accept $\mathbf{f}$ **do** 7: Starting from s, learn a compensatory policy π_g for achieving the sub-goal defined by g. 8: **end for** 9: Merge π_g with π to form a new hybrid policy π'.

they are available and needed to trigger contingencies for handling exceptions to the basic plan.

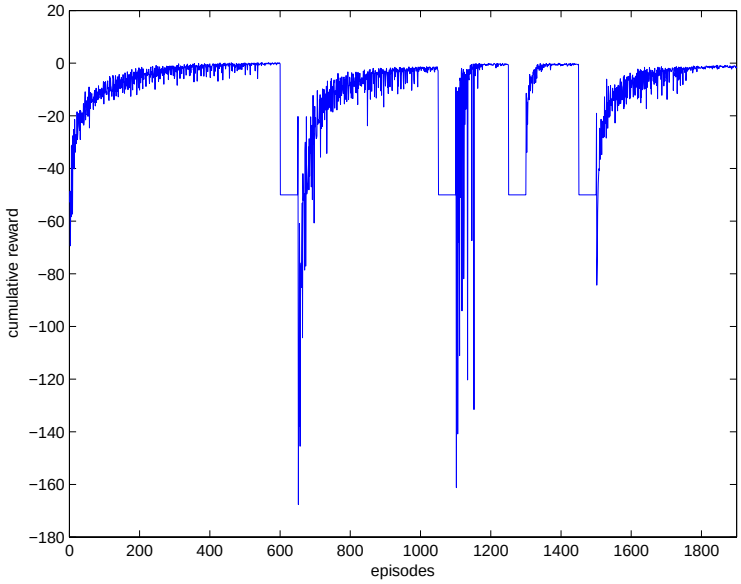


Figure 5.4. Average cumulative reward over 100 trials using the prospective repair approach. Each dip in the learning curve corresponds to a task change that leads to a specific type of failure in the previously learned policy. Results show that the prospective repair algorithm allows the robot to quickly adapt to each new context.

Figure 5.4 shows the resulting learning curve from the prospective repair/generalization approach applied to the navigation scenario. The action set $\mathcal{A}$ remains the same as in the flat learning formulation. Once again, the robot receives 1 unit of reward for achieving the goal and -0.01 units of reward for every action it takes. The learning parameters, $\gamma = 1.0, \alpha = 0.1$, and $\epsilon = 0.1$ remain the same as in the flat learning problem. In the first stage, a path toward the goal is learned with all the doors open. The initial policy, π, for traversing the unobstructed environment is illustrated in Figure 5.5. It depends on (x, y) state information exclusively and serves as the initial general-purpose solution. As Figure 5.4 illustrates, in each subsequent stage, a new context is introduced wherein exactly one of the doors is closed causing the cumula-

tive reward to decline sharply. At this point, a new learning problem is initiated to recognize the new context and to repair the general strategy. Under the experimental conditions described, the reward begins to climb until it converges once again as the robot quickly adapts to the new context. For the particular map used, the closing of some doors do not cause the general policy to fail, therefore there are only 4 dips in the learning curve. The prospective repair process is complete after less than $2,000$ episodes compared to $30,000$ episodes for the flat-learning approach. We can extrapolate these results and conclude that the advantage would be even more significant as more doors are added to the map, or when the robot has to pay attention to more perceptual features.

Figure 5.6 illustrates learned paths to button 1 from any location on the general policy π where the status of the corresponding door can be observed. The path that is the shortest is selected as the compensatory behavior and integrated with the original behavior to achieve a new and more comprehensive behavior.

Several design elements contributed to the performance improvement. First, the choice of the initial state description does indeed provide a policy that serves the task well from many positions in the map—there are only a small number of special cases that the robot must handle. As a result, there is a significantly smaller state-action space than there is with the flat learning approach. All guidance from a human teacher that has this property is expected to produce the same utility in learning performance. Moreover, the search for the prospective behavior is initiated as a separate learning problem with an independent goal and state transition structure, thus enhancing re-use. When multiple doors are closed simultaneously, the prospective repair approach naturally decomposes the original problem into sub-problems associated with navigating to buttons corresponding to closed doors en-route to the goal. The robot can reuse previously learned contingencies for relevant doors rather than having to learn them from scratch as in the case of the flat learning design.

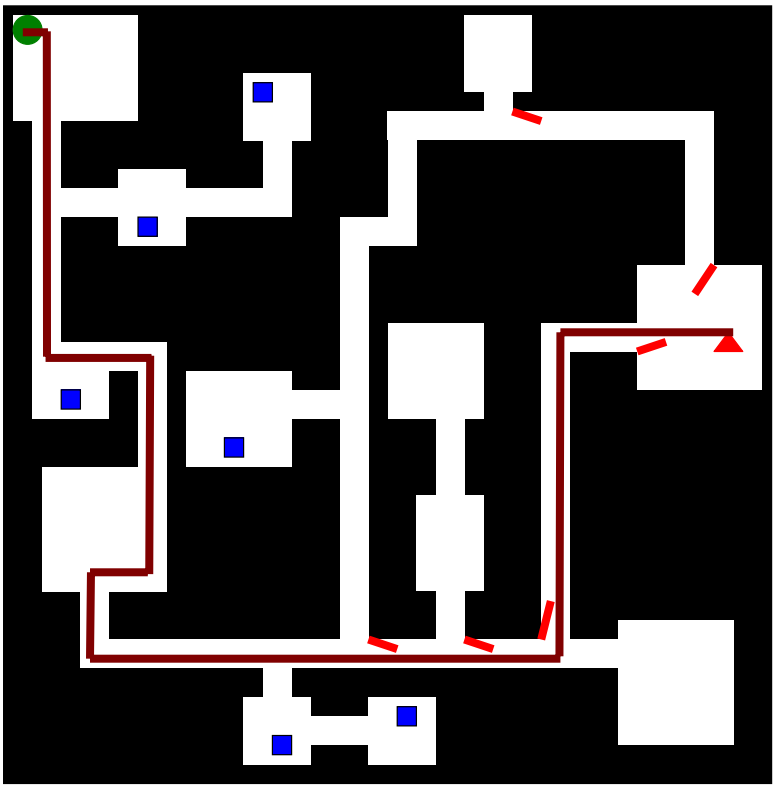


Figure 5.5. Learning result from stage 1: an unobstructed path π to the goal that functions as the general-purpose policy.

5.2.3 Discussion

This work advocates an incremental learning paradigm towards behavior acquisition in robots, where a human user can teach robots skills interactively, using a sequence of increasingly challenging tasks. This is an open-ended process that requires learning framework designers to build systems that can act based on incomplete information and that adapt to new situations where previously learned behavior fails.

In this work, human guidance first comes in the form of training guidance—structuring the environment and focusing exploration on a restricted set of sensors

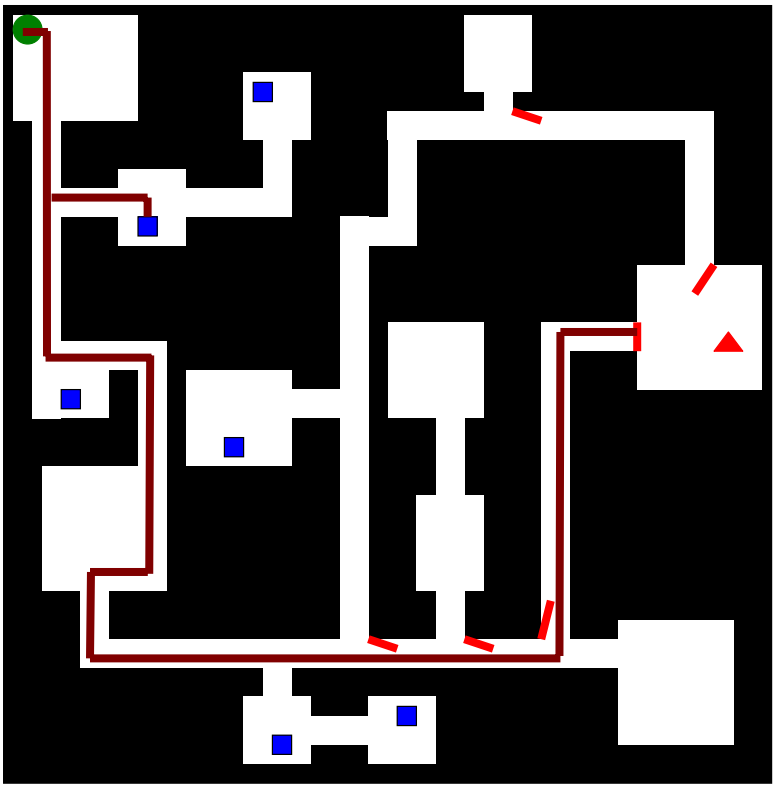


Figure 5.6. Learned paths to the button 1 for opening door 1 from any location on the general policy π where the status of the corresponding door can be observed. By integrating this policy with π, a new, more comprehensive policy for handling the contingency of the closing of door 1 can be created.

and effectors and thus states and actions in order to facilitate the formation of new skills. In subsequent stages, constraints are incrementally removed.

The proposed prospective repair algorithm has significant learning performance advantage over the flat Q-learning approach for solving tasks that can be decomposed into a series of problems and presented to the robot in an incremental fashion. The significant improvement is the result of knowledge reuse including the preservation of much of the previously learned path in the new strategy and only focused learning on a new compensatory policy to open doors that block the path to the goal. Once the robot has learned how to open any door individually, this knowledge is reused

again for the case where multiple doors are closed simultaneously, thus minimizing duplicated learning.

5.3 Case Study: Learning Expressive Pointing Gesture

Figure 5.7 illustrates how the prospective repair algorithm is applied within the control basis framework. Through experience with success and failure, the algorithm first learns a decision boundary $g(\mathbf{f})$ is computed (using a standard decision tree algorithm) for separating contexts that succeed, $g(\mathbf{f}) = 1$, from those that fail, $g(\mathbf{f}) = 0$. A new learning problem is automatically generated with $g(\mathbf{f}) = 1$ as the (sub)goal. Prospective learning back-tracks along the original policy until the earliest instance of the context, $\mathbf{f}$, can be observed. The robot explores the available actions and attempts to find a policy that leads to the (sub)goal. After learning, the newly acquired repair policy is incorporated into the original policy (Figure 5.7). Thus, prospective learning enables the robot to adapt to the new context while maintaining the structure of the previously learned program.

5.3.1 Experimental Setup

To verify the proposed approach for repairing defects in manual behavior due to underactuation using communicative gestures, we conduct a series of demonstrations using our bimanual upper-body humanoid, Dexter as show in Figure 5.8.

The same experimental setup for learning manipulation skills is used, however, we change the learning context and move the objects away from the robot until they are all out-of-reach. A human peer is brought into the experiment to interact with the robot. This scenario naturally satisfies the conditions of *underactuation* and *mutual reward*. The robot is *underactuated* since it is unable to reach the desired object unless it is possible for the robot to influence the human to bring the object closer somehow. The robot is motivated by achieving a TOUCH reward and the human is instructed

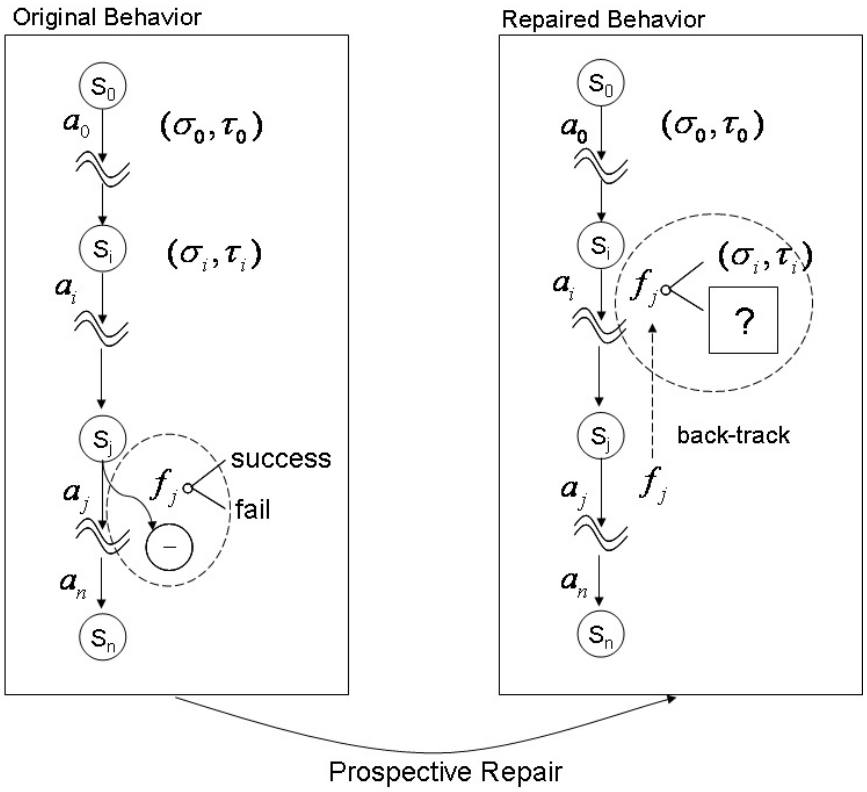


Figure 5.7. Prospective learning. Left: a context change f_j alters transitions generated by the existing policy π and results in an unrewarding absorbing state '$-$' (dotted circle region on the left). Right: the prospective learning algorithm attempts to handle this context change by searching for repairs earlier on in the policy.

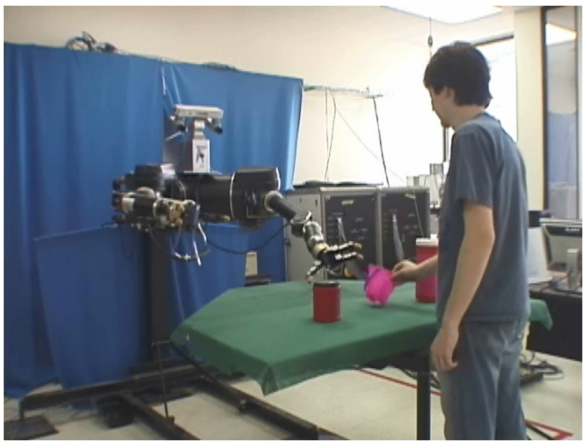


Figure 5.8. Robot learning to gesture in the presence of a human

to interact with the robot and/or the object as they see fit, thus establishing, at least for a time, the condition of mutual reward. Our goal is to determine whether the robot can learn to reliably compel the human to help the robot acquire the object.

To facilitate the learning process, initially the robot explores actions associated exclusively with motor variables in its head. In the second stage, this constraint is lifted and the robot is allowed to use both its arms and its head to communicate. The goal of the experiment is to see if the learning framework enables the robot to learn sequences of actions that are useful for soliciting assistance from the human, even though these actions derive from motor skill learning tasks.

Eighteen subjects of convenience tooks part in the evaluation process. Seven were computer science students, including 2 lab members with extensive knowledge of Dexter. The remaining 11 were diverse in educational background, in major, and in level of education, ranging from high school students to undergrads, to graduate students and working professionals. They were simply told to interact with robot for a number of rounds, and that "the robot will randomly pick an object of interest in each round, observe and help when necessary." All interactions between the robot and the subjects were recorded with consent for the purpose of offline analysis. Human

detection in this work is achieved using the human model acquired in the previous stage.

5.3.2 Prospective Learning and Communicative Behavior

Suppose that the robot detects that its strategy for acquiring an object has failed. It discovers the failure when, in the course of intrinsically motivated exploration, a primitive TOUCH controller fails to reach a rewarding transition to "1" as expected. When the contact feedback force is not detected, the TOUCH controller enters an unrewarding absorbing $'-'$ state.

Figure 5.9 illustrates how the prospective learning algorithm is applied in this case to learn communicative behavior for soliciting human assistance. When the target object is unreachable using autonomous options[1], then the prospective learning algorithm attempts to assimilate the new context into the existing motor behavior. It does so by gathering positive and negative examples of the TOUCH transition in question, and uses a discriminative learning algorithm (decision tree $C4.5$) to extract feature $\mathbf{f}$ such that classifier function $g(\mathbf{f})$ correctly predicts the outcome of TOUCH. In this case, $\mathbf{f}$ corresponds to the x-coordinate of the object in the robot's coordinate frame, and the decision boundary classifier function $g_h(1.2m - x)$, is discovered where $g_h()$ is the hard limiter function. The classifier returns 0 when the argument of $g_h()$ is negative ($x > 1.2m$) and 1 when the object is within reach and TOUCH produces the anticipated intrinsic reward.

The prospective learning (PL) algorithm back-tracks through the greedy rollout of negative examples of TOUCH reward and finds the earliest state where the context $x > 1.2m$ can be observed, and PL formulates a subtask learning problem for the task defined by $g_h(1.2m - x)$, and learns programs capable of achieving the sub-goal by

[1]We adopt the term "autonomous" to denote that the option depends only on existing skills and robot resources.

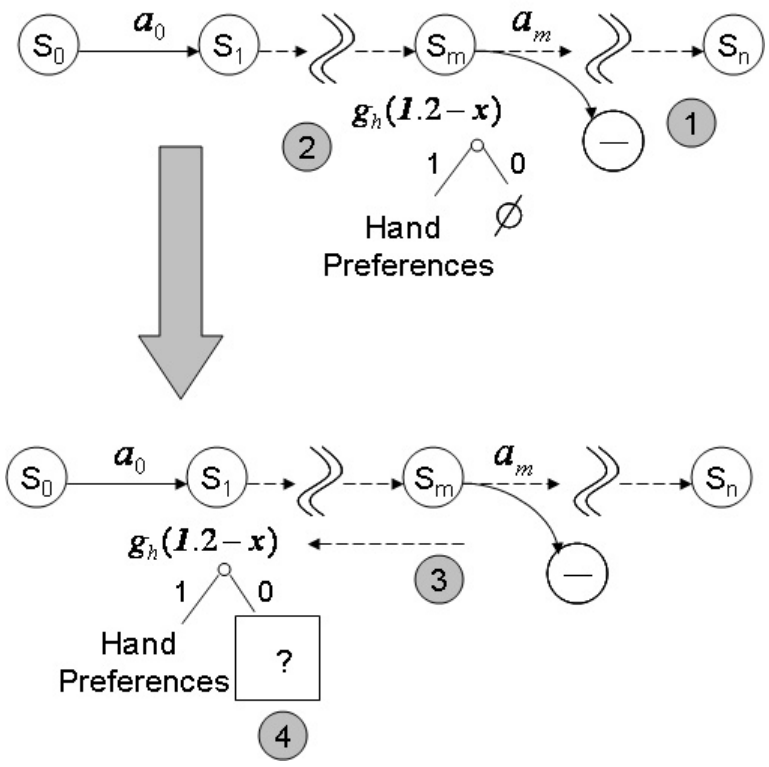


Figure 5.9. Prospective human recruitment. When an object is out-of-reach, (1) the robot detects the failure as it enters an unrewarding absorbing $'-'$ state, (2) it then uncovers a decision boundary ($x > 1.2m$) regarding when its knowledge of hand preferences can no longer lead to the rewarding TOUCH event, (3) the robot backtracks through the program and finds the earliest state where the context $x > 1.2m$ can be observed, and (4) formulates a subtask learning problem.

conveying the intention to TOUCH to the human partner and successfully recruiting their assistance.

5.3.3 The Emergence of Gaze Pointing

The robot must now learn behavior for achieving the rewarding $g_h(\mathbf{f}) = 1$ condition. To do so, it explores actions for recruiting human assistance efficiently. In this developmental stage, Dexter is constrained to actions (and therefore states) that use head degrees of freedom exclusively. This is a severe constraint that permits only a single type of action; one that moves the head so as to track segments of the retinal image—an action Hart called SEARCHTRACK (ST), often parameterized by the visual feature in question, i.e. ST(motion)[40]. Dexter can implement this type of behavior in many different control circuits, directing attention to visual segments distinguished by hue, saturation, and intensity. It may sample these referents from the background, the object in question, or other objects. When a human enters the context, Dexter can direct its gaze to the large motion cue as well as other elements of the HRI context. These actions are the results of prior learning using the same framework [40].

In addition, a *monitor* (see Chapter 3) is configured as a "off-policy" controller (with no effector resources) that observes the X coordinate of the desired object. The monitor reports 1 when the object is inside decision boundary $x < 1.2$ and reports "X," "$-$," or "0" otherwise. For short, this monitor is denoted as ϕ_m^{obj}. The resulting action set $\mathcal{A}$ available to Dexter is: $\mathcal{A} \in \{ST(human), \phi_m^{obj} \triangleleft ST(obj)\}$, where the monitor is concurrently executed with the SEARCHTRACK action associated with the object. According to the control basis (Section 3.1), from this action set, $\mathcal{A}$, a 3-predicate state space $\mathcal{S}$ is automatically formed for the subtask: $\mathcal{S} : \{p_{ST_{human}}, p_{ST_{obj}}, p_{m_{obj}}\}$ with one predicate ST(motion) directed at the human, one ST predicate directed to any feature associated with the object, and one pred-

icate describing the status of the monitor. The robot is rewarded for reaching any state where the $g_h(\mathbf{f}) = 1$ subgoal can be observed.

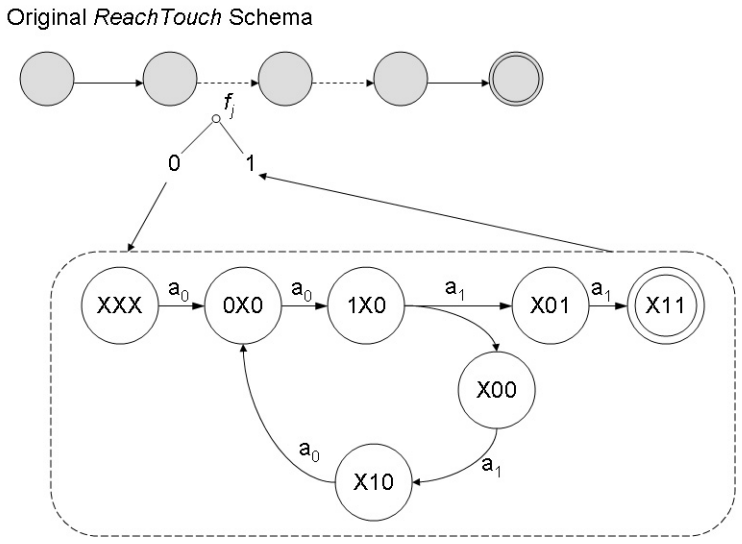


Figure 5.10. New policy for touching a target object, with a new modular gaze gesture acquired by prospective learning. In the repair policy MDP, a_0 corresponds to behavior that searches for and tracks large scale motion cues and a_1 is the same behavior directed toward an object. Each state predicate in the MDP corresponds to the dynamic state of the action and monitor. This policy alternates visual attention directed at the human and the object in a cycle.

Given the state-action space, the goal and the reward, Dexter learns a policy for reliably causing the object to be moved closer, using standard Q-learning with $\alpha = 0.1,\ \gamma = 0.9,$ and $\epsilon = 0.9$. After training, 10 more episodes were conducted using 10 subjects (three of whom were also involved in training) to test the performance of the resulting policy. The learning curve is shown in Figure 5.11. This curve is the average over 5 training subjects for the training phase and 10 evaluation subjects for the testing phase. The dip in average reward at the beginning of the testing phase is caused by the ambiguity of the gaze gesture because some subjects, who did not participate in the training phase, are initially confused about where to place the object.

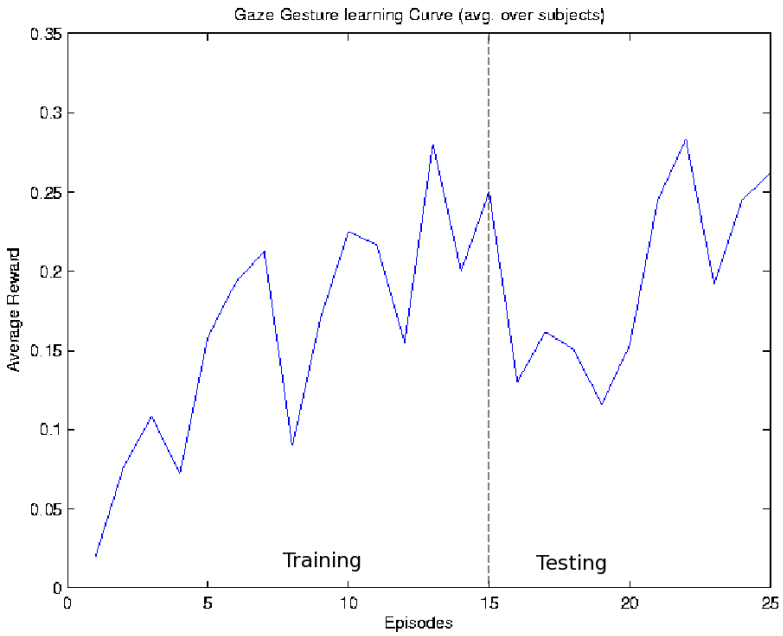


Figure 5.11. Gaze gesture learning curve, averaged reward per state transition over all subjects. The first 15 episodes are the training phase.

Figure 5.12 shows the performance of the "repaired" policy that makes use of gaze gestures and human assistance. Even though the new policy is not yet ideal for acquiring the appropriate response from the human, the standard deviation in the performance plot shows that it is a significant improvement over random. The recorded video footage reveals that a policy of sustained gaze, at either the motion cue or the object, did not cause subjects to respond and was, therefore, not rewarded as often as an alternating gaze strategy that provoked some response from everyone. Gaze is imprecise as a deictic pointer because the motion is subtle and can therefore result in inaction on the part of the human (especially is executed only once) or cause the human to pick the wrong/adjacent object. Alternating gazes were much more conspicuous and led to more reward. Most subjects quickly got the idea that the robot was attempting to communicate with gazes within about 10 episodes, however, they still made mistakes due to ambiguity in the robot's gaze direction. Finally, even

when the subject identified the target object correctly, it remained ambiguous as to where the object should be placed to assist the robot. 60% of the subjects took several tries to place the object within the reachable region of the robot. For these reasons, one person showed confusion about the gaze gesture throughout his interaction with Dexter, and managed to help only once.

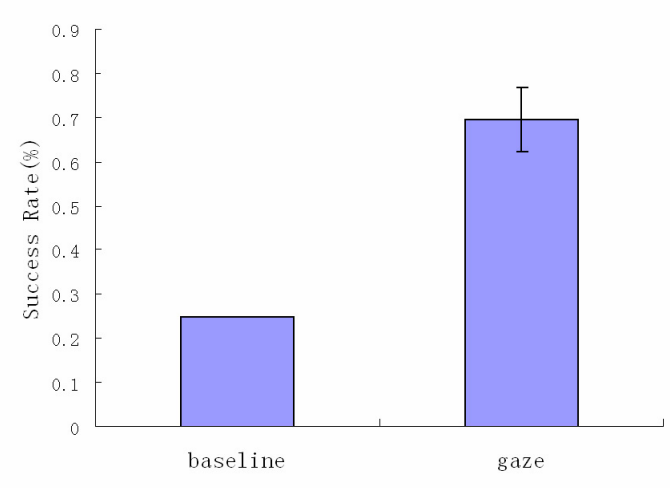


Figure 5.12. Learned gaze gesture performance for acquiring human selects an object at random. The expected random performance for 4 objects is 25%.

A more surprisingly observation is that, those subjects who presumed to be novices with no experience with Dexter or robots in general had more successful rounds of interaction than supposedly more "knowledgeable" subjects (Figure 5.13). A possible explanation is that since this is a such a simple scenario, over-analyzing (speculating on how Dexter receives reward, or what actions Dexter will take) tends to cause more confusion and hesitation than if the person simply acted out instinctively. The result is even more significant when we further divide the "knowledgeable" subjects into two categories, one group contains people who have worked with Dexter and the other contains the rest. The "Dexter-experienced" group performed the worst of all subjects because they are used to Dexter gazing at objects with one of its eyes

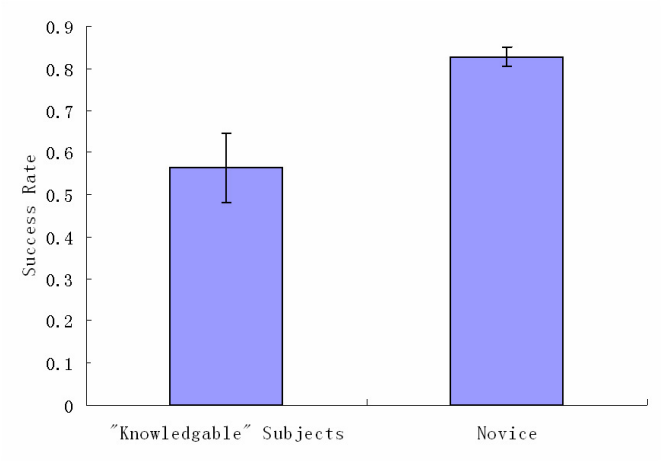


Figure 5.13. Comparison between "Knowledgable" subjects with robot experience and naive subjects

and therefore attempted to parse the direction of the gaze using the dominant eye. However, unknown to them, for this experiment, Dexter was configured to track using both of its cameras and as a result, its gaze direction keeps the object in-between its eyes. One of these "Dexter-experienced" students realized this in the middle of the experiment and corrected accordingly, while the other persisted till the end and made a few incorrect guesses, thus lowering the overall statistics.

5.3.4 Learning Arm Pointing

During the second stage, Dexter explored augmenting its strategy for recruiting precise human assistance by using arms and previous manipulation behavior. In addition to the gaze policy from the previous stage, an existing policy for reaching to a triangulated visual target [40] is now permitted. Therefore the action set is: $\mathcal{A}_2 \in \{ST(human), ST(obj), \phi_m^{obj} \triangleleft RT(obj)\}$, where RT denotes the learned ReachTouch policy. The state space is augmented accordingly: $\mathcal{S} : \{p_{ST_{human}}, p_{ST_{obj}}, p_{RT_{obj}}, p_{m_{obj}}\}$, where m_{obj} is the monitor predicate. As in the previous stage, the robot is rewarded for reaching any state where $g(\mathbf{f}) = 1$, where $\mathbf{f}$ is the monitor feature.

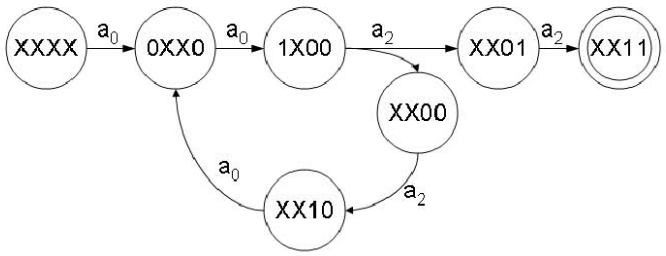


Figure 5.14. Pointing gesture policy for repairing the original manual program. The robot has learned to alternate between gazing at the human (a_0) and reaching for the object (a_2). Each state predicate in the MDP corresponds to the dynamic state of the actions and monitor.

Dexter learned an extended policy (Figure 5.14) within 30 additional training episodes. Three people took part in the training process for 10 episodes each. The resulting policy was tested by *eight* more subjects. Three subjects in this experiment also participated in the previous gaze experiment.

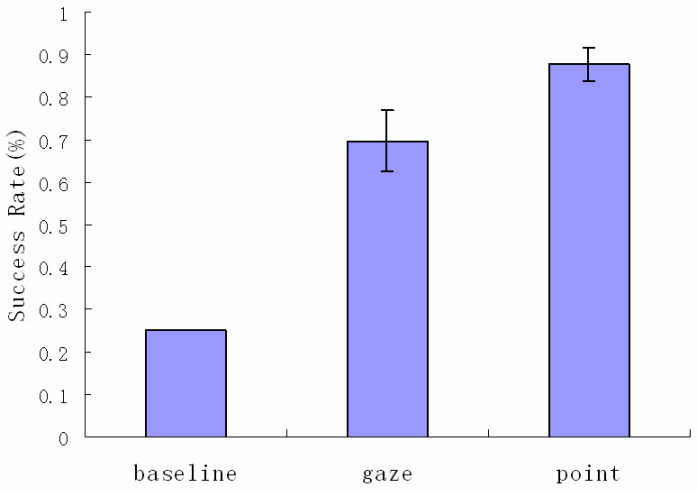


Figure 5.15. Pointing policy performance in comparison with the previously learned gaze policy

The resulting policy has the same structure as the previously learned gaze gesture. This implies that if we reuse the structure of the gaze gesture and simply exchange the

actions that direct gaze with actions for reaching to the object, it is possible for the robot to obtain a skeleton of the arm pointing gesture with no additional training. Of course, further training can be performed for the policy to be refined over time. This time, as a natural outcome of exploring learned manipulation behavior, Dexter found the failed attempt to reach and grab the desired object a more effective alternative to the gaze gesture (Figure 5.15). This is expected because the arm pointing gesture is less subtle, and less ambiguous regarding the target object and where it should be placed. In fact, even the person who failed to attend to Dexter in the previous stage, responded almost immediately in this stage.

5.3.5 Potential Issues of the Learned Pointing Gesture

The pointing experiment revealed a pathological flaw of the learned pointing gesture: when the human handed the object to the robot's out-stretched hand, sometimes the object was visually occluded by the hand. As a result, the robot retracted its arm and confused the subject who thought that they had selected the wrong object. Although this did not occur often enough to prevent the robot from learning the pointing gesture, it is conceivable that if smaller objects were used, more unsuccessful attempts would arise.

This problem can be resolved if the robot develops the understanding of occlusion as part of its manipulation skill set. One such possible alternative could be achieved when a new manipulation behavior, i.e., to "pick and place" becomes available from manual skill learning (see Section 3.6). This is because when parameterized properly, the *pick* goal of the new behavior can indicate the object of desire, while the *place* goal designates the placement location. Thus reducing the likelihood of occlusion that exist for the pointing gesture.

5.3.6 Maintaining Human Interest

For this set of experiments we assumed that the human subjects were benevolent and therefore always behave to help the robot whenever possible. We also made sure the training sessions were short enough so that human subjects would not lose patience and violate the mutual reward assumption.

During the course of the experiments, we noticed that for most people, once they discovered the general strategy for recognizing the robot's intention, they patiently repeated the strategy, placing the object in the same place until all required rounds were completed. For these people, the general assumption of mutual reward is automatically met.

However, two people behaved differently. They soon exhibited signs of boredom after discovering the general strategy for helping the robot and started experimenting with different options to test the capability of the robot by hiding the desired object from the robot's view, placing the object in random locations, moving the object while the robot is pointing, or swapping objects or stacking them up. Due to robust motor behavior, Dexter was able to handle most of testing situations posed by the human and acted "sensibly," i.e. using the left hand for objects placed on the left side and the right hand for objects placed on the right, and the "point" dynamically followed the object if it was moved. This intentional testing kept the subjects interested. One subject performed 5 more rounds of training beyond the requested amount. These observations lend support for the use of existing manual behavior as the basis of communicative gestures as our results suggest that a robot with comprehensive manual skills keeps the human mutually rewarded and engaged, and thus preserves the constructive hiuman-robot dyad.

5.4 Extending the Human Affordance Model—Object with Agency

At the end of Chapter 4, although the robot has acquired a kinematic affordance model of humans, it remained a passive observer of human behavior. However, in this chapter, after learning expressive communicative behavior, the robot has acquired the ability to use actions to influence the behavior of a human and is able to observe the change in the environment.

As mentioned earlier, we hypothesize that although humans are objects with "agency", human behavior is predictable under the social contexts summarized as conditions of *underactuation* and *mutual reward*. It has been established that the experimental setup of this stage naturally satisfies both conditions of underactuation and mutual reward. The question is "can Dexter reliably engage the human being as an *actuate-able* resource such that the rewarding touch response can be achieved?" If so, then such behavior can be considered as an affordance property of the human and therefore can be added to the human affordance catalog.

The evaluation stages of the GazePoint and ArmPoint behavior were also used as the affordance modeling process, through which Dexter gathered data to learn the probability distribution for the concept "how likely would the large motion segment respond to a gaze of arm point gesture and help me get reward?". Mathematically, this affordance is denoted $Pr(r|f, a)$, where f in this case is the feature human-scale motion, a is either the GazePoint or ArmPoint behavior. The results from the evaluation stage show that the affordance model for GazePoint $Pr(r|f,\text{GazePoint})$ habituates around 0.7, while $Pr(r|f,\text{ArmPoint})$ habituates around 0.87. As shown in Figure 5.16, the human affordance catalog is augmented with two affordances: both *Actuate-able(Gaze)* and *Actuate-able(Point)* are reliable behaviors afforded by humans, indicating in addition to previous concepts of humans, to the robot, humans are also "objects" that are "actuate-able" via pointing gestures.

The hierarchical version of the affordance model is shown in Figure 5.17 where a new branch of "actuate-able" affordances has been added.

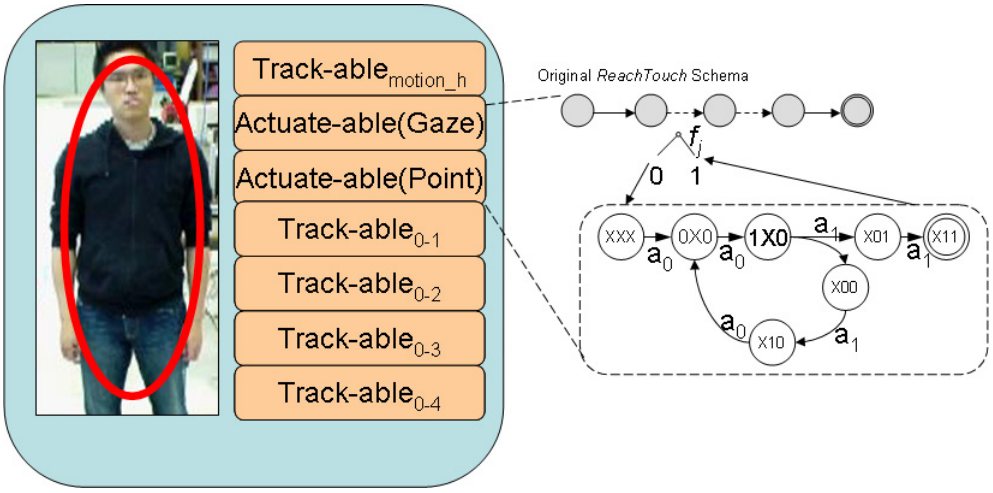


Figure 5.16. The human affordance model after this stage. The robot has found two reliable behavior for "actuating" the human resource. Thus they are behaviors humans afford and are then added to the human affordance catalog.

Finally, the resulting affordance of this stage can be formulated succinctly as a logical assertion:

$$((Trackable_{motion_h} \wedge Trackable_{obj}) \wedge !(X_{obj} < 1.2) \wedge (Gaze \vee Point)) \rightarrow (X_{obj} < 1.2)$$

This assertion represents an intuitive "actuate-able" concept regarding humans. Due to the systematic nature of the prospective learning algorithm, this logical assertation can be derived automatically since the first part, $(Trackable_{motion_h} \wedge Trackable_{obj})$, represents the prerequisite conditions before the repair behavioral module can be executed—the corresponding objects must be present. The second part $(!(X_{obj} < 1.2))$ corresponds to the failure conditions the robot has encountered that warrants the repair policy, while the third part $(Gaze \vee Point)$ is the repair policy. Finally, the last part $(X_{obj} < 1.2)$ corresponds to the condition where the rewarding event can be achieved.

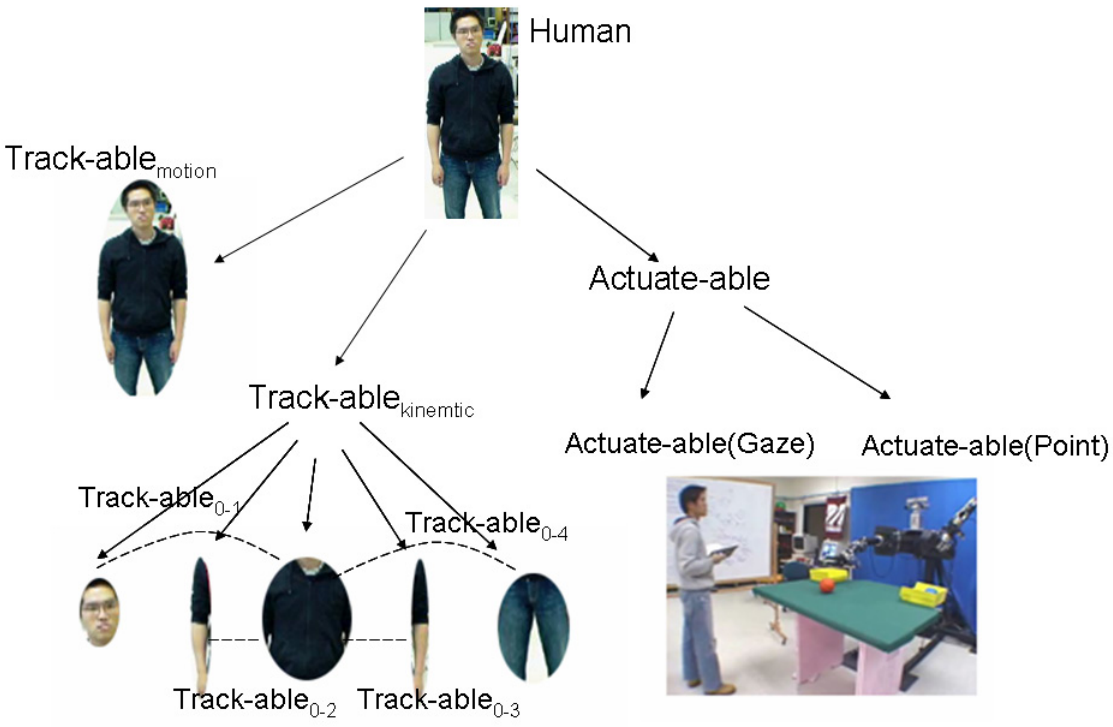


Figure 5.17. The hierarchical And-Or Graph of learned human affordances is augmented with new expressive behavior affordances.

5.5 Discussion

This chapter proposes a grounded approach to the acquisition of expressive communicative behavior in robots and presents a framework in which a robot can learn communicative actions and manual skills in conjunction. A human interaction case study is presented to demonstrate the feasibility of this approach. The approach enabled the reuse of manual skills acquired from previous intrinsically motivated behavior for interacting with objects in the environment.

Using manual behavior as the basis of communicative gesture, the robot was able to learn behavior programs that effectively convey its intentions to humans in very few on-line interactions with the human subjects. The robot learned in stages; initially employing gaze exclusively and subsequently integrating pointing gestures with its arms.

Possible learning stages to further improve the effectiveness of the pointing gesture are also suggested. The experiments provide support for using robust manipulation behavior as the basis of socially interactive behavior. This approach can be beneficial for maintaining the interest of the human subject and thus prolonging the interaction.

Finally, human detection in this work is achieved using the simple motion model acquired during the first stage of the human modeling process. This provides support evidence for the incremental human modeling approach employed by this thesis—even simple models are useful in some cases for the robot to learn useful behavior. In turn, learned behavior improves the robot's ability to build more sophisticated models. For instance, with the result of this stage's behavior learning, the human affordance catalog is augmented two new affordances, allowing the robot to acquire a new concept that "humans are actuate-able via pointing gestures." This raises the robot's understanding of humans from a passive notion of "moving objects that are track-able" to objects with "agency".

In future work, we hope to observe the emergence of more communicative gestures by subjecting the robot to more challenging scenarios, while it acquires more complex manuals skills. We expect that the *size-hinting* gesture can arise from the two-handed grasping behavior when the object becomes too large for one hand. *Beckoning* can emerge as the robot attempts to bring the human closer with the manual behavior for bringing graspable objects closer; lastly, a "no" negation gesture can emerge as the robot discovers the communicative utility of the *push* behavior when it attempts to push unwanted objects away.

CHAPTER 6

LEARNING RECEPTIVE BEHAVIOR

Previous chapters have demonstrated how a consistent learning framework and knowledge structure enabled expressive behavior to emerge from manual behavior. I have also demonstrated how these behaviors can be used to form affordance models of humans. In the same spirit, this chapter focuses on receptive behavior and discusses how a robot can take advantage of the previously learned programs for the purpose of inferring intentions from human partners. A case study is presented to demonstrate the feasibility of this approach: a robot interacts with a number of participants who require the robot's assistance for obtaining the object they desire. From these studies, we wish to evaluate:

- whether the proposed approach can transfer background information from expressive behavior to infer the intentions of naive humans,
- whether the robot can learn to engage the appropriate behavior in response to communicative action from human beings.

6.1 Related Work

In computer vision, many algorithms have been developed for gesture recognition. For static gestures, recognition is generally achieved by using template matching, Principle Component Analysis (PCA) [42], or Elastic Graph Matching [114]. Algorithms for recognizing dynamic gestures often employ Hidden Markov Models (HMMs) to parse sequences and observations. A time-series of hand or body postures

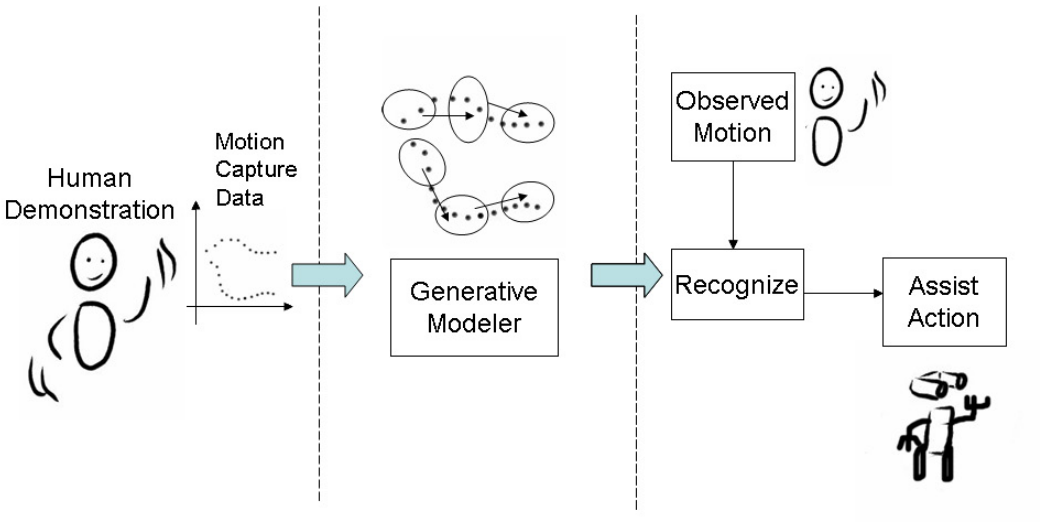


Figure 6.1. Conventional approach for human gesture recognition where models are learned from passive observation of human demonstrated motion, for gesture recogition and imitation. The generation of assistive behavior (assist action selection component) is normally not considered as part of learning process.

are used as training inputs and a model that fits the training examples is acquired. After different models have been learned, the system can then be used for recognition by computing the match likelihood of the time series. This approach has been demonstrated by Nam [79] and in Starner's work [107] where real-time recognition of 40 words in American Sign Language was achieved. This approach works best when gestures are purely postural and do not refer to environmental entities. Other variations of the HMM method such as Dynamic Time Warping (a simplification of HMM) [17] and Bayesian time delay networks [119] have been proposed to simplify the process.

In robotics, motor primitives are extracted from demonstrations and used to construct adaptive behavior in novel contexts. Several methods have been introduced to enable robots to learn models capable of recognizing novel, more complex behaviors. Researchers [62, 51] have developed methods to represent high-dimensional motion-captured data from human demonstrations by automatically segmenting the data and

encoding them into motor primitives. Therefore novel demonstrations can be automatically segmented, recognized in terms of their parts, and mapped onto the known motion primitives. Kulic's method [62] is HMM-based, while Jenkins [51] augmented a manifold learning approach called Isomap by incorporating spatial and temporal relationships and developed ST-Isomap. Although these methods were primarily developed to enable robots to learn by imitation or learn by demonstration (LbD) using relatively few demonstrations, they can be also used for recognizing human gestures as shown in [52].

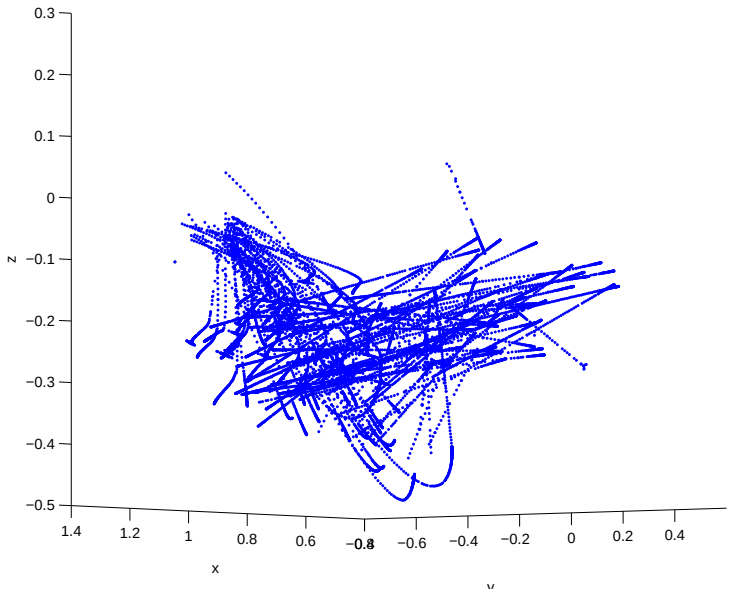


Figure 6.2. A number of tracked trajectories for the PickPlace behavior in Cartesian space. It would difficult to model these trajectories as they cover much of the Cartesian space. The model can only become more ambiguous when more data is captured. This example shows that motion trajectory data is not uniformly informative and are inherently ambiguous, since all actions share trajectory to some degree.

Figure 6.1 shows the commonalities among these methods. First, these approaches depend on human demonstrations, in the form of dense motion capture data. These motions are not uniformly informative and are inherently ambiguous since all actions share trajectories to some degree (Figure 6.2). Also, learning how to help once the

primitives are recognized is generally not a focus of methods currently in the literature. However, I argue that there exists a more fundamental problem with these approaches: the emphasis has been put on motion patterns which when performed out of context does not directly communicate any information.

In contrast, the method proposed by this work focuses on the purposeful act itself as well as the associated environment contexts. This approach begins from the robot actively examining its own behavior under the context where itself requires assistance. The reasoning behind this approach and its benefits are discussed in the next section.

6.2 Methodology: Learn to Infer Intention

Figure 6.3 provides an overview of the proposed approach for robots to learn to recognize and understand intentions from humans. Comparing with Figure 6.1, its key distinctions from previously mentioned methods are as follows:

1. the process begins with the robot building proprietary world knowledge and models through autonomous exploration,

2. recognition is based on hierarchical structure of behavior and puts emphasis on isolated events pertinent to the functional outcomes of purposeful behavior,

3. and finally, the robot actively explores and learns the compensatory behavior for participating in a potential dyadic relationship.

In the next few sections, we expand and elaborate on each of these components in detail.

6.2.1 Capturing Intentions by Building Proprietary World Models

In this work, we conjecture that robots can recognize and interpret intentions in terms of their proprietary world knowledge—knowledge that the robot has acquired while interacting objects and humans in the environment, using techniques we

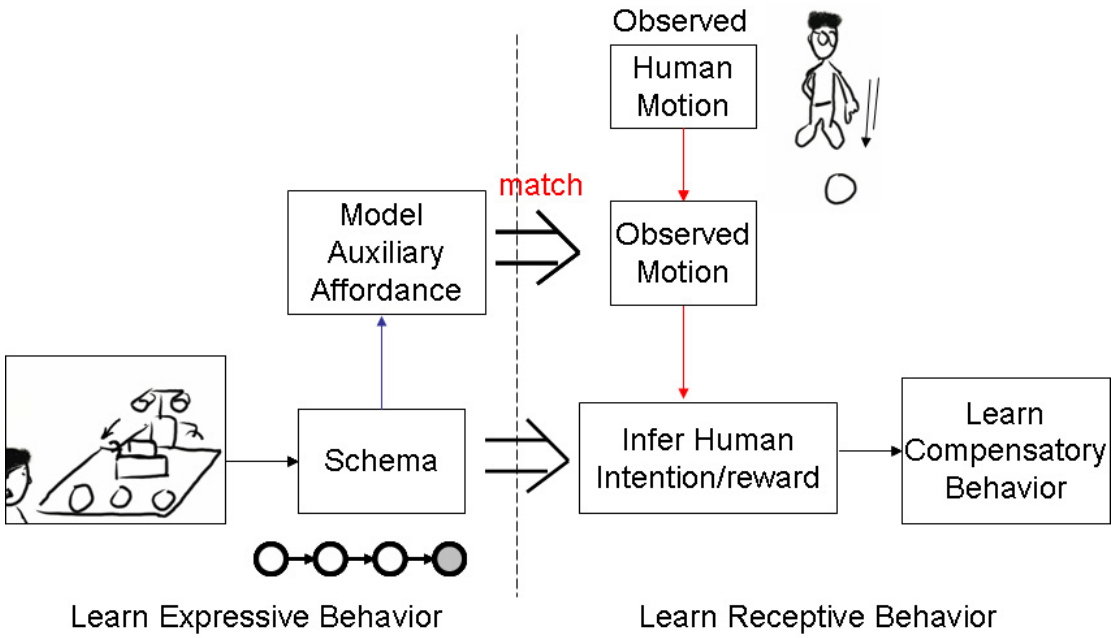


Figure 6.3. The proposed approach for robots to recognize intentional behavior from other agents, by reusing knowledge acquired from prior learning sessions.

have discussed in earlier chapters. I hypothesize that these behavioral programs and knowledge structures already captured much information regarding the behavior, the associated intention and the environmental context. Though from the robot's own perspective, this information is valuable for bootstrapping the learning process of receptive behavior. Compared with prior work where learning begins from human motion demonstrations, this approach takes advantage of exiting knowledge and is grounded in nature since in this case, the robot gathers knowledge about the behavior in situ and therefore can associate the learned behavior with goals and intentions.

This approach is inspired in part by the discovery of mirror neurons [89]. Scientists have found that certain neurons in monkeys exhibit similar activity when the animal observes the goal-directed action of another agent as when it carried out that action itself. This observation has led researchers to hypothesize that there exists a common coding between perceived and generated actions [86, 10]. Therefore, these neurons

may play an important role in how humans and other animals relate their own actions to actions of others.

The proposed approach is based upon the same principle. Consider recognizing a "pointing" behavior in human beings. In my framework, the robot first learns a suite of pointing behavior itself. In the process, it makes the appropriate associations between its actions, its own goal and intention, and the behavior of a peer. Later, when the role is reversed and the human peer selects communicative actions to solicit help from the robot, this background knowledge can be used by the robot to infer the intentions of the peer in order to learn how to participate in the dyadic relationship.

The idea of role switching has been explored by Berlin et al. in [5] and Roy et al. in [91], where the robot switches to the perspective view point of the other agent to determine object visibility [5] or the object's relative position [91]. This work pushes the idea further and explores how role switching can be applied to a more complex case where rather than a simple view-point change, reasoning with the help of a schematic behavioral program is involved.

In the proposed control-basis (Chapter 3) approach, generative models are acquired from first principles, using active exploration and learning in the context of intrinsic motivations. These models exist in the form of behavioral schemas that include information regarding observable states and the transitional probabilities between these states. The control basis provides mechanisms for automatically generating control actions for exploration in the combinatoric space of available sensors Ω_σ, effectors Ω_τ, and potential functions Ω_ϕ. It also has built-in mechanisms for estimating state by observing control action dynamics. Given predefined resources, $\Omega_\sigma \times \Omega_\tau \times \Omega_\phi$ and a intrinsic reward function, the robot automatically formulates a MDP and begins to learn the optimal policy for maximizing reward using Q-learning. These schematic programs contain state transition probabilities that are similar to a HMM model learned using an approach dedicated to model human motion demon-

strations. More importantly, behavioral programs captured from first principles using the control basis specifically encodes the relationship between action and goal by computing the expected reward for each action at any given state. If this can be applied for recognizing the same behavior from a human, then learning from scratch can be minimized. This topic is discussed next.

6.2.2 Intention Recognition using Hierarchical Structure of Proprietary Behavior

For recognition, in a teleoperation scenario, where a human "shows" the robot how to perform a task either via motion capture devices or literally by holding the robot's hand, the robot can directly take advantage of these these schematic programs by matching sensory observations to them in an on-line manner. For instance, Figure 6.4 shows the robot's learned policy ReachTouch for reaching out and touching objects of desire. The circles represent the states of the agent and the directed edges are actions that cause the agent to transition from one state to the next. In the definition of each control action, the appropriate sensor, effector and potential function are specified to ensure the appropriate resources are allocated so that the transition to the next state can be monitored. For recognition, the robot simply keeps track of the state action transition of the current control policy and matches it against the learned transitions in the behavioral program. As the robot transitions from one state to the next, the traversed transition probabilities are multiplied to give a likelihood match score. At any given time, the behavior that has highest score is considered as the intended behavior by the human demonstrator.

However, for my work, I consider a different scenario where the human stands across the table to the robot, in a face-to-face situation. In this case, the robot observes the action performed by someone who is not "holding" its hand. While the declarative structure of the behavior program is still useful, e.g. in the case of

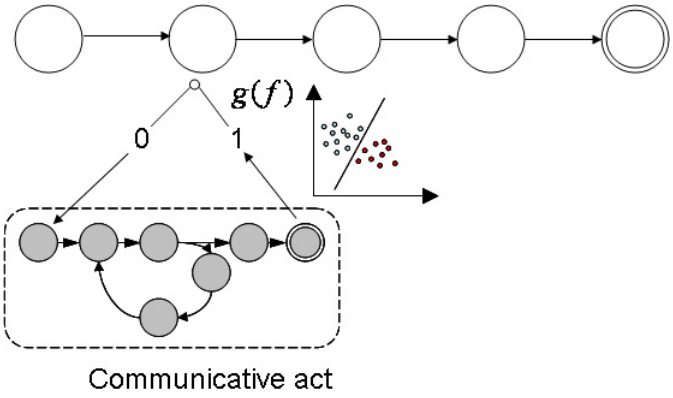


Figure 6.4. The learned hierarchical program, ReachTouch, although can be used for intention recognition in a teleoperated task demonstration scenario, in a face-to-face interaction scenario some procedural knowledge (such as the learned decision boundary $g(f)$ for determining the reachable regions) cannot be generalized to third-party agents.

ReachTouch that if the agent cannot touch the desired object, a communicative act is needed to "repair" the policy, some of the procedural knowledge does not generalize to third-party agents. For instance, in the ReachTouch program, the rewarding Touch event cannot be detected when the behavior is performed by a human, and similarly, the decision boundary regarding whether a communicative action is needed is specifically with respect to the kinematics of the robot and therefore does not generalize to the length of arm for the human. The following section presents a solution.

6.2.3 Focusing on Goals

The solution is inspired by the *teleological stance* proposed by psychologist Gergely [33], who describes the development during the first year of an infant's life as a process that extracts goals, means and constraints to explain the behavior of others (for details refer to Section 2.3). This is in contrast to other approaches, often collectively referred to as the Theory of Mind (TOM) hypothesis [96], that requires a more complete

mental state of the other agent. The recognition process is simplified when teleological principles are applied.

Based the experimental observations and principles, in the context of robot learning, I argue that isolated goal-related events and context cues are useful for inferring intentions. As mentioned before, although similar information can be directly extracted from existing behavior programs and affordance models from manual and expressive behavior learning, some of the knowledge does not generalize to humans. Fortunately, using the same mechanisms we have described before the robot can also build auxiliary off-policy affordance models that correlate with the on-policy events. This can be achieved by creating off-policy *monitors* and attaching them to the existing behavioral programs. As discussed in Chapter 3, in the control basis, monitors can be created the same way as controllers, by combining sampled resources, i.e. denoted $C_M(f_\sigma, \phi)$. One distinction is that no effector resource τ is attached since a monitor is a passive observer. Other crucial distinction is that features are sampled exclusively from the operational space, as oppose to features extracted from sensors that are internal to the robot, e.g. proprioceptive joint angle information. This distinction is the key to abstracting away from the robot's own body and acquiring knowledge that generalizable for the recognition of behavior from a peer. To utilize this knowledge, rather than matching the entire program state-by-state, transition-by-transition, an agent using this approach simply looks for similar affordances in the stream of observation generated by the peer to identify the goal. This process is reflected in the "Model Auxiliary Affordance" component as illustrated in Figure 6.3.

More specifically, given an existing behavioral program, goals are easily identifiable since they are the terminating states where rewards occur. When the robot encounters constraining conditions, a communicative act, e.g. arm pointing, is needed. To learn auxiliary affordances that are highly correlated with each of these rewarding events, Algorithm 2 is presented.

Algorithm 2 A Sampling-based Algorithm for Building Auxiliary Affordances

1: Given a operational space feature set $\mathcal{F} = \{f_1, ..., f_n\}$, a behavior program a and its rewarding event e,
2: Sample m features $\mathcal{F}_s = \{f_i\}_m$, where $f_i \sim \mathcal{F}$.
3: **for all** $f \in \mathcal{F}_s$ **do**
4: Create a monitor C_{Mi} for each $f_i \in \mathcal{F}_s$, where each predicate p_i corresponds to the dynamic state of each monitor.
5: Attach monitor to action a, s.t. $C_{Mi} \lhd a$.
6: **end for**
7: Executing action a.
8: **for all** $f \in \mathcal{F}_s$ **do**
9: **if** $p_i : 0 \rightarrow 1$ **and** $e : 0 \rightarrow 1$ **then**
10: Update affordance model $Pr(r|f_i, C_{Mi} \lhd a)$
11: **end if**
12: **end for**

This is a sampling-based method where the robot creates monitors for sampled features from the operational space feature set $\mathcal{F} = \{f_1, ..., f_n\}$ (as shown in line $1 \sim 2$). To abstract away from the robot's own body, each operational space feature f describes the relative property between a feature internal to the robot and a feature from the external environment, e.g. relative distance between the hand of the robot and a feature on the desired object. In general, feature set $\mathcal{F}$ can be either automatically generated from given the sensor resources Ω_σ, or hand-picked by the designer to reduce the search space as part of a developmental learning strategy. While the robot interacts with objects and humans in the environment, this algorithm runs repeatedly in the background collecting statistics to update affordances that correlate with the rewarding event e (line $7 \sim 12$). The statistics update in line 10 is triggered by the detection of the convergence event of a monitor C_{Mi} and the co-occurence of the goal event $e : 0 \rightarrow 1$ (line 9).

During recognition, these highly correlated affordances are used as indicators for predicting the rewarding event e when it is not directly observable by the robot. An example is given next section, where a robot uses this algorithm to extract cues

for detecting the pointing behavior from a human, and for inferring the rewarding TOUCH event for the human which itself cannot directly experience.

6.2.4 Learning Reciprocal Behavior

Much of the previous research on learning by demonstration stops after the recognition step and replays the motor pattern or executes a fixed response. In this work however, we consider the recognition step as the first stage of a receptive behavior where actions prepare the robot for recognition, as well as behavior for re-orienting if an initial attempt at recognition fails. Furthermore, more complicated scenarios may demand the robot to alternate between intention recognition and manual behavior as the context requires. Such behavior involves sequencing other existing behavioral programs, a skill that a general purpose behavioral learning framework such as the control basis is designed for. A case study to demonstrate how the framework enables the robot to learn the receptive behavior is described next.

6.3 Case Study: Learn to Recognize Pointing Gesture and Assist Behavior

In this case study, we employ a bimanual mobile robot, the uBot-5, as shown in Figure 6.5. The uBot-5 is a small and lightweight dynamically balancing mobile manipulator with 13 DOF. It is designed to perform work with a whole-body approach to mobility and manipulation, e.g. by exploiting the mass and dynamics of its entire body to improve pushing and throwing performance [19]. Multiple hands have been designed for the uBot and can be swapped in and out as the task requires. For the purpose of this work, a light-weight simple 1-DOF servo-motor hand is used for the uBot to perform simple grasping tasks. The uBot observes the world through its stereo camera pair mounted on a pan/tilt head.

Figure 6.5. The uBot performing plowing, stacking, pushing, and throwing tasks.

The behavioral programs that form the action primitives for this work are acquired using a non-mobile upper-torso humanoid robot named Dexter. As a result, these programs do not yet incorporate the uBot's mobility. However, the programs for manual and expressive communicative skills transferred to the uBot in a straight forward manner. Only definitions of the uBot's kinematics was required. To learn receptive communicative behavior, the uBot was placed on a stand with its wheels turned off to simulate the conditions in which the programs were originally learned.

We replicate the same experimental setup with which Dexter learned to point. However, rather than placing the objects out of the uBot's reach, the objects are placed within reach of the robot and out of reach of the human. The uBot is already familiar with these objects from the previous studies and therefore only finds activities associated with the human rewarding. This scenario naturally satisfied our previously defined conditions for natural emergence of communicative behavior: *underactuation* and *mutual reward*. First, one of the agents, in this case the human, is unable to reach the object unaided and is underactuated. However, it is possible for the robot to assist the human to reach the desired object by virtue of the experimental statement. Secondly, the robot and the human are *mutually rewarded* since the human is rewarded for acquiring the out-of-reach object. On the robot's side, it is possible for the robot to infer the goals of the human in terms its own experienced goals and rewards under similar circumstances, thus it receives reward that it does not directly experience.

This reward model can be implemented to motivate the robot to learn the appropriate receptive behavior for assisting the human, without the need of developing a complete mental model for him.

Ten subjects of convenience are used for this study, half of whom were are computer science students, while the rest are various members of the campus community. three out of the ten were female. The humans are asked to stand across the table to the robot one by one, randomly pick an object and try to enlist the robot's help. Four subjects are involved in the training session, using two objects. During training, the robot has to learn the appropriate behavior to assist the human. Once trained, all ten subjects are asked to participate in the evaluation phase where four objects are used and the robot simply executed its learned behavior. All interactions between the robot and the subjects are recorded with consent for the purpose of offline analysis. Humans motions are detected and tracked as a whole and individually with the stereo camera-pair, using the multi-body kinematic model learned in Chapter 4. Using stereo triangulation, each Cartesian coordinate of the human body is computed and used as features to match against auxiliary affordance models for the human's intention to be recognized.

6.3.1 Recognizing Human Pointing

In the proposed approach, the first step towards recognizing the pointing gesture from a human is to acquire the same behavior from the robot's own perspective. It is hypothesized that this allows the robot to make the association between the action taken and the eventual reward when it own goal is satisfied.

For the uBot, the pointing schema was learned in a prior study (Chapter 5) on a different robot—Dexter. Dexter was situated in a social context where it needs to formulate its own behavior for soliciting human assistance in order to achieve its goal. In such a situation where the robot's previously learned ReachTouch behavior

failed, the robot adapted and learned a repair strategy for the original ReachTouch by negotiating two different ways of communication with nearby humans. One of which is the pointing gesture policy where the robot uses alternating gaze and arm pointing actions to convey the intention.

The advantage of the control basis framework is demonstrated as the pointing schema learned on Dexter was transfered to the uBot with little effort. This is because only the declarative structure of the schema was transferred. At run-time the uBot specific procedural knowledge, such as length of arm or handedness, was applied such that the the appropriate resources can be instantiated. For instance, while Dexter never performed a point using two arm, for the uBot, the underlying manual behavior automatically gives rise to a two-arm pointing gesture for certain objects. Next, we will show how the ReachTouch program can be used to bootstrap the learning of the uBot for recognizing the same behavior performed by humans.

As shown in Figure 6.6, given the learned ReachTouch behavioral program with built-in contingencies to point, the teleological processing begins by extracting goals, means and constraints. In this case, the *goal* is the absorbing state in Figure 6.6 where the robot is able to reach the object and detect *Touch* sensor responses from the finger tip tactile sensors. Next, while the robot uses the pointing gesture to solicit nearby human assistance for the out-of-reach objects, Algorithm 2 is applied to sample and monitor configurations to learn auxiliary affordances that highly correlate with the rewarding events via constructing off-policy monitors.

The monitor configuration is sampled from the feature set $\mathcal{F}$ that consists of the relative position between features derived from the human partner and the experimental objects on the table. In theory, the robot applies Algorithm 2 and samples features to monitor while it executes the known behavior until correlated features have been found. However, to expedite the learning process, a developmental learning strategy is employed to focus the robot's attention on the operational space feature

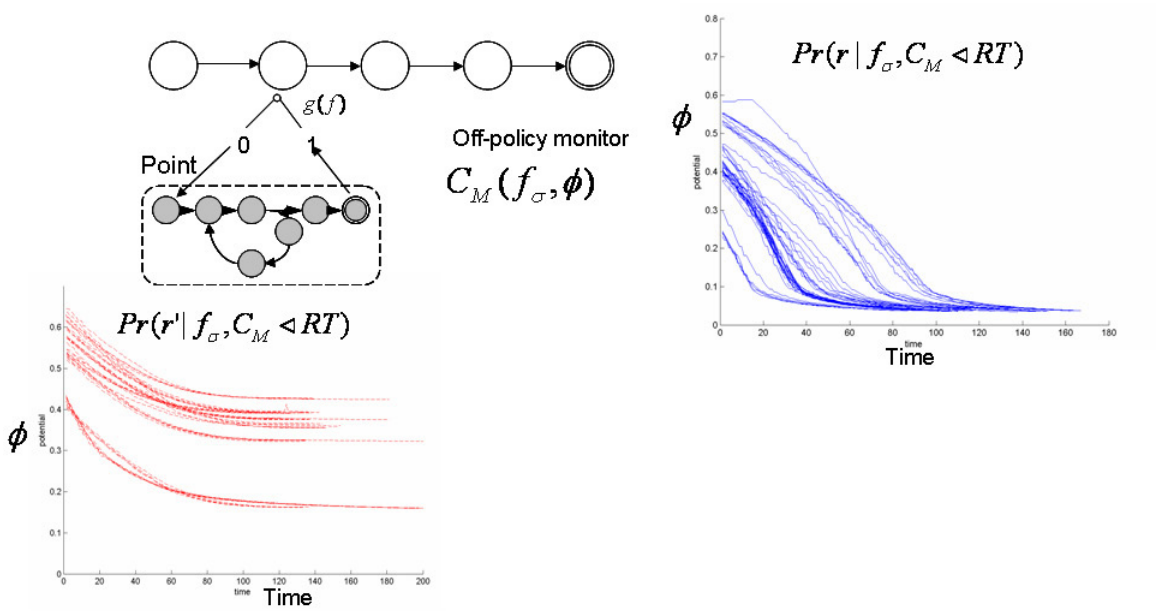


Figure 6.6. The learning of auxiliary affordances for REACHTOUCH. One affordance, $Pr(r|f_\sigma, C_M \triangleleft RT)$ (shown on the right), highly correlates with the TOUCH event of the REACHTOUCH behavior, while the other correlates with the rewarding sub-task event ("object is within reach") when communicative point gesture is performed.

that describes the relative properties between a pair of features: one from the object of interest and another from the catalog describing the robot. More specifically, a monitor $C_M(f_\sigma, \phi)$ is created for each pair of these features, e.g., $f : \{X_{obj}, X_{hand}\}$, where X_{obj} and X_{hand} respectively represents the cartesian positions of the object and the robot's hand. A quadratic potential function $\phi = \epsilon^T \epsilon$ is used to compute a error signal, where $\epsilon = X_{obj} - X_{hand}$.

The robot updates the affordance models as explained in Algorithm 2 every time a co-occurrence of the convergence event on its monitors and a rewarding event is observed. After about 30 interactions, the collected statistics indicate that the relative Cartesian position of the desired object and the hand of the robot exhibit a reliabe relationship to the rewarding TOUCH event, such that whenever the relative distance drops to zero, a TOUCH event is always observed. In a similar manner, another

affordance, where the monitor converges but the relative distance does not reach zero, is found to be highly correlated with the rewarding event for the repair task where a pointing gesture is used (Figure 6.6).

For recognition, when the role is switched between the robot and the human, if similar error dynamics is observed between the a part of the human's body and an object where the error converges at a non-zero value, the robot can thus infer that the intention of the human is to also reduce the error to zero since it is sympathetic to the need from its own experience. After obtaining an estimate of the intention of the human, the next section discuss how the robot learns the appropriate behavior for helping out.

6.3.2 Learning Receptive Behavior

After the uBot has learned cues for recognizing pointing behavior from a human, it explores previously learned behavior to form a new integrated behavior for recognizing the human's need and acquiring a policy for assistance. The action set $\mathcal{A}$ available to the uBot at this stage is: $\mathcal{A} \in \{ST(human),\ ST(obj),\ \phi_m^{obj} \triangleleft RT(obj),\ \phi_m^{obj} \triangleleft PP(obj)\}$, where RT denotes the learned ReachTouch program and PP denotes the PickPlace program for picking up and transferring an object to a designated location. Attached to the $PP(obj)$ and the $RT(obj)$ is a monitor ϕ_m^{obj} for the Cartesian distance between the out stretched human hand and the object, ϵ. Corresponding to the action set, the state space is therefore: $\mathcal{S} : \{p_{ST_{human}}, p_{ST_{obj}}, p_{RT_{obj}}, p_{PP_{obj}}, p_{m_{obj}}\}$. From the previous step, the robot has inferred the intention of the human is to touch the object. However, since it cannot directly observe the tactile event from the human's perspective, the robot is implicitly rewarded for observing the alternative event it has found to highly correlate with the goal Touch event in the recognized program—when the object and human hand distance remains $\epsilon < th$ (where th is a small positive constant) when the object has been passed to him.

For this experiment, 4 subjects of convenience participated in the training process, while the learned policy was tested on 10 people. During the training process, 2 objects were used to ensure the robot was exposed to sufficient positive experience to facilitate behavior formation. Once the robot acquired a stable policy for handling the situation, 4 objects were placed on the table to evaluate the performance of the policy.

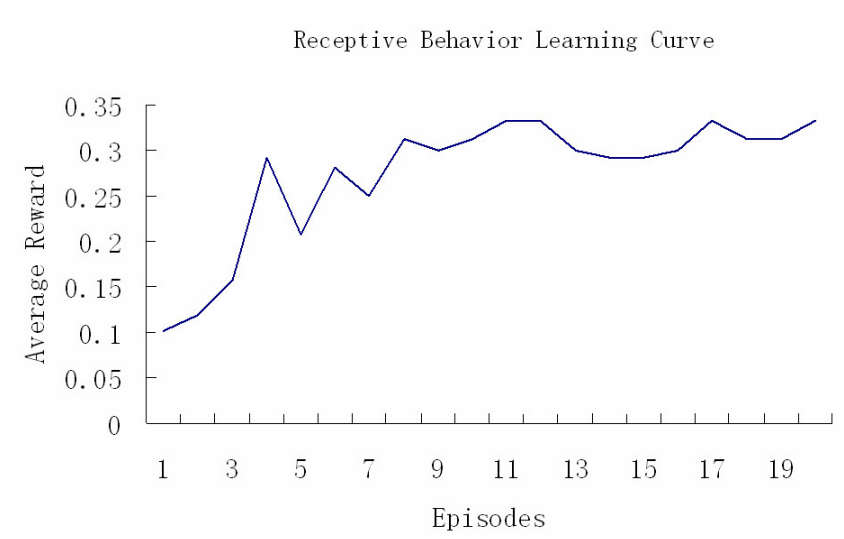


Figure 6.7. Receptive pointing assist behavior learning curve, averaged reward per state transition over all subjects.

Figure 6.7 shows the learning curve of the receptive behavior training process. It can be observed that in the initial stages the average reward the robot achieved is low since the robot was exploring different actions in a random fashion. The number of actions it took for the robot to stumble upon the goal state was high, and thus lowering the initial average reward. However, as the robot gained more experience and began to propagate reward throughout the MDP, its value function improved until greedy behavior was appropriate in most situations. As a result, the average reward per episode for the robot rose and became stable.

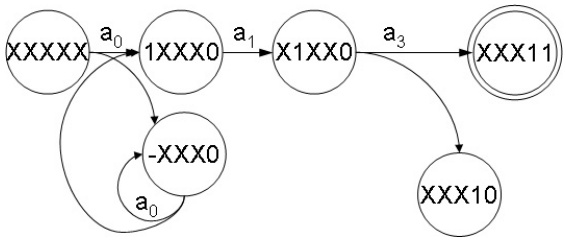


Figure 6.8. Receptive assist policy for recognizing the need of a human for acquiring an out-of-reach object and original ReachTouch program. The robot has learned to use gazes between the human and the objects (a_0 is $ST(human)$ and a_1 corresponds to $ST(obj)$) to recognize the human's pointing gesture and identify the the object of desire, followed by the PickPlace behavior (a_3) to transport the object to the human.

The uBot learns the policy, shown in Fig. 6.8, within a reasonable 20 training episodes with 4 subjects. Due to developmental structuring, the resulting policy uses only 3 actions and has a simplistic structure and therefore is easy for the robot to discover. The final policy involves first a repeated ST_{human} action for a match of the human affordance catalog and if any part of the observed human catalog moves towards the objects, a ST_{obj} action is executed such that the distance between the hand and the objects can be monitored and the desired object can be identified. Finally, a PickPlace can be executed in order for the object to be passed to the human.

After training, the 2-object setup is replaced with 4 objects and the effectiveness of the learned policy is evaluated using all 10 subjects. As expected, the average reward drops as the test setup is undoubtedly a more difficult task since more objects generally lead to more mishaps such as grasp failure and objects being accidentally knocked down by the robot.

Figure 6.9 provides a finer analysis of the success rate of the learned behavior. It shows that although the overall success rate of the learned behavior is only around 64%, it is significantly higher than the 25% chance of picking the right object if the

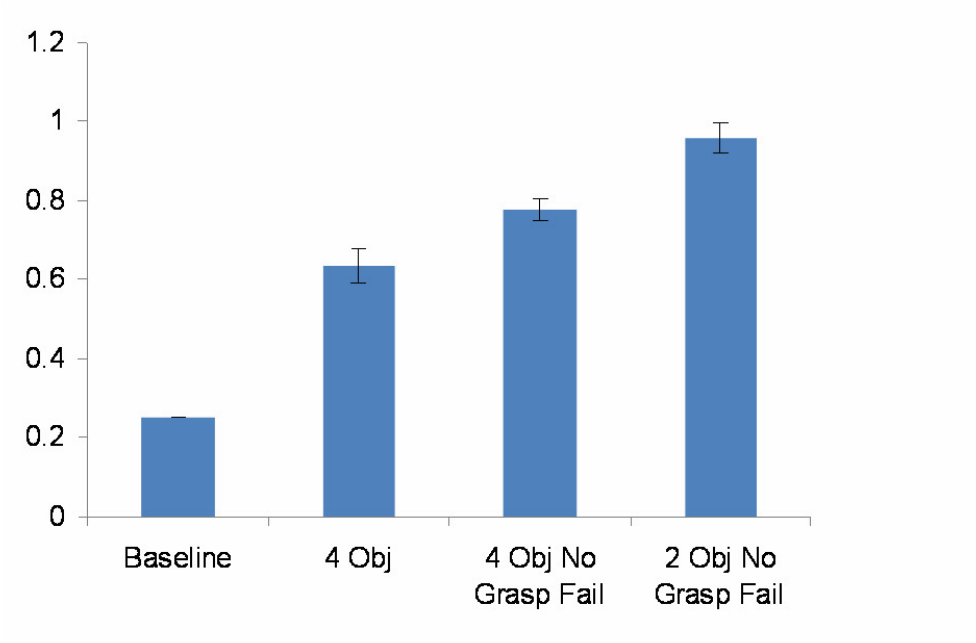


Figure 6.9. Point assist receptive behavior policy performance plot provides a finer analysis of the success rates of the learned behavior

human's attempt to communicate his intention was not recognized and the robot had to pick an object at random. Furthermore, if eliminating the manipulation behavior failures such as the grasping mishaps, the success rate is raised to 78% for 4 objects and 96% for 2 objects respectively. These are reasonable results given that highly ambiguous and coarse hue color features were used as the perceptual basis for these experiments, and the experiments were performed in a naturally cluttered lab environment where noisy background features and lighting changes can all easily contribute to mistakes in object or human detection as well as stereo triangulation error. Performance improvement can be expected when more robust features such as edges or textures are used.

Finally, we discuss the observed behavior of the humans and their implications with respect to future studies. Although no instructions were given to the humans regarding how to communicate with the robot, we expected all humans would use

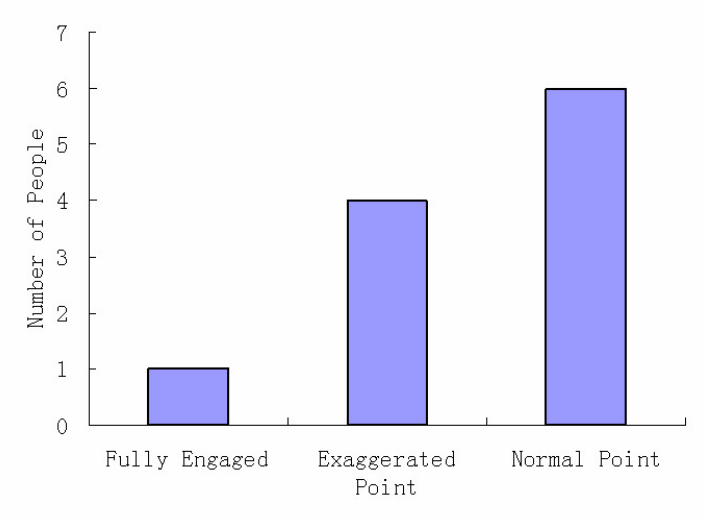


Figure 6.10. Human social behavior comparison plot shows distribution of people who exhibited different social behavior during the course of the experiment.

with the same pointing gesture since the scenario is simple. The outcome supports our hypothesis and at the same time offers some interesting insights. Shown in figure 6.10, while all subjects used pointing gestures to convey the intention, there were variations. For instance, while most people performed pointing by lifting the arm and extending it in the direction of the desired object. However, one person used exaggerated motions to emphasize the direction of the point when 2 objects were used. More interestingly, when 4 objects were placed on the table, and the robot began to falter in recognitions the intended object, 3 more people altered their point gesture and exhibited exaggerated point behavior, seemingly adjusting their behavior in attempt to be more conspicuous. Of the ten people participated in the experiment, only one person was fully engaged, exhibiting multiple social behaviors that include exaggerated pointing, nodding and praise when the robot correctly identified the object and transported to her hand, and head-shaking when the robot chose the wrong one. The other participants seemed much less engaged and used exhibited only pointing behavior.

One possible explanation for this observation is that because the uBot was placed on a wooden stand and the actions were limited, most people did not quite view the uBot as a social partner but rather a machine capable of moving objects. The person who interacted with the uBot enthusiastically has a one-year old child and therefore automatically entered into her mother-ese mode where she would use exaggerated motion as well as different pitch of voice for interacting with children. While it is sufficient for this experiment if the human only used one modality for conveying the intention, it is however desirable to have people who exhibit more of their natural social skills and participate actively for future learning sessions. This Give Robot opportunities to learn about different social behaviors and how to react to them. Our experience from this experiment may influence our experimental design decisions in future studies.

6.4 Summary

This chapter is part of an effort to create a consistent framework for robots to learn communicative behavior for assisting with humans in daily lives. While previous chapters focused on how the framework enables a robot to learn expressive behaviors and behavioral affordance-based models of humans in an incremental manner, this work concentrates on the reciprocal problem: how can a robot recognize the same behavior performed by a human, and learn the receptive behavior to address that need. For this study, knowledge reuse continues to play a key role as it did in our previous examples. With the shared knowledge representation, we extended the idea to enable robots to reuse existing skills as behavioral templates and extract important cues for the purpose of recognition. Upon detecting similar cues from a human, the robot can then infer the human's goal by reflecting on its own goals and intentions when it performed the behavior. Given the extracted goal, the robot can thus explore and find policies to meet the human's needs. A case study has been presented to

demonstrate the feasibility of this approach. It shows that sufficient information can be extracted from a robot's own pointing behavior for the bi-manual robot to recognize pointing behavior exhibited by various subjects, which led to the successful negotiation of proper assistive behavior that meets the human's needs. Observations of human behavior from this study also provide insights for the design of future, more complicated learning stages.

CHAPTER 7

CONCLUSIONS

Communication skills are needed for robots to collaborate and assist humans in daily activities. Despite several state-of-art robots have incorporated a number of social skills for this purpose, these skills are either hard-coded behavior or simple replays of pre-defined motion trajectories. For my dissertation, I have presented an approach that enables robots to learn these communication skills from first principles and showed that a robot can learn these skills in an incremental fashion by adapting to increasingly challenging interactions with humans.

On a technical level, the presented approach tackled the following important challenges: first, human beings are sophisticated objects that have been proven difficult to model. Second, humans are independent agents with its own goals and activity, robots need ways to learn skills for directing human attention, expressing intention and soliciting human assistance. Third, previous attempts on developing communicative skills for robots have different methodologies and representations for expressive and receptive behavior. As a result, knowledge reuse is rare. However, for humans, to develop complex skills or solve challenging problems we often take advantage of previously acquired skills and knowledge. A unified framework for developing communication skills that supports knowledge reuse and transfer has been presented in this dissertation.

7.1 Contributions

To address these challenges, this dissertation has made contributions in the following areas::

1. **A unique approach for robots to learn and represent humans in terms of behavior the humans afford**. Although this affordance modeling approach has recently gained attention in the developmental robotics community, most work still only focuses on simple objects. Few attempts have been made on complicated, articulated objects, let alone "objects" with independent motions such as humans. This work has both outlined a computational framework for the affordance modeling of humans, and provided learning examples to show how a robot can construct an increasingly complex model of humans as the robot accumulates manual, communicative skills. Beginning with an initial concept that a human is just a big motion segment that moves, the robot extended this simple concept into complex kinematic structures that afford simultaneous tracking. As the robot gathers more interaction experience, the visual tracking model was expanded to include social behavioral patterns such as the observation that a human is likely to offer assistance to out-of-reach objects if a pointing gesture is used.

2. **The extension and application of behavioral learning framework intended for developing manual skills for the purpose of learning communicative behavior**. A change of operating context for teaching a robot communicative behavior in the presence of humans presented several challenges: to adapt the new contexts the robot must learn in a new state space, consider using new actions and handle situations where local adjustments of the original policy is no longer adequate. To address these issues, a prospective learning algorithm has been presented to enhance the framework's ability to adapt to new contexts while maintaining as much of the previous acquired knowledge

structure as possible. This framework compliments other efforts in the field by providing a grounded means of learning social behavior. For learning receptive behavior, this work also proposed a unique off-policy affordance modeling approach that enables the robot to exploit the symmetry in communicative behavior such that knowledge in existing behavior can be reused.

3. **The proposal of a developmental trajectory for robots to acquire communication skills to interact with humans**. Previous attempts at developing communication skills treats expressive and receptive skills separately. The use of the proposed developmental trajectory facilitates learning and promotes knowledge reuse and transfer between these interrelated processes. The process begins with the robot first learning various manual skills through intrinsically motivated exploration. Next, by subjecting the robot to conditions of mutual reward and underactuation, expressive communicative behavior emerges naturally as the robot discovers the utility of manual behavior under the new context. Finally, receptive behavior is learned by reusing existing manual skills and knowledge structure gathered during the expressive behavior learning process.

7.2 Discussions and Future Work

Together, these contributions form a unique approach for robots to autonomously develop non-verbal communication skills from on-line interactions with human partners. Compared with the prevalent programming approaches, the approach presented in this dissertation has the advantage of being adaptive to unexpected human responses, to skill transfer onto a different robot, and to different preferences of human users. More importantly, compared with existing learning-based methods such as programming-by-demonstration (PbD) or learning from imitation, the approach presented in this dissertation is grounded. While the end goal of PbD and learning from demonstrations research generally focus on producing models to classify motion

trajectories, poses or configurations, the learning framework in this work focuses on associating actions with goals and intentions using intrinsically motivated learning: the robot simply seeks to sequence actions that increase its chances for achieving reliable reward. On the other hand, as the result of the reward/goal oriented behavior learning, expressive behavior is used as a template that forms the basis for receptive behavior learning. Using methods presented in this thesis the existing methods for producing expressive communicative gestures can be also grounded, since it is conceivable for robots to explore and learn to associate reward and intentions with actions produced by, PbD for instance, using the same behavioral learning framework.

Second, this work demonstrated the benefits of the proposed approach of studying several problems—the learning of manual skill, expressive and receptive communicative behavior—in a consistent framework. Effective knowledge reuse and transfer has been illustrated as manual behavior is reused in the context of mutual reward and underactuation such that its communicative nature is revealed and archived. Then, the archived expressive communicative behavior is again reused as a template to facilitate the recognition process of the same behavior when performed by a human. Furthermore, this work demonstrated efficient knowledge transfer of learned behavior from one robot to another. Since only the structure of the behavior is transferred, the robot is able to select the appropriate resources based on its own experience, rather than simply applying the resource that may not be applicable. For instance, when transferring pointing from Dexter to the uBot, the uBot can rely on its own procedure knowledge for determining when a two-handed reach is appropriate rather than rigidly use Dexter's knowledge. Lastly, as the robot accumulates manual skills for interacting with objects, and expressive and receptive behavior for interacting with humans, this work also showed how these behaviors can be stored incrementally and reused for the purposes of detecting humans and tracking articulated human motions.

Observations of human reactions from the experiments offer interesting insights. For instance, results from expressive behavior learning showed that in some situations, knowledge about robots in general may bias the human's expectation of the robot and therefore can affect learning performance negatively. Naive participants typically reacted instinctively and performed better. The same set of experiments also revealed that communicative behavior built on top of robust manual skills is useful for maintaining the interest of the human participant. Also, evaluations of the learned expressive behavior showed how timing of actions from the robot can cause unexpected responses from naive humans and therefore lowering the overall effectiveness of certain gestures. Although not yet addressed in this work, it can be extrapolated that once more manipulation behavior becomes available to the robot, the problem will be eventually resolved as the robot explores the utilities of these new actions. This further strengthens the necessity of a learning approach taken by this work. Similarly, results from the receptive behavior experiments suggested that people who have young children are more engaged with the robot and thus more often produced a wider variety of social behavior in response to the robot's actions. Therefore they may be more preferable participants for later stages since this would give the robot more opportunities to learn about different social behaviors and how to react to them. However, more carefully designed human subject studies would be needed for these conclusions to be evaluated in a more rigorous manner .

This works shows an interesting developmental trajectory suitable for robots to develop non-verbal communication skills: from manual skills emerge expressive communicative behavior, and that the learning of expressive skills preceded receptive skills. This trajectory is supported by evidence from the psychology literature that motor skill development preceded the emergence of communication behavior, and that the concept of others is not developed till later stages. Even when different modalities of communication are considered, this trajectory still seems to apply. For instance,

even crying in some sense is also a motor skill, albeit a built-in one, this basic motor skill has to exist before the infant can learn to use it as an effective form of communication. However, it is unclear whether the reason for such a trajectory existed is the same as the reasons (benefits) that motivated this work. There may also exist other trajectories that lead to the same results. It is conceivable that some gestures may be developed as the result of motor-babbling and some are simply genetically built-in through evolution. For instance, infants at a very young age has been shown to be responsive to the adult's protruding tongue motion with that of their own. This is a possible future research direction for this work.

Another question that was not addressed by this thesis is related to a similar, but much more complex skill—verbal communication, i.e., language. Although this thesis focuses on the development of non-verbal communication skills only, the proposed approach as well as the findings may extend to the verbal domain, e.g. the advantage of a consistent framework or the robot's demonstrated ability to sequence actions to maximize reward. This is because studies from psychology have suggested that the human infants' capacity to learn complex sequence actions in manipulation tasks and their subsequent interest in object-object relationships allowed humans to eventually develop complex systems of communication, including language, since sequencing behavior (utterances) and associating the causal outcome are the key to developing effective verbal communication skills as well. Although these issues are not addressed in this dissertation, the resulting robotic platform and the learning framework provide a formal vehicle using which these questions can be studied.

BIBLIOGRAPHY

[1] Asada, M., MacDorman, K., Ishiguro, H., and Kuniyoshi, Y. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems 37*, 2 (2001), 185–193.

[2] Asadi, M., Papudesi, V. N., and Huber, M. Learning skill and representation hierarchies for effective control knowledge transfer. In *ICML 2005 Workshop on Structural Knowledge Transfer for Machine Learning* (Pittsburgh, PA, 2006).

[3] Atkeson, C., and Schaal, S. Robot learning from demonstration. In *Intenational Conference on Machine Learning* (1997).

[4] Bates, E., and Dick, F. Language, gesture, and the developing brain. *Developmental Psychobiology* (2002).

[5] Berlin, Matt, Gray, Jesse, Thomaz, Andrea Lockerd, and Breazeal, Cynthia. Perspective taking: An organizing principle for learning in human-robot interaction. In *AAAI* (2006), AAAI Press.

[6] Blake, Andrew, and Isard, Michael. The CONDENSATION algorithm - conditional density propagation and applications to visual tracking. In *NIPS* (1996), Michael Mozer, Michael I. Jordan, and Thomas Petsche, Eds., MIT Press, pp. 361–367.

[7] Bradshaw, J.L., and Rogers, L.J. *The evolution of lateral asymmetries, language, tool use, and intellect.* Academic Press, 1993.

[8] Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Lieberman, J., H. Lee, A. Lockerd, and Mulanda, D. Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robotics* (2004).

[9] Breazeal, C., Gray, J., and Berlin, M. An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research* (2009).

[10] Breazeal, Cynthia, Buchsbaum, Daphna, Gray, Jesse, Gatenby, David, and Blumberg, Bruce. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artif. Life 11*, 1-2 (2005), 31–62.

[11] Calinon, S., and Billard, A. Stochastic gesture production and recognition model for a humanoid robot. In *Proceedings of the international Conference on Intelligent Robots and Systems* (2004), pp. 2769–2774.

[12] Chamero, A. An outline of a theory of affordances. *Ecological Psychology 15*, 3 (2003), 181–195.

[13] Coelho, J., and Grupen, R. A control basis for learning multifingered grasps. *Journal of Robotic Systems 14*, 7 (1997), 545–557.

[14] Coelho, J.A., Piater, J.H., and Grupen, R.A. Developing haptic and visual perceptual categories for reaching and grasping with a humanoid robot. In *First IEEE-RAS International Conference on Humanoid Robots* (Cambridge, MA, September 2000).

[15] Cohen, P., Chang, Y. H., and Morrison, C. T. Learning and transfering action schemas. In *Proceedings of IJCAI* (2007).

[16] Cootes, T. F., and Taylor, C. J. Active shape models: Smart snakes. In *British Machine Vision Conference* (1992), pp. 267–275.

[17] Darrell, Trevor, and Pentland, Alex. Space-time gestures.

[18] Davis, E. *Representations of Commonsense Knowledge*. Morgan Kaufmann, San Mateo, CA, 1990.

[19] Deegan, P., Thibodeau, B., and Grupen, R. Designing a self-stabilizing robot for dynamic mobile manipulation. In *Robotics: Science and Systems* (2006).

[20] Demiris, J., and Hayes, G. Imitation as a dual-route process featuring predictive and learning components: A biologically plausible computational model. 321–361.

[21] Dornaika, Fadi, and Davoine, Franck. Simultaneous facial action tracking and expression recognition using a particle filter. In *ICCV* (2005), IEEE Computer Society, pp. 1733–1738.

[22] Duhon, David, Weinman, Jerod, and Learned-Miller, Erik. Techniques and applications for persistent backgrounding in a humanoid torso robot. In *IEEE International Conference on Robotics and Automation (ICRA)* (2007).

[23] Edsinger, A., and Kemp, C. Human-robot interaction for cooperative manipulation: Handing objects to one another. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)* (2007).

[24] Edsinger, A., and Kemp, C. C. What can i control? a framework for robot self-discovery. In *Proceedings of the 6th International Workshop on Epigenetic Robotics* (2006).

[25] Fergus, Robert, Perona, Pietro, and Zisserman, Andrew. Object class recognition by unsupervised scale-invariant learning. In *CVPR (2)* (2003), pp. 264–271.

[26] Fitzpatrick, P., Metta, G., Natale, L., Rao, S., and Sandini, G. Learning about objects through action: Initial steps towards artificial cognition. In *IEEE International Conference on Robotics and Automation* (Taipei, May 2003).

[27] Fitzpatrick, Paul, Metta, Giorgio, Natale, Lorenzo, Rao, Ajit, and Sandini, Giulio. Learning about objects through action -initial steps towards artificial cognition. In *ICRA* (2003), IEEE, pp. 3140–3145.

[28] Frey, Scott .H. Tool use, communicative gesture and cerebral asymmetries in the modern human brain. *Philosophical Transactions of the Royal Society* (2008).

[29] Gallese, Vittorio, Rochat, Magali, Cossu, Giuseppe, and Sinigaglia, Corrado. Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Developmental Psychology 45*, 1 (January 2009), 103–113.

[30] Gaver, and W., William. Technology affordances. In *Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems* (1991), Use of Familiar Things in the Design of Interfaces, pp. 79–84.

[31] Gavrila, D. M. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding 73*, 1 (Jan. 1999), 82–98.

[32] Geib, C., Mourão, K., Petrick, R., Pugeault, N., Steedman, M., Krüger, N., and Wörgötter, F. Object action complexes as an interface for planning and robot control. In *Workshop 'Toward Cognitive Humanoid Robots' at IEEE-RAS International Conference on Humanoid Robots* (Genoa, Italy, 2006).

[33] Gergely, G. What should a robot learn from an infant? mechanisms of action interpretation and observational learning infancy. *Connection Science 15* (2003), 191–209.

[34] Gibson, J. J. The theory of affordances. In *Perceiving, acting and knowing: toward an ecological psychology* (Hillsdale, NJ, 1977), Lawrence Erlbaum Associates Publishers, pp. 67–82.

[35] Gibson, K.R., and Ingold, T. *Tools, language and cognition in human evolution.* Cambridge University Press, 1993.

[36] Gomez, G. Simulating development in a real robot. In *Proceedings of the 4th International Workshop on Epigenetic Robotics* (2004).

[37] Gray, J., Breazeal, C., Berlin, M., Brooks, A., and Lieberman, J. Action parsing and goal inference using self as simulator. pp. 202–209.

[38] Greenfield, P.M. Language, tools and the brain: the development and evolution of hierarchically organized sequential behavior. *Behavioral and Brain Sciences 95* (1991), 531.

[39] Hart, S., Sen, S., and Grupen, R. Generalization and transfer in robot control. In *Epigenetic Robotics Annual Conference* (2008).

[40] Hart, S., Sen, S., and Grupen, R. Intrinsically motivated hierarchical manipulation. In *Proceedings of 2008 IEEE Conference on Robots and Automation (ICRA)* (2008).

[41] Hart, Stephen. *The Development of Hierarchical Knowledge in Robot Systems*. PhD thesis, University of Massachusetts Amherst, 2009.

[42] Hasanuzzaman, M., Ampornaramveth, Vuthichai, Kiatisevi, Pattara, Shirai, Yoshiaki, and Ueno, Haruki. Gesture based human-robot interaction using a frame based software platform. In *IEEE International Conference on Systems, Man and Cybernetics* (2004), IEEE, pp. 2883–2888.

[43] Hayes, P. J. The naive physics manifesto. In *Expert Systems in the Micro-Electronic Age*, D. Michie, Ed. Edinburgh University Press, 1978.

[44] Hjelmås, Erik, and Low, Boon Kee. Face detection: A survey. *Computer Vision and Image Understanding 83*, 3 (2001), 236–274.

[45] Hobbs, J.R., and Moore, R.C., Eds. *Formal Theories of the Commonsense World*. Ablex, Noorwood, NJ, 1985.

[46] Hogg, D. C. Model-based vision: A program to see a walking person. *Image and Vision Computing 1*, 1 (Feb. 1983), 5–20.

[47] Huber, M. *A Hybrid Architecture for Adaptive Robot Control*. PhD thesis, Department of Computer Science, University of Massachusetts Amherst, 2000.

[48] Ioffe, Sergey, and Forsyth, David. Human tracking with mixtures of trees. pp. 690–695.

[49] Jacob, and K., Robert J. What you look at is what you get: Eye movement-based interaction techniques. In *Proceedings of ACM CHI'90 Conference on Human Factors in Computing Systems* (1990), Eye, Voice and Touch, pp. 11–18.

[50] Jenkins, O., and Mataric, M. Primitive-based movement classification for humanoid imitation. In *Tech. Report IRIS-00-385* (2000).

[51] Jenkins, O., and Mataric, M. Primitive-based movement classification for humanoid imitation. In *Tech. Report IRIS-00-385* (2000).

[52] Jenkins, Odest, and Mataric, Maja J. A spatio-temporal extension to isomap nonlinear dimension. In *In The International Conference on Machine Learning (ICML 2004* (2004), pp. 441–448.

[53] Jones, J.L., and Lozanzo-Perez, T. Planning two-fingered grasps for pick-and-place operations on plyhedra. In *Proceedings of 1990 Conference on Robotics and Automation* (1990).

[54] Jones, Michael J., and Rehg, James M. Statistical color models with application to skin detection. In *CVPR* (1999), IEEE Computer Society, pp. 1274–1280.

[55] Kaplan, F., and Hafner, V. V. Mapping the space of skills: An approach for comparing embodied sensorimotor organization. In *Proceedings of the 4th IEEE International Conference on Development and Learning* (2005).

[56] Katz, Dov, and Brock, Oliver. Extracting planar kinematic models using interactive perception. In *Robotics: Science and Systems* (2007).

[57] Kimura, D. Neuromotor mechanisms in the evolution of human communication. *In Neurobiology of Social Communication in Primates: An Evolutionary Perspective* (1979).

[58] Kjellström, Hedvig, Romero, Javier, Mercado, David Martínez, and Kragic, Danica. Simultaneous visual recognition of manipulation actions and manipulated objects. In *ECCV (2)* (2008), pp. 336–349.

[59] Koenderink, Jan J. The structure of images. *Biological Cybernetics* (1984).

[60] Kolsch, M., and Turk, M. Fast 2D hand tracking with flocks of features and multi-cue integration. In *Vision for Human-Computer Interaction* (2004), p. 158.

[61] Kölsch, Mathias, and Turk, Matthew. Robust hand detection. In *FGR* (2004), IEEE Computer Society, pp. 614–619.

[62] Kulic, Dana, and Nakamura, Yoshihiko. Scaffolding on-line segmentation of full body human motion patterns. In *The IEEE/RSJ International Conference on Robots and Systems (IROS)* (2008), IEEE, pp. 2860–2866.

[63] Lee, M. H., and Meng, Q. Psychologically inspired sensory-motor development in early robot learning. *International Journal of Advanced Robotics Systems 2*, 4 (2005), 325–333.

[64] Lee, Mun Wai, and Cohen, Isaac. Proposal maps driven MCMC for estimating human body pose in static images. In *CVPR (2)* (2004), pp. 334–341.

[65] Lenat, D.B. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM 38*, 11 (1995), 33–38.

[66] Lozano-Perez, T. Automatic planning of manipulator transfer movements. In *Trans. Syst. Man, Cybern.* (oct 1981), vol. SMC-11, pp. 681–698.

[67] M. Doniec, G. Sun, and Scassellati, B. Active learning of joint attention. In *IEEE/RSJ International Conference on Humanoid Robotics* (2006).

[68] Mataric, M. Sensory-motor primitives as a basis for imitation: Linking perception to action and biology to robotics. In *Imitation in animals and artifacts*. MIT Press, 2002, pp. 391–422.

[69] McCarty, M.E., Clifton, R.K., and Collard, R.R. Problem solving in infancy: The emergence of an action plan. *Developmental Psychology 35*, 4 (1999), 1091–1101.

[70] McGrenere, Joanna, and Ho, Wayne. Affordances: Clarifying and evolving a concept. In *Graphics Interface* (May 2000), pp. 179–186.

[71] Metta, G. *Babybot: A Study Into Sensorimotor Development*. PhD thesis, LIRA-Lab (DIST), 2000.

[72] Min, F., Suo, J. L., Zhu, S. C., and Sang, N. An automatic portrait system based on and-or graph representation. In *EMMCVPR* (2007), pp. 184–197.

[73] Minsky, M. A framework for representing knowledge. Memo 306. MIT-AI Lab, MIT, June 1974.

[74] Mori, Greg, Ren, Xiaofeng, Efros, Alexei A., and Malik, Jitendra. Recovering human body configurations: Combining segmentation and recognition. In *CVPR (2)* (2004), pp. 326–333.

[75] Morris, D., Collett, P., and Marsh, P. *Gesture, their origins and distribution*. Gestures, their origins and distribution, 1979.

[76] Mutlu, B. A storytelling robot: Modeling and evaluation of human-like gaze behavior. under review. In *International Conference on Humanoid Robots* (2006).

[77] Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (2009).

[78] Nakamura, Y. *Advanced Robotics: Redundancy and Optimization*. Addison-Wesley, 1991.

[79] Nam, Yanghee, Korea, Taejeon, and Wohn, KwangYun. Recognition of space-time hand-gestures using hidden markov model. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology* (1996).

[80] Natale, L. *Linking Action to Perception in a Humanoid Robot: A Developmental Approach to Grasping*. PhD thesis, LIRA-Lab, DIST, University of Genoa, 2004.

[81] Newell, A., and Simon, H. A. GPS: A program that simulates human thought. In *Lernende Automaten*. MIT Press, Munich, Oldenbourg KG, 1961.

[82] Oates, Tim. *Grounding Knowledge in Sensors: Unsupervised Learning for Language and Planning*. PhD thesis, University of Massachusetts Amherst, 2001.

[83] Oztop, E., and Arbib, M. Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics* (2002), 116–140.

[84] Pfeifer, R. Robots as cognitive tools. *International Journal of Cognition and Technology 1* (2002), 125–143.

[85] Piater, J. H., and Grupen, R. A. Constructive feature learning and the development of visual expertise. In *Proceedings of the Seventeenth International Conference on Machine Learning* (Stanford, CA, 2000).

[86] Prinz, W. A common coding approach to perception and action. 167–201.

[87] Ramanan, D., and Forsyth, D. Finding and tracking people from the bottom up, 2003.

[88] Ren, X. F., Berg, A. C., and Malik, J. Recovering human body configurations using pairwise constraints between parts. In *International Conference on Computer Vision* (2005), pp. I: 824–831.

[89] Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* (1996).

[90] Roy, D. *Learning Words from Sights and Sounds: A Computational Model*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1999.

[91] Roy, D. Grounding words in perception and action: computational insights. 389–96.

[92] Roy, D., and Pentland, A. Learning words from sights and sounds: A computational model. *Cognitive Science* (2003).

[93] Sandini, G., Metta, G., and Konczak, J. Human sensori-motor development and artificial systems. In *Proceedings of AIR & IHAS* (Japan, 1997).

[94] Sanmohan, K., and Kruger, Volker. Primitive based action representation and recognition. In *SCIA* (2009), pp. 31–40.

[95] Saxe, David M., and Foulds, Richard A. Toward robust skin identification in video images. In *FG* (1996), IEEE Computer Society, pp. 379–384.

[96] Scassellati, Brian. Theory of mind for a humanoid robot. *Autonomous Robots 12*, 1 (2002), 13–24.

[97] Schaal, S. the new robotics - towards human-centered machines.

[98] Schmidhuber, J. Curious model-building control systems. In *Proceedings of the International Joint Conference on Neural Networks* (1991).

[99] Schmidhuber, J. A possibility for implementing curiosity and boredome in model-building neural controllers. In *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (1991).

[100] Schmidhuber, J., and Storck, J. Reinforcement driven information acquisition in nondeterministic environments. Tech. rep., Fakultat fur Informatik, Technische Universit at Munchen, 1993.

[101] Shapiro, S. D. Generalizing the hough transform. In *ICPR* (1978), pp. 710–714.

[102] Sigal, L., and Black, M. J. Measure locally, reason globally: Occlusion-sensitive articulated pose estimation. In *IEEE Computer Vision and Pattern Recognition or CVPR* (2006), pp. II: 2041–2048.

[103] Sinapov, J., and Stoytchev, A. Toward interactive learning of object categories by a robot: A case study with container and non-container objects. In *Proceedings of the IEEE International Conference on Development and Learning (ICDL)* (2009).

[104] Sinapov, J., Wiemer, M., and Stoytchev, A. Interactive learning of the acoustic properties of household objects. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA)* (Kobe, Japan, 2009).

[105] Singh, P. The public acquisition of commonsense knowledge. http://www.openmind.org/commonsense/pack.html, 2001. The Open Mind Commonsense project, 2001.

[106] Smets, Gerda, Overbeeke, Kees, and Gaver, William. Form-giving: Expressing the nonobvious. In *Proceedings of ACM CHI'94 Conference on Human Factors in Computing Systems* (1994), vol. 1 of *Expressive Interfaces*, pp. 79–84.

[107] Starner, T. E., and Pentland, A. P. Real-time american sign language from video using hidden markov models. In *Motion-Based Recognition* (1997), p. Chapter 10.

[108] Steels, L. The talking head experiments. *Laboratorium. Limited-Preedition* (1999).

[109] Steels, L., and Kaplan, F. Aibo's first words: The social learning of language and meaning. *Evolution of Communication* (2001), 4:3–32.

[110] Steels, L., and Vogt, P. Grounding adaptive language games in robotic agents. In *Proceedings of the 4th European Conference on Artificial Life* (1997).

[111] Stoytchev, A. Toward learning the binding affordances of objects: A behavior-grounded approach. In *Proceedings of the AAAI Spring Symposium on Developmental Robotics* (Stanford University, 2005).

[112] Sutton, R., and Barto, A. *Reinforcement Learning*. MIT Press, Cambridge, Massachusetts, 1998.

[113] Trafton, J. Gregory, Cassimatis, Nicholas L., Bugajska, Magdalena D., Brock, Derek P., Mintz, Farilee, and Schultz, Alan C. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics 35*, 4 (2005), 460–470.

[114] Triesch, J., and von der Malsburg, C. Classification of hand postures against complex backgrounds using elastic graph matching. *Image and Vision Computing 20*, 13-14 (Dec. 2002), 937–943.

[115] Viola, Paul A., and Jones, Michael J. Robust real-time face detection. In *ICCV* (2001), p. 747.

[116] Vygotsky, L. *Mind in society*. Harvard University Press, 1930.

[117] Wheeler, D., Fagg, A., and Grupen, R. Learning prospective pick and place behavior. In *Proceedings of the IEEE/RSJ International Conference on Development and Learning* (2002).

[118] Woodward, Amanda. Infants' grasp of others' intentions. *Current Directions in Psychological Science 18* (2009), 53–57.

[119] Yang, M. H. *Hand Gesture Recognition and Face Detection in Images*. PhD thesis, University of Illinois, Urbana-Champaign, 2000.

[120] Zhai, Shumin, Milgram, Paul, and Buxton, William. The influence of muscle groups on performance of multiple degree-of-freedom input. In *CHI* (1996), pp. 308–315.

[121] Zhu, S. C., and Mumford, D. *A Stochastic Grammar of Images*. World Scientific, 2007.