Cognitive Modeling of Human Intention Recognition

Elisheva Bonchek-Dokow

Interdisciplinary Studies Unit
Brain Sciences

Ph.D. Thesis

Submitted to the Senate of Bar-Ilan University

Ramat-Gan, Israel June, 2011

This work was carried out under the supervision of Prof. Gal A. Kaminka (Department of Brain Sciences), Bar-Ilan University

Acknowledgements

This work would not have come to fruition without the material and emotional support of close friends and family. Most notably, my parents, Avigdor and Shulamit Bonchek, and my parents-in-law, Shmuel and Yona Dokow, and of course my husband, Gidon. I am indebted to them for all they have done and gone through for me during the past six years—and before that, as well. I hope their pride in me will serve as adequate reward. If not, I do not know how I can ever repay them.

Thanks also to the past and present members of the Maverick lab, especially to my room-mates, Natalie, Mor and Avishai, whose encouragement and sympathy, as well as their practical assistance, were crucial to me, and to Gabriella, whose positive demeanor and help with technical matters were a blessing. Most of all, to my advisor, Prof. Gal Kaminka, who was enthusiastic and understanding throughout my work. I do not know if I would have reached this final stretch with any other advisor.

Last, but not least, I am full of gratitude to Hashem for the truly meaningful outcome of these past years—our three beautiful children, Amitai, Shaked and Talya. I hope that when they grow up and understand the price they paid, they will agree that it had been worthwhile.

I would like to acknowledge the generous sponsorship of the **Paula Rich Foundation** whose support has made possible my advanced-degree studies and research within the Doctoral Fellowships of Excellence Program.

This research was supported in part by grants from IMOD, the Israeli Science Foundation (ISF), and AFRL/EOARD.

Contents

List of Figures

List of Tables

Abstract

Human beings, from the very young age of 18 months, have been shown to be able to extrapolate intentions from actions [Meltzoff, 1995]. That is, upon viewing another human executing a series of actions, an observing child can guess the underlying intention, even before the goal has been achieved, and even when the performer failed at achieving the goal. In this work, we propose a cognitive model of this human ability, namely, that of intention recognition.

The proposed model deals with the challenge of recognizing the intention of an observed sequence of actions, performed by some acting agent. Intention recognition is apparently one of the core components of social cognition. Such a model is therefore important both from a cognitive science point of view and from an engineering perspective. It could provide a deeper understanding of normal and pathological development of human social cognition processes, as well as allowing for artificial implementation of this ability in software agents and physical robots, towards the end of creating more socially intelligent artificial beings.

Much work has already been done in all the many areas touching upon this topic, from psychology through neuroscience to artificial intelligence and engineering. In this work we aim to address those aspects of intention recognition which have not yet been treated satisfactorily. We provide a high-level overview of the process as a whole, and detail this model in a way which can explain how *failed* sequences of actions can be dealt with and their underlying intention extracted, and how *novel* objects can be dealt with, and goals regarding them predicted, although there is—seemingly—no prior knowledge about them.

We elaborate on two components of our proposed model, which we believe to be at its core, namely, those of intention detection and intention prediction. By *intention detection* we mean the ability to discern whether or not a sequence of actions has any underlying intention at all, or whether it was performed in an arbitrary manner with no goal in mind. By *intention prediction* we mean the ability to extend an incomplete sequence of actions to its most likely intended goal.

The overall structure of the model, i.e. its components and the connections between them, is justified by psychological theories and supported by a plethora of empirical results reviewed in the relevant literature. These

theories and experiments are referred to appropriately throughout this work. As for the two core modules on which we elaborate—Intention Detection and Intention Prediction—we present results from several experiments which we have designed and implemented for this purpose.

The Intention Detection module is based on a measure of intention, which captures a notion of efficiency, in keeping with the Principle of Rational Action [Gergely and Csibra, 2003], which states that intentions are brought about by the most rational means available to the actor. This module is validated by two experiments. The first is an artificial emulation of the original intention re-enactment procedure by Meltzoff [1995]. The results show that the proposed measure of intention indeed succeeds at categorizing streams of action according to the extent to which they convey an underlying intention.

The second experiment validating the Intention Detection module is closer to real life. It uses surveillance videos taken from an online database, and analyzes them according to the proposed measure of intention. This analysis is then compared to human judgment of intention on the same videos. The resulting correlation between the output of our module and that of the human subjects is high, showing once again that our measure of intention indeed captures the notion of intention present in action.

Like the Intention Detection module, the Intention Prediction module is based on a measure of intention as well. This measure is also designed to be in line with the Principle of Rational Action, however, it is formalized differently, for reasons which will be discussed. The Intention Prediction module also makes use of the psychological notion of *affordances*, for extracting goal states from objects in the environment, which the observed actions might possibly be intending to realize.

In order to test this second measure of intention as far as its usefulness for predicting intention, we designed an online experiment in which human subjects were presented with abstract objects (various geometric shapes), and were asked to predict the end-configuration of the objects, which observed sequences of movements were aiming to achieve. The predictions arrived at by our measure of intention were compared to the human results, and proved to be highly reliable. Other possible measures, such as proximity of the terminal state arrived at by the actions to the various goals, were also considered. However the success of these measures at predicting the intended goal was inferior to that of our measure of intention, and they at

most play a secondary role in the process.

To conclude our work, we summarize our findings and propose several directions for future research on intention modeling. We hope this work will be of interest and of use to researchers in the multidisciplinary communities dealing with intention recognition, and look forward to seeing the ideas proposed here implemented in socially cognitive artificial systems.

Chapter 1

Introduction

This work proposes a cognitive model of human intention recognition. It deals with the problem of recognizing the intention of an observed sequence of actions, performed by some acting agent. Intention recognition is apparently one of the core components of social cognition. Such a model is therefore important in at least two ways. First, from a cognitive science point of view, it could provide a basis for understanding human social cognition processes, perhaps leading to diagnosing and treating cases in which this ability is impaired, as well as furthering research on normal development. Second, from an engineering perspective, it could allow for artificial implementation of this ability in software agents and physical robots, thus enhancing their cognitive capacities, as well as socially natural interactions with humans.

While much work has been done both psychologically and computationally to further the understanding of the process of intention recognition, a complete model has not yet been proposed. Such a model should fit the vast amount of empirical data collected over the years, which describe the various aspects of the process—in particular, the ability of dealing with failed actions and novel objects, as will be explained. It should also be complete in the sense that it describes the process as whole, from a high level, as well as going down to the details of the mechanisms of each of the sub-processes of which it is composed.

For example, many experiments with children have been conducted in order to elucidate how and when intention recognition abilities develop. In particular, Meltzoff and colleagues have devoted much of their research to these questions.

A number of his experiments [Meltzoff, 1995], which serve as a main motivation for the work described here, highlight two specific challenges which have not yet been satisfactorily resolved by the research community. The first stems from the fact that children as young as 18 months of age are able to understand the intention underlying the actions of an observed adult even when these actions *fail* at bringing about the adult's intention. An added conundrum is the fact that the children are able to do this also when the actions are performed on *novel* objects, thus—seemingly—not relying on prior information regarding the objects.

This is what we have attempted in this work. Firstly, to outline the necessary components of a cognitive model of human intention recognition, and the interactions between them, as arises from the psychological data. Secondly, to detail those components which are at the core of the process, and for which existing models in the literature are lacking.

We next give an overview of our proposed model (Section 1.1), and then briefly describe the two main components of the model, which are at the core of the process Section 1.2). To conclude this introduction, we describe the organization of this dissertation.

1.1 A Model of Human Intention Recognition

The proposed model is schematically described in Figure 1.1. It consists of several modules—Intentional Being Detector, Intention Detector, Affordance Extractor, Success Detector, and Intention Predictor—connected between them by flow of relevant information from one to another. The input to the process as a whole consists of the acting agent A and the state-trace induced by its observed actions $s_0, s_1, ..., s_n$. The desired output is a goal state most likely intended by the acting agent. In the following chapters theoretical justification will be given for the modules and the connection between them, and empirical evidence will be provided for those two modules which are at the core of the process, as we understand it. We present here a brief overview of the model, with elaboration on, and justification for, each of the modules left for Chapter 3.

As Figure 1.1 shows, the process begins with the perception of an agent performing actions within an environment. This is the input. The expected output is a goal which is most likely intended by the actor. First, the observing agent

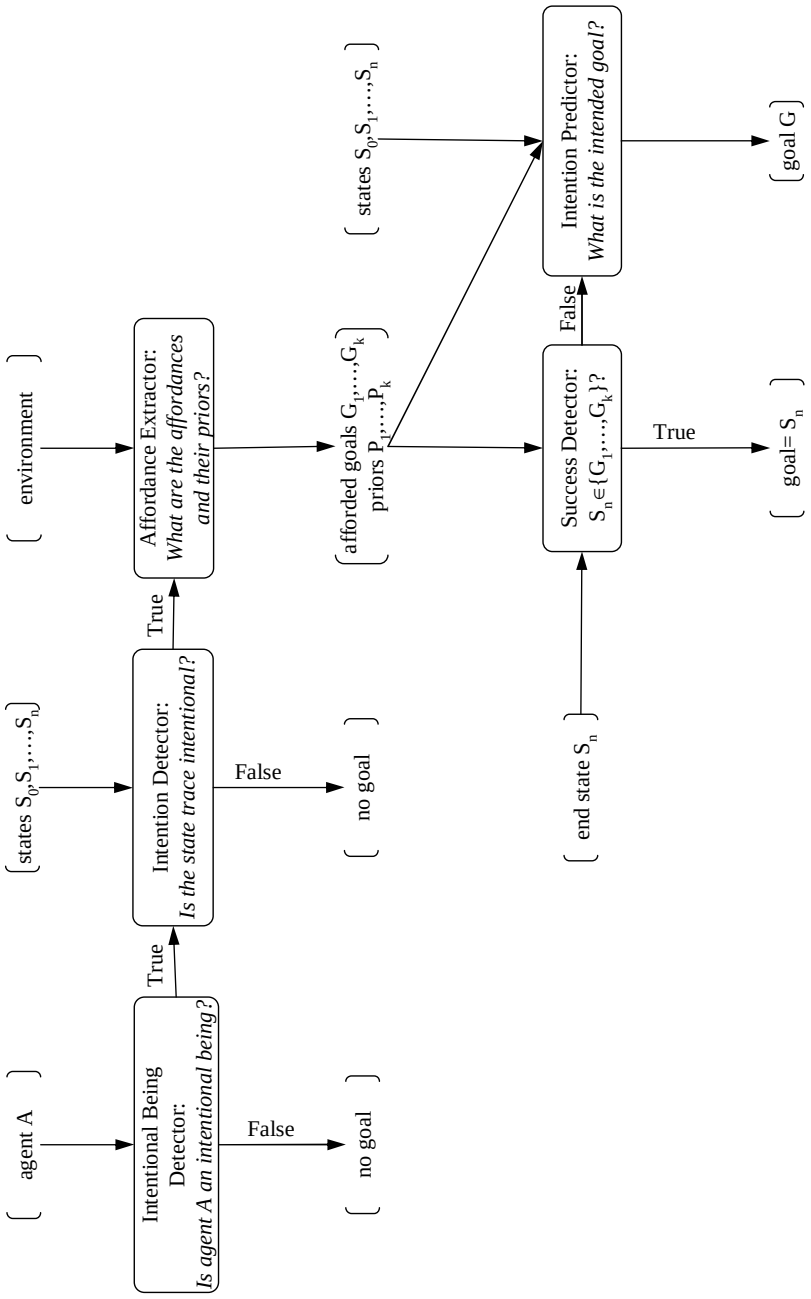


Figure 1.1: Scheme of Proposed Model.

determines whether or not the acting agent is at all capable of intention. If the answer is ”no”, there is no point in continuing the process, and it is terminated.

If the answer is ”yes”, the observing agent determines whether this particular instance of actions is being performed intentionally or not. Answering this question—detecting intention—is one of the core modules we elaborate upon in this work. Again, if the answer is ”no”, the process is terminated, since there is no goal to look for.

If the answer is ”yes”, that is, the actions are identified as intentional, the intended goal must now be predicted. This can be done online—while the actions are being performed, before the acting agent has achieved its goal, or offline, after the acting agent has stopped acting, and the observing agent can ask whether or not the terminal state at which the actor has stopped is its intended goal. We specifically deal with the possibility that the actor failed at bringing about its goal, and want our model to be able to detect these cases and ”fix” them, i.e., correctly predict what the actor was intending to do.

In order to answer the ”success or failure” question, we propose using a notion of affordances, which will discussed in detail later. Basically, affordances in our context are possible goal states which are likely to be performed on the objects in the environment. These are extracted from the environment, and then made use of in answering the discussed question. If the actions are deemed successful, the process can terminate with the answer that the achieved terminal state is the intended goal.

In the case that the actions are deemed to have failed, the observing agent must now guess what the intended goal was. This is the second of the two modules which are at the focus of this work.

The final output of the model is the intended goal—whether it has been successfully achieved by the acting agent or not—or an answer indicating either that the acting agent is not an intentional being, or that its actions were not performed intentionally.

1.2 Detection and Prediction of Intention

The phrase ”recognition of intention” could mean either recognition of the presence of intention, i.e. recognizing that the observed actions are intentional, or

prediction of the intention, i.e. recognizing what the intention of the actions actually is. In this work, we deal with both of these meanings, as two separate, sequential, processes. Each of these processes consists of one of the core modules, as mentioned in the precious section.

These two processes of detection and prediction are conceptually and practically distinct: according to the proposed model, given an observed sequence of actions, the observing agent first decides whether the actions were performed intentionally or not. This is what we refer to as *detection of intention*. Next, the agent goes on to determine the content of the intention, a stage which we name *prediction of intention*. Prediction—since the agent must determine the intention before it has been realized, as in the case where the actions resulted in failure.

The importance of this distinction is first and foremost in explaining and describing the process of intention recognition, as it appears in humans. An experiment in developmental psychology, which serves as a major motivation for this work (and is brought in detail later on in Section 2.4), shows that children do not always choose to imitate the intended goal of an acting adult. We propose that the determining factor in the decision to imitate or not to imitate is the perceived presence of intention. When the participating children detected intention in the actions of the adult, they made the effort to guess what that intention was, and then went on to imitate it. while, if no intention was detected by them, they did not trouble themselves to imitate the actions of the adult.

In addition, this distinction could prove useful in computational implementations of the process. Attempting to predict the intention of an acting agent when no such intention is present would both be wasteful in terms of computational resources, as well as result in a wrong answer.

The contribution of this work is meant to be both in the model itself, i.e. in the description of the process as whole, as well as in the detailing of the two main modules of detection and prediction.

1.3 Organization of this Dissertation

In Chapter 2, we bring the necessary theoretical background for understanding and justifying the proposed model, as well as work relating to it. We highlight what has already been done and what is yet considered as open. We touch upon

several seemingly unrelated topics, and therefore ask the reader to keep in mind the above outline of the model, in order to understand how these topics are related to the subject of intention recognition.

Next, in Chapter 3, we make use of the theoretical background in elaborating on the different modules of the model, with focus on the two main modules of intention detection and intention prediction. For these two modules, we describe the methods used for evaluating them.

We then present two chapters of experiments, Chapter 4 one for intention detection and Chapter 5 for intention prediction. Each contains a description of the experimental setup and the results along with their consequences.

Finally, in Chapter 6, we discuss the significance of the results, highlighting several aspects such as the use of different measure of intention and the role of repetition. Possible directions for future research are also suggested.

1.4 Publications Resulting from this Dissertation

- ”Distinguishing Between Intentional and Unintentional Sequences of Actions”, with Gal A. Kaminka and Carmel Domshlak. In Proceedings of the 9^{th} International Conference on Cognitive Modeling (ICCM-09), 2009.

- ”Distinguishing Between Intentional and Unintentional Sequences of Actions”, with Gal A. Kaminka and Carmel Domshlak. In Proceedings of the IJCAI-09 workshop on Plan, Activity, and Intention Recognition (PAIR-09), 2009.

- ”Distinguishing Between Intentional and Unintentional Sequences of Actions”, with Gal A. Kaminka. In preparation, to be submitted to the Journal of Cognitive Systems Research.

Chapter 2

Background and Related Work

Intention is at the focus of many fields of study, ranging from psychology and philosophy through neuroscience to artificial intelligence and robotics. In Section 2.1 we open the following review with references to those areas which touch upon intention, either researching it directly or using it in service of other branches of study.

After establishing the importance and applications of the study of intention, we turn to inquire into the different meanings of the term, since it is used in different contexts, and not always with the same meaning. Various characteristics of intention serve as potential candidates for solving the problem of intention detection (Section 2.2).

Next, in Section 2.3, we survey some recent major contributions to the subject of the prediction of intention. Some of these are in the context of the study of imitation, which is closely related to that of intention prediction. Most relevant is the work of Meltzoff, whose intriguing experiments on imitation serve as motivation for this dissertation. As such, a section is dedicated to a more detailed review of his experiments and findings (Section 2.4).

Finally (Section 2.5), we touch upon the notion of *affordances*. While not directly related to the subject of intention, use is made of it in the framework proposed here. For that reason, a brief review of this notion and its relevance to us is included here.

For each of the reviewed topics, we bring theoretical background which justifies its relevance to the task of intention recognition. In addition, we survey

computational models which attempt to implement them. Finally, after gathering information pertaining to the task at hand, we will go on, in the next chapter, to describe how all this information comes together in our framework for detecting and predicting intention.

2.1 Importance and Applications of Intention

Being able to infer the intention underlying the actions of agents within the environment is a valuable asset. This ability is part of what is known in psychology as *theory of mind*—the ability to understand the intentions, beliefs and desires of another. Theory of mind is what enables humans to interact socially in an intelligent manner. Individuals who posses a theory of mind which is impaired to a substantial degree—specifically in their ability to understand intentions—are often diagnosed as suffering from a disorder in the autism spectrum [Baron-Cohen, 1995].

Understanding the mechanisms underlying this ability is important for gleaning insights regarding autistic and normal development. In addition, they have practical significance in the field of engineering, software and hardware: the insights gained in neuroscience and psychology can be implemented in artificial systems, to the end of enhancing their interaction with humans in a socially intelligent, natural way. Such systems are useful for artificial household help, such as invalid assistance [Miro et al., 2009] as well as security systems, such as interception of enemies and tracking of adversaries [Foo et al., 2007].

2.2 Detecting Intention

To the best of our knowledge, most of the research conducted on computational recognition of intention deals with the problem of predicting the content of the intention, not with the problem of detecting whether or not intention is present at all in the observed actions. Yet several past investigations show that humans react differently to sequences of observation which convey some intention, than to sequences of arbitrary actions [Woodward, 1998, Gergely et al., 1995]. To address this challenge—that of discriminating between intentional and unintentional action—we must first define more rigorously what we mean by *intentional*

action, and what characterizes intentional actions as opposed to unintentional ones.

2.2.1 Definitions of Intentional Action

Throughout this work, the term *intentional action* refers to action which is performed by an agent with the purpose of bringing about some desired goal-state. Three key words should be emphasized here: action, purpose, and goal-state. Each such key word distinguishes our definition of intentional action from other possible definitions, and should be kept in mind throughout this work. This is important since the experiments presented here were designed according to this understanding of the term, and the results might not necessarily be relevant to other understandings of it.

As opposed to intention in action, other forms of intention, such as intention in thought or intention in speech, do not bring about observable changes in the world state (generally modeled as relationships between entities in the world). As such, they require different tools and mechanisms, and are out of the scope of this work.

The key word *purpose* is relevant to the relation of the notions of success and failure to the notion of intention. In this work we specifically address the possibility of failure in the execution of action, however, this does not render the actions intentionless. The actual outcome of the actions might indeed be unintended, i.e., a failed goal, yet the actions themselves were nevertheless performed with an intention in mind. According to this understanding, intentional actions terminating in failure or accidents would still be considered as intentional. The important criterion here is that there was a purpose which drove the actions, even if that purpose was not realized.

Compare this viewpoint to that of Harui et al. [2005], for example. While they too aim to distinguish between intentional and unintentional action in their proposed implementation, their distinction is actually between intentional outcomes and accidental outcomes. For this they made use of prosody and verbal utterances (such as ”oops”) and their timing within the stream of action. The distinction we wish to make here is of another kind: between action performed with a specific intention in mind, to action performed without any intention in mind.

There is a philosophical opinion, held by such as Husserl, which views all conscious action as intentional [Banchetti-Robino, 2004]. While this may be so, it is possible to distinguish between two types of intention, as expressed in action. There is one type of intentional action, in which it is the motion itself which is the goal of the actor. Consider, for example, a dancer: there is no end-state which the dancer is aiming to bring about, rather, the motion itself is the goal. Another type of intentional action is that which is performed for the purpose of bringing about a certain goal-state. In the context of this work, we limit ourselves to the second type, and in this sense, the first type would be considered unintentional.

2.2.2 Characteristics of Intentional Action

A preliminary condition that actions must fulfill in order to have intention attributed to them, is that they be performed by an intentional being. Experiments with children have shown that when the same actions are performed by a human and by a mechanical arm, the observing children tend to attribute intention only to the human, and not to the mechanical being [Meltzoff, 1995]. Apparently, from a very young age children are able to distinguish between intentional beings and non-intentional ones, and employ this distinction before attempting to decide whether the actions per se are intentional or not.

Several factors have been suggested in the literature which are relevant to the task of distinguishing between intentional and non-intentional agents. Among them are biological motion [Blakemore and Decety, 2001], texture of the appendage performing the action [Guajardo and Woodward, 2004] (i.e. skin versus gloved or metal object) and presence of eye gaze [Itakura et al., 2008].

Once the actor has been determined as capable of intentions, the actions themselves can be inspected for underlying intention. In order to be able to discriminate between intentional actions and unintentional ones, we must have a better understanding of what exactly intention is, and how it manifests itself in action. For this we turn to the psychological literature, mostly in the cognitive and developmental fields, in which there is an elaborate and extensive ongoing discussion on the topic, from Smith [1978] through Dennet [1989] to Gergely and Csibra [2003], to name just a few. Much insight can be gleaned from such research for the purpose of modeling and implementing the capacity of intention detection. We refer here to those works which supplied us with motivation and

justification for our work.

There are several psychological theories regarding the stance taken when dealing with intention. Meltzoff [2002] takes the mentalistic stance, that infants' ability to interpret intention makes use of an existing theory of mind—reasoning about the intents, desires and beliefs of others. Gergely and Csibra [2003], on the other hand, take a teleological stance, that infants apply a non-mentalistic, reality-based action interpretation system to explain and predict goal-directed actions. As Gergely and Csibra say themselves, this teleological evaluation should provide the same results as the application of the mentalistic stance as long as the actor's actions are driven by true beliefs. In their own words, "... when beliefs correspond to reality, the non-mentalistic teleological stance continues to be sufficient for interpreting action even after the mentalistic stance, which includes fictional states in its ontology, has become available" [Csibra and Gergely, 1998, p. 258]. The teleological interpretation would break down, however, if the interpreted actions were based on pretense or false beliefs. Since the scenarios we address here do not deal with false beliefs, and assume that the agent's beliefs correspond to reality, we can ignore this distinction for now and take Gergely and Csibra's psychological theories as motivation for our model, without decreeing which of the two stances humans actually take.

So what are the criteria which enable the attribution of intention—or, in teleological wording, goal-directedness—to movement? According to Gergely and Csibra, the *Principle of Rational Action* [Gergely and Csibra, 2003, Watson, 2005] plays a major role here. This principle states that intentional action functions to bring about future goal states by the most rational means available to the actor within the constraints of the situation.

Kiraly et al. [2003] break down the rationality principle into two assumptions which respectively provide two perceptual cues indicating goal-directedness. The first assumption is that the basic function of actions is to bring about some particular change of state in the world. This specifies that the outcome of the action should involve a salient change of state in the environment. When trying to determine whether the end-state arrived at is the intended goal or whether it is a failure or an accident, this could come in handy. The second assumption is that agents will employ the most efficient (rational) means available to them within the constraints of the situation. This specifies that the actor should be

capable of equifinal variation of actions, meaning that when the situational constraints change, the agent will take different actions in order to reach the goal efficiently. It is this second assumption which we will take advantage of here for our purposes. In attempting to determine whether an action sequence is intentional or not, we will be looking for efficiency—in time, space, effort, or any other resource utilized in the process.

Besides saliency of change of state and equifinality, several other features of intentional action have been proposed. Among them are affective vocal and facial cues [Carpenter et al., 1998], animate movement which is self-propelled, possibly along a nonlinear path and undergoing sudden changes of velocity [Blakemore and Decety, 2001], persistence (repeating actions over and over), and expending of effort [Heider, 1958]. In our proposed model we focus on the Principle of Rational Action and ignore for now other features, for reasons that are both practical and theoretical. On the practical side, the notion of efficiency is captured in a very straightforward way computationally. This, as opposed to such notions as ”effort”, ”animate movement” and ”affective cues”. On the theoretical side, we would like to isolate the minimal components which characterize intentional behavior. For example, research has shown that eliminating affective vocal and facial cues from the demonstration does not impair infants' ability to discern intention [Meltzoff, 1995]. Our results show that indeed rationality of action might be a strong enough indication of intention.

2.3 Predicting Intention

As opposed to intention detection (i.e. determining whether or not the observed actions were performed intentionally), intention prediction (i.e. guessing the goal which the observed actions were aimed at bringing about) has been the focus of much research, regarding both its appearance in humans and its implementation in artificial systems. We review here the major findings from psychology and neuroscience, and several important implementations in engineering.

2.3.1 Cognitive and Developmental Psychology

From the moment they are born, humans display a remarkable ability to imitate. Imitation is closely related to inference of intentions, in the sense that the latter is

a preliminary requirement for the former, when dealing with high-level imitation, i.e., goal imitation. Understanding how this mechanism works in humans, and modeling it, is one of the topics dealt with in the cognitive and developmental psychology research community.

2.3.1.1 Stages of Imitation

Our main inspiration and motivation comes from work by Meltzoff and colleagues [Metlzoff and Moore, 1983, Meltzoff, 1998, 1988, Meltzoff and Moore, 1989, Hanna and Meltzoff, 1993, Meltzoff, 1995, Meltzoff and Moore, 1997]. Based on many studies done on neonates, infants and toddlers, as well as adults, Meltzoff distinguishes between four stages in the development of imitative capabilities.

- Body Babbling: Coined by Meltzoff and Moore [1997], body babbling is the process of learning the correspondence between various muscle movements and their resulting body configurations, which begins already in utero. Much as babies start out with vocal babbling before they proceed on to comprehensible speech, so too they begin with random movements of limbs and facial parts before they are able to intentionally achieve a specific motion or pose.

- Imitating Body Movements: Once the infant has learned the basic motor primitives which enable it to achieve intended body configurations, it learns to recognize these when they are performed by an adult, and subsequently imitate them. Neonates have been documented [Metlzoff and Moore, 1983, Meltzoff and Moore, 1989], displaying the ability to imitate such facial expressions as tongue protrusion and frowning.

- Imitating Actions on Objects: At the age of several months, infants expand their imitation repertoire to include not only gestures and expressions, but also object manipulations [Meltzoff, 1998, 1988, Hanna and Meltzoff, 1993]. This enables the acquisition of knowledge of tool use and other important skills.

- Inferring Intentions: This last stage is the one which interests us. Meltzoff [1995] showed that 18-month old infants can imitate an intended act, even

when that act failed to be completed for some reason. The experimental setup had an adult attempt to demonstrate a novel action on a novel object, for example, pulling apart a two-piece dumbbell, whereby his hand slipped and the dumbbell remained as it was. When handed to the child, the child conveyed his ability to understand the intended action by re-enacting it successfully. Since this experiment serves as motivation for our work, we bring it in detail in the next section (2.4).

Other studies relating to this last stage of inferring intentions were done by Bekkering and colleagues. In one such study they showed that imitation of gestures in children is goal directed [Bekkering et al., 2000]. In another [Gergely et al., 2002], they showed how children could reason about whether the means were actually goals in and of themselves, and therefore worth imitating, or whether they were secondary to the true goal, and therefore could be substituted for by other means. These studies demonstrate the ability of young children to infer and reason about goals. However, we refer back to the above-mentioned work by Meltzoff [1995], which has the added value of showing how children can infer the goal even when it is only implied, and not successfully completed, and when it is performed on objects which are novel to the children. These are the two main challenges we deal with in our model.

2.3.1.2 Imitation and Theory of Mind

First introduced by Premack and Woodruff [1978], *theory of mind* (aka *folk psychology* and *mentalizing*) is the ability to attribute mental states (beliefs, intents, desires, etc.) to oneself and to others. As originally defined, it enables one to understand that mental states can be the cause of others' behavior, thereby allowing one to explain and predict the observed actions produced by others. As Meltzoff categorizes it, this ability enables a psychological attribution of causality to human acts, rather than the physical causality generally attributed to inanimate objects.

Different accounts are given by psychologists for the mechanism underlying this ability. One of them, known as simulation theory [Gordon, 1986, Davies and Stone, 1995, Heal, 2003], has gained popularity and credibility lately, in part due to the discovery of mirror neurons (see next section). In the words

of Breazeal et al. [2005], simulation theory posits that by simulating another person's actions and the stimuli they are experiencing using our own behavioral and stimulus processing mechanisms, humans can make predictions about the behaviors and mental states of others based on the mental states and behaviors that they themselves would possess if they were in the other's situation. In short, by thinking "as if" we were the other person, we can use our own cognitive, behavioral, and motivational systems to understand what is going on in the heads of others.

According to this explanation, the mutual relationship between imitation and theory of mind is clarified, and further elucidated by many studies [Meltzoff and Moore, 1992, 1994, 1995, Meltzoff and Decety, 2003, Meltzoff and Gopnik, 1993]. On the one hand, basic imitation of movement is a precursor to the development of theory of mind skills, by laying the foundations for what Meltzoff calls the "like me" framework for recognizing and becoming an intentional agent [Meltzoff, 2007]. Once the infant learns by imitation that his body, along with its inputs and outputs, is similar to those of the adults he sees around him, then he can simulate their behavior within his own mind. On the other hand, once this capacity is developed, theory of mind can be put to use for the explanation and prediction of actions observed. This would enable the level of imitation which requires inference of intentions and goals.

2.3.2 Neurophysiology

Mirror neurons, found in the premotor cortex of the macaque monkey, are activated both when the monkey performs a goal directed action, and when it perceives a conspecific performing the same actions. In humans, the existence of mirror neurons has not been verified, however mirror systems (i.e. brain regions, as opposed to single neurons) have been shown to be active during both perception and generation of motor actions. This was done using various methods of brain imaging, such as EEG [Cochin et al., 1998, 1999, Altschuler et al., 1997, Bekkering et al., 2000], MEG [Hari et al., 1998], TMS [Fadiga et al., 1995] and PET [Arbib et al., 2000]. This discovery paved the way for many computational models and implementations, attempting to produce imitation based on intention prediction.

Mirror neurons are one of the main neural mechanisms proposed for explain-

ing theory of mind capabilities in general, and understanding of intentions in particular [Rizzolatti et al., 1996, Gallese et al., 1996]. These neurons are especially important for imitation, where the ability to predict the intention at which the actions are aimed is necessary. Therefore, as in the domain of psychology, so in neurophysiology, research on imitation is tightly coupled with research on intention prediction.

It has been suggested that in addition to action recognition [Gallese et al., 1996], these neurons contribute to the functioning of imitation [Rizzolatti et al., 2001, Williams et al., 2001] and understanding of intentions [Fogassi et al., 2005], as well as to various other theory of mind abilities [Dapretto et al., 2005] and theories, such as the simulation account of theory of mind [Gallese and Goldman, 1998].

2.3.3 Implementations in Artificial Systems

2.3.3.1 Goal and Plan Recognition

A closely related yet conceptually distinct area of research is that of plan and goal recognition, in the field of artificial intelligence. Here too, the aim is to develop a system which is able to correctly understand the goal underlying an observed sequence of actions. However, while in cognitive modeling the purpose is to approximate and explain the given task as humans perform it, in artificial intelligence the purpose is generally to create a system which performs the task in the best possible way (according to some specific performance criteria), though not necessarily in the same way humans do. While some of the algorithms developed in artificial intelligence are motivated by findings from cognitive sciences, this is not the general rule. We bring only a small sampling of the vast amount of work done in this area, highlighting how it differs from the work presented here.

First, we note that most recent plan recognition works focus on using probability distributions over possible explanations for an observed sequence of actions [Charniak and Goldman, 1993, Geib and Goldman, 2005]. Using consistency rules [Lesh and Etzioni, 1995, Hong, 2001] and learning [Blaylock and Allen, 2006], earlier goal recognition systems return a likelihood-ranked set of goals consistent with the observed sequence. We too propose using a probability distribution over possible goals, however, as we show, people utilize additional

information (aside from a-priori likelihood and consistency) in making their inference. Avrahami-Zilberbrand and Kaminka [2007] discuss additional ways, such as a bias towards hypotheses that signify threat.

More recently, an approach similar to ours has been suggested by Ramirez and Geffner [2010]. In both systems—ours and theirs—the actions are used in order to determine which of a predefined set of goals is the one intended by the actor. While their work goes down to the details of the computational implementation, we put emphasis on giving a cognitive justification of the system, and put it in the context of recent findings in psychology and neuroscience. This is true both of the set of goals (we propose that the notion of *affordances* plays a role here), as well as of the criteria of efficiency for choosing among the goals (we propose the Principle of Rational Action as a psychological justification of its use).

Another system which recognizes intentional actions has been implemented by Hongeng and Wyatt [2008] on a robot. Their work differs from ours in several respects. First and foremost, they emphasize the visual input analysis, which is not of interest to us in the scope of this work. Second, they aim to identify *action-goals*, such as grasp, reach, push, and not *state-goals*, i.e. desired end-states of the world, which is what we do. Towards the end of their work they point out that their system behaves in a way which fits the Principle of Rational Action. However, this principle is not explicitly part of their system, as it will be shown to be in ours. Last, since their work is in artificial intelligence and not in cognitive modeling, they do not compare the performance of their system to that of humans, as we do here.

2.3.3.2 Robot Imitation

Another field which has much in common with ours, is that of robot imitation. When dealing with robot imitation, two problems must be addressed: the problem of recognizing the goal to be imitated, in our words, ”intention prediction”, and the problem of executing the recognized goal using the robot’s physical configuration and its action repertoire—which are not necessarily the same as those of the actor’s. This second problem is known as the body correspondence problem. Only the first problem is relevant to our work.

Affordances are becoming more and more popular in the field of robotics.

The work of Lopes et al. [2007] applies them to imitation. They show how a robot can learn a task, or a policy, after observing repeated demonstrations by a human. As defined above, the term *intention* in our context does not include tasks, or sequences of actions, but rather end-states. More importantly, our experiments show that in our model, observation of one demonstration is enough for predicting intention. Another difference is that the repeated demonstrations familiarize the robot with the objects, thus allowing it to learn the relevant affordances. We suggest that affordances can be extracted from *novel* objects, based on findings from the psychological literature, as will be explained when we introduce affordances in detail (Section 2.5).

Many other robot imitation systems have been developed (for a review, see Breazeal and Scassellati [2002]). Some of them concentrate on the problem of robot body correspondence mentioned above [Schaal et al., 2003]. Others deal with imitation of the kind related to a different definition of intention than that dealt with in this work, such as movement per se [Schaal et al., 2003, Billard and Mataric, 2001], gestures [Calinon and Billard, 2007] and emotion [Breazeal, 2003].

2.3.4 Cognitive Modeling

Putting the hardware aside, cognitive modeling of intention recognition aims at uncovering the core cognitive abilities required for the task, and the way in which they interact to produce the desired effect. To this end, empirical psychological data is made use of, as well as many computational tools developed in recent years.

For example, research in psychology attempts to pinpoint the age at which intention understanding matures. By correctly placing it within the context of other developing skills—be they social, motoric or cognitive—speculations can be explored regarding the various relationships between the different skills. One such study has shown that understanding failed reaching actions is present at 10 months of age [Brandone and Wellman, 2009], and is preceded by the understanding of successful reaching actions. In addition, development of the understanding of failed actions has been shown to occur at the same time as initiation of joint attention and the ability to locomote independently [Brandone, 2010].

Identifying the relationship between various skills, i.e. knowing what skills

are required for understanding intentions, and what skills make use of understanding intentions, enables correctly identifying and implementing the building blocks of artificial cognitive systems with intention understanding abilities. Nehaniv and Dautenhahn [2007], Meltzoff and Decety [2003] and Meltzoff et al. [1999] are examples of this approach.

Another example of utilizing psychological theories for cognitive modeling is that of [Oztop and Kawato, 2005], who have implemented systems based on the simulation theory account of theory of mind. In this work, we too make use of a psychological principle—that of Rational Action (Section 2.2.2)—and show how it can be translated into a computable form.

Computational tools are also made use of in the field of cognitive modeling. Cuijpers et al. [2006] is one of many who have used neural networks to solve problems related to goal recognition. Meltzoff and colleagues [Rao et al., 2007] have employed Bayesian learning to implement a system based on his four-stage paradigm of imitation (see above, Section 2.3.1.1). Inspiration from mirror neurons (see above, Section 2.3.2) has been drawn by several researchers to the end of creating artificial systems which exhibit imitative behavior. For a review, see Oztop et al. [2006].

We focus here on two challenges posed by Meltzoff [1995]'s experiments: How is intention prediction possible when only a failed sequence of actions is demonstrated? And how is intention prediction possible when the actions are performed on novel objects, about which the observer seemingly has no prior knowledge? These two challenges have not yet been satisfactorily addressed in the cognitive modeling research community, and that is what we attempt to do in this work. Our aim is to model the phenomenon of intention recognition, in a way that best fits the data accumulated in the various fields.

2.4 Meltzoff's Experiment

In order to understand the motivation for our model, as well as the setup used to evaluate it, we elaborate here on a description of Meltzoff [1995]'s experiment. The purpose of his experiment was to test whether children of 18 months of age are able to understand the underlying intention of a sequence of actions, even when that intention is not realized, i.e. when the acting agent failed to achieve

the goal. Since children of such young an age are not verbally proficient, he used a re-enactment procedure which builds upon the tendency of toddlers to imitate adults.

For each of five different novel toy objects, a target action was chosen. For example, for a two-piece dumbbell-shaped toy, the target action was pulling it apart. For a loop and prong device, the target action was to fit the loop onto the prong. The children were divided into four groups: Demonstration Target, Demonstration Intention, Control Baseline and Control Manipulation. Each child was seated in front of an adult with a table between them, on which lay one of the five objects, and was exposed to a demonstration, depending on the experimental group to which he or she belonged:

- The children in the Demonstration Target group were shown three repetitions of a successfully completed act, such as pulling apart the dumbbell, or hanging the loop on the prong; their voluntary response was to reproduce the same act when the objects were handed to them.

- The children in the Demonstration Intention group were shown three *failed attempts* of the adult to produce the goal, where the adult (seemingly) failed at reaching it, and they never saw the actual goal. *These children's re-enactment of the goal reached a level comparable to that of the children who saw the successful attempts*. This shows that children can see through the actions to the underlying intention, and extrapolate the goal from the failing actions.

- The children in the Control Manipulation group saw the object manipulated three times in ways that were not an attempt to reach the chosen target act. This was done in order to make sure that mere manipulation of the object is not enough for the children to reproduce the goal.

- A second control group—Control Baseline—had the children just see the object, without it being manipulated at all, in order to test whether they would reproduce the goal on their own. Both control groups did not show significant success at reproducing the target act.

When do children choose to act in a way that imitates the adult, and when do they choose to remain passive and not act? The experiment of Meltzoff [1995]

shows that when children discern an underlying intention, as in the two Demonstration groups, they attempt to imitate it. When they do not detect such an intention, as in the Control groups, they do nothing, or sometimes mimic the arbitrary acts of the adult (in the Control Manipulation group; obviously, children were imitating *what they understood to be* the intention of the adult). Only when no intention was apparent from the actions of the adult did the children remain passive and not produce any action.

Thus a complete model of intention recognition must first be able to model the ability to discern whether or not there is an underlying intention. Only then is it relevant to attempt to infer what that intention is. Allowing for such a preliminary stage would explain why children in both Demonstration groups were motivated to look for an underlying intention, while children in the Control Baseline group were not. This also explains why children in the Control Manipulation group sometimes reproduced the actions of the adult, even when it was not exactly what the experimenter had in mind. We propose that what characterizes those demonstrations which the children chose to re-enact is a rationality, or efficiency, which hints at an underlying intention worth imitating. We will make this notion more concrete, and show how to make use of it in order to computationally detect intention.

Meltzoff's original ground-breaking experiment intrigued other researchers, and served as a basis for many follow-up experiments, exploring various aspects of the understanding of intentions in order to hone in on the exact mechanism enabling this ability. One such work is that of Huang et al. [2002]. They suggest several candidate "clues" which the infants might make use of in their attempt to identify the intention underlying the observed actions. One clue which they confirmed plays an important role is stimulus enhancement by spatial contiguity, i.e. the proximity of the object parts relevant to the realization of the intended goal. This clue will also be made use of in our model. For this, infants—and artificial agents with the same social abilities—must be able to decompose objects to their parts, and identify what actions can be performed with them. This is where the notion of *affordances* comes in, which will be taken up in the next section.

2.5 Affordances

An affordance is a quality of an object, or an environment, that allows an individual to perform an action. For example, ”sitting” is an affordance offered by a chair. In the present work, we claim that affordances play a role in the process of intention prediction. In order to lay the ground for the understanding of this role, the following section presents a short review of the topic, which refers only to those aspects which are relevant to the current work. For a more complete review see for example St. Amant [1999] and, more recently, Sahin et al. [2007].

The notion of affordances was first introduced by Gibson [1977], in the context of visual perception. Since his definition of affordances as ecological properties of the environment which depend on the perceiver, the concept has evolved into various forms and uses in many different fields of study. The ecological, perceptual and cognitive psychology literature all deal with affordances, as does research in several computer science and engineering fields, from object-oriented programming languages through human-computer interaction and artificial intelligence to robotics and industrial engineering. The term ”affordance” is often used loosely, and the different contexts in which it appears possibly refer to different meanings of it. Therefore, any work which makes use of the notion of affordances should begin with a clarification of what exactly is meant by the term. In the following we do this, while putting the notion into the context of intention prediction.

2.5.1 Action-State Duality of Affordances

One level of abstraction of the notion of affordances, which follows naturally from the original definition, tends to blur the distinction between affordances and actions. On this level, every affordance is an action. See for example Gaver [1991], who defines affordances as ”potentials for action”. The same is true of Cisek [2007], who straightforwardly refers to potential actions as affordances. Neurophysiological data supports this abstraction. Using fMRI, Grezes et al. [2003] have shown that viewing an object potentiates brain activity in motor areas corresponding to the actions that the object affords.

The action-state duality familiar in the Artificial Intelligence planning community, suggests viewing affordances from the point of view of states, rather than

actions. Since every sequence of actions has a sequence of states induced from it, and vice versa, every executed sequence of states has a sequence of actions which induced it, we propose here to view affordances not as possible actions which can be performed on the environment, but as possible states which the environment can be brought to. This duality allows us to refer to possible goal states as affordances. In other words, when attempting to recognize the intention underlying a sequence of actions, we can consider the affordances available in the environment, in the form of possible goal states. Although this is not a common view in the affordance literature, we exploit this duality and suggest that findings regarding affordances as actions are valid regarding affordances as states.

2.5.2 Affordances as Interactions and Relationships Between Objects

While the framework described here is applicable to affordances in general, when dealing with the prediction of intentions, our experiments deal with a specific subset of affordances, namely, those which can be described as interactions and relationships between objects in the environment. This subset has been dealt with in the context of object-oriented programming [Baldoni et al., 2006], and fits in well with our view of affordances as states: two objects can define different states, depending on the relationship they hold with each other. Several examples studied by developmental psychologists are ”passing-through” and ”support” [Sitskoorn and Smitsman, 1995], ”containment” [Carona et al., 1988, Chiarello et al., 2003], ”above” and ”below” [Quinn, 1994] and ”tight-fit” [Casasola and Cohen, 2002].

2.5.3 Development of an Affordance Library

Regarding how and when the ability to recognize affordances is acquired, much research has been done in the field of developmental psychology. The works quoted above [Sitskoorn and Smitsman, 1995, Carona et al., 1988, Chiarello et al., 2003, Quinn, 1994, Casasola and Cohen, 2002] attempt to determine the age at which various spatial relationships are incorporated into the cognition of the normally developing infant.

Learning functional categorization of objects based on object parts is also seen as acquisition of affordances, and has been extensively studied from a developmental perspective. Infants as young as ten months old, who have been familiarized with the same action performed on different objects, increase their attention when a familiar object is combined with a novel action [Horst et al., 2005]. By 14 to 18 months, infants who have been familiarized with two objects, each of which was combined with a certain action, dishabituate to novel combinations of the familiar objects and actions [Madole et al., 1993, Madole and Cohen, 1995]. These findings indicate that objects become associated with actions through experience. Infants aged 14 and 18 months can also attend to relations between function and the presence of certain object parts [Booth and Waxman, 2002], thus confirming that generalization can be made and applied to novel objects, based on familiar functional parts.

While there is ongoing debate as to the exact developmental time-line, all agree that throughout infancy and toddler-hood these and other concepts of functions and spatial relationships which objects afford are incorporated into the cognition of the developing child. We refer to this dynamically growing structure as an ”affordance library”. The existence of such a library enables humans to recognize possible actions which can be performed on objects—including novel ones—and possible states to which these objects can be brought about to, in relation to other objects in the environment.

2.5.4 Accessing the Affordance Library

Studies in experimental psychology support the claim that perception of an object serves as a prime which can potentiate or inhibit reaction time to commands to execute afforded actions on the object. Craighero et al. [1996] have shown how a prime visually congruent with an object to be grasped markedly reduces the reaction time for grasping. Tucker and Ellis [1998] employed a stimulus-response compatibility paradigm whose results were consistent with the view that seen objects automatically potentiate components of the actions they afford, even in absence of explicit intentions to act. This behavioral data shows that the perception of an object automatically potentiates motor components of possible actions toward that object, irrespective of the subject's intention. In terms of an affordance library, we interpret this as having the library accessed and the

relevant affordance extracted and made available upon perception of the object.

Neurophysiological experiments complement the above results. Fogassi et al. [2005] showed how mirror neurons encode goals (such as eating an apple or placing it in a cup). These neurons fire upon view of the grasping configuration of the actor's hand on the object, and so prove how the type of action alone, and not the kinematic force with which actors manipulated objects, determined neuron activity. Other research goes further, to state that even before an action is initiated, merely the observation of the object itself is enough to cause neuronal activity in specific motor regions. Among others, Grezes and Decety [2002] used positron emission tomography for exploring neural correlates of object perception. They found increased regional cerebral blood flow in areas known to serve motor representation. These activations are congruent with the idea of an involvement of motor representation already during the perception of an object and thus provide neurophysiological evidence that the perception of objects automatically affords actions that can be made toward them. Functional MRI was used by Grezes et al. [2003] to show increased activation in specific brain regions when the action subjects were asked to perform on an object clashed with the action the object afforded. Thus, results from both behavioral and neuroimaging studies confirm that affordances of an object become available to the observer upon the object's perception—even before action has been initiated on the object, and before the observer formulates an intention to do so or recognizes such an intention by a confederate. In other words, perception of the environment causes constant access to the affordance library—at every given moment, the perceiver has at hand possible affordances which are compatible with the current perception of the environment.

2.5.5 Probability Distribution Over Affordances

Having established that affordances are made available upon perception, we go on to claim that more than one affordance can be invoked by an object, and these multiple affordances have a probability distribution over them. In a hypotheses formulated and tested behaviorally and neurophysiologically, namely, the affordance competition hypothesis, Cisek [2007] sets forth a parallel mechanism by which biological agents choose actions. According to this hypothesis, at every given moment, when receiving input from the environment, an agent

is presented with several action possibilities, and must choose between them in order to act. Disregarding the action selection stage, we borrow from here the notion that upon observing the environment and the objects present in it, an agent is aware of several possible affordances competing between them. In the work of Cisek [2007] this competition is settled for the purpose of action selection, while in ours it is used for the purpose of intention prediction. Ye et al. [2009] have recently shown how the perception of one affordance can interfere with the possibility that another affordance will be detected for the same object. Based on their findings, we conclude that several different affordances can be invoked simultaneously with different likelihoods.

2.5.6 Applications of Affordances in Engineering

Recently, AI experts and roboticists have turned to affordances, understanding their potential for enriching an agent's interaction with its environment. Affordances can be applied to software and hardware agents in two ways. The first is concerned with developing the ability to automatically learn affordances. There has been extensive research in this area [Stoytchev, 2005, Erdemir et al., 2008, Fitzpatrick et al., 2003, Hart, 2009, Ridge et al., 2009, Dogar et al., 2007]. The related field of object categorization [Pinz, 2005] has also been explored, in particular, functional object categorization [Rivlin et al., 1994], which builds upon Biederman [1987]'s Recognition By Components theory. The present work does not attempt to deal with this application of affordances, rather, we shall assume it is well developed enough in order to be incorporated into a model such as the one suggested in our work, as will be shown.

Once the ability of affordance learning and recognition is incorporated into the artificial agent, it can be applied in a second way: it can be used to enhance the cognitive repertoire of the agent. This possibility has already been pointed out by Murphy [1999], and has recently gained more popularity, e.g. in the work of Dogar et al. [2007]. In particular, affordances can be applied to intention prediction, which is what we propose here. To the best of our knowledge, there has not been much work in this area so far. One notable exception is the work of Lopes et al. [2007], which similarly suggests using affordances for robot imitation. Their work differs from ours in several aspects, mentioned above (Section 2.3.3.2).

Chapter 3

A Model of Intention Recognition in Humans

In the introduction, we presented an outline of a proposed cognitive model of human intention recognition. We are now ready to go into the details of the model. Several of the modules, as important as they are to the complete process, deal with aspects which are not directly related to intention, such as gaze detection and affordances. Therefore, aside from placing them in context and describing their contribution to the model, we do not further analyze their possible underlying mechanisms. Rather, we rely on others' work in the relevant fields for this. The two main modules which concern us are those of intention detection and intention prediction.

The next sections provide a sequential description of the process of intention recognition, beginning with the input of the perceived environment and the acting agent, along with the trace of observed actions (this input is accessible to all modules), and ending with the output of the most likely goal intended by the acting agent. We bring here again a graphic description of the model, in Figure 3.1. The input and output of each module will be stated, with the output of each module contributing to the input of a subsequent module, thus forming the flow of the model. Details of each module will be given, with the reservation mentioned above: elaboration will be provided only for those two modules which are at the core of the process, as we understand it. For these two modules we will describe their underlying mechanisms and the methods used to evaluate them.

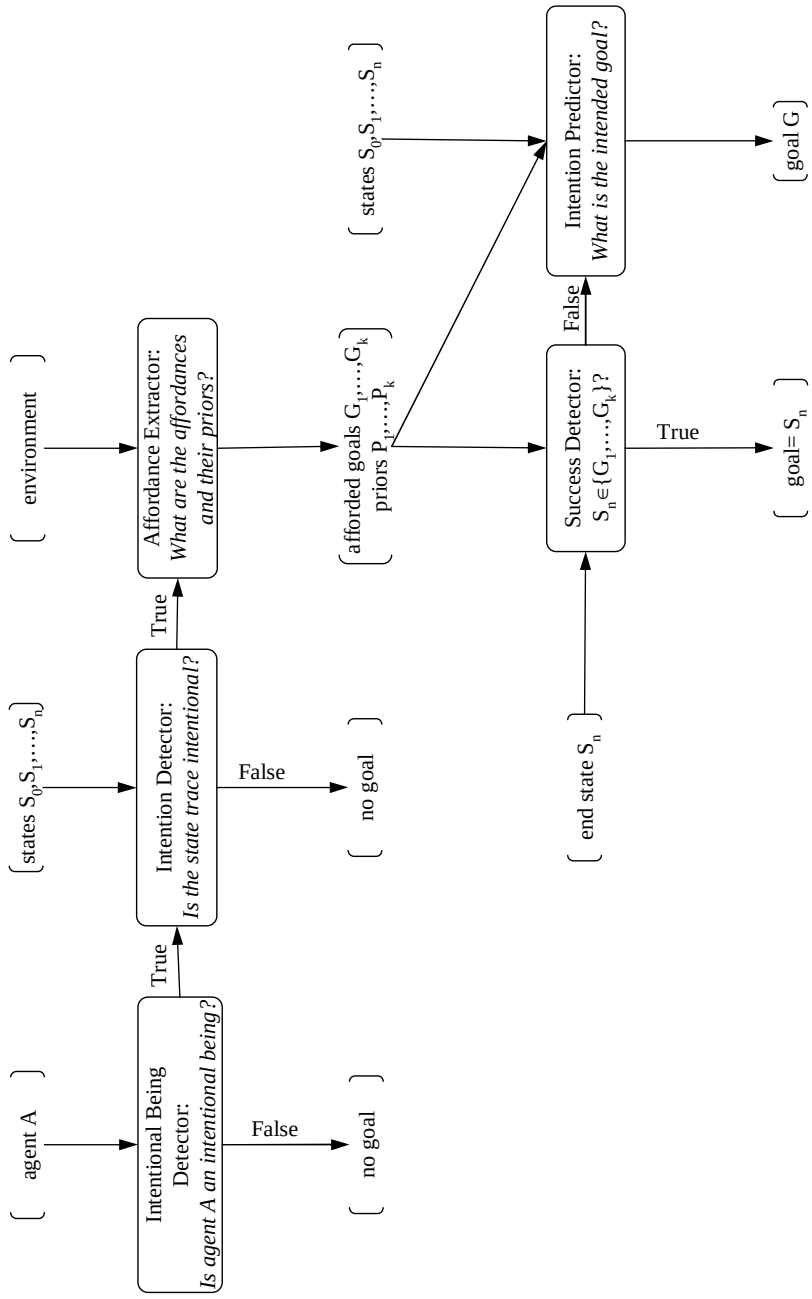


Figure 3.1: Scheme of Proposed Model.

3.1 Intentional Being Detection

The input relevant for this module is the perception of the acting agent. The module answers the question: Is the observed agent an intentional being? The output is a binary answer: `True` if the agent is deemed capable of intention, and `False` otherwise.

As noted above (Section 2.2.2), detecting whether or not the observed acting agent is capable of intention is a necessary preliminary stage. This is true both conceptually and practically. Conceptually, since apparently this is how humans process movement performed by other agents: they first determine whether or not an underlying intention should be searched for. And practically, since it would be futile and misleading to attempt to decipher a sequence of actions regarding its underlying intention, when it was performed by an agent not capable of intention.

Evidence for the fact that humans do indeed first determine the capability for intention of the observed acting agent arises from several experiments in developmental psychology. In his original re-enactment procedure, Meltzoff [1995] himself showed that when a mechanical arm performed the same actions as the human, the observing children did not attempt to reproduce the failed goal. A set of studies by Woodward and colleagues [Woodward et al., 2001] indicate that agents lacking certain specific human-like characteristics do not induce imitative behavior in children observing them. Hofer et al. [2005] have shown that while 12-month-old infants relate to a mechanical claw as possessing intentions, 9-month-old infants do not do so unless they are first shown that a human hand is activating the claw. All this goes to show that in order to be able to attribute intentions to an acting agent, humans must first possess an understanding regarding the ability of that agent to act intentionally.

As to the exact characteristics of the agent which invoke that understanding in the observer, there are various candidates. Among them are texture (e.g. metal versus skin [Woodward et al., 2001], and eye gaze [Itakura et al., 2008]. The latter has proved to be a determining factor in the attribution of intentions by children to an acting robot: only when the robot made eye contact with the children before acting, did the children interpret its actions as intentional.

Since extracting human-like characteristics from the observed agent is not directly related to the subject of intention, we do not go into implementation

details of this module. Rather, we assume it to be given. If the output of the module is `False` , then the process stops here. Only if the output is `True` , will the process continue and go on to the next module: once the acting agent has been determined to be of the intentional kind—such as a human or a cognitive robot—the next stage is to analyze the particular instance of executed action for the presence of intention.

3.2 Intention Detection

This module consists of the first of the two main processes we identify in the problem of intention recognition. The question it answers is whether the observed sequence of actions was performed intentionally or not. Again, the answer is binary: `True` if the action sequence is deemed intentional, `False` otherwise.

How can this question be answered? As described above (Section 2.2.2), a necessary condition of intentional action is that it is efficient, in the sense of the Principle of Rational Action [Gergely and Csibra, 2003, Watson, 2005]. We propose that this condition is also sufficient. In order to establish this, we make the notion of efficiency more concrete, so that it can be translated into a computable form. To this end, we introduce a *measure of intention*, described and formalized next.

We denote the observation trace by $t = s_0, ..., s_n$, i.e. a sequence of states, brought about by the actions of the demonstrating agent. s_0 is the initial state, and s_n is the terminal state. The task of the observing agent is to decide, given this trace, whether there was an underlying intention or whether the acting agent behaved unintentionally.

Inspired by the Principle of Rational Action [Gergely and Csibra, 2003], we check for some form of efficiency in the trace. It is reasonable to expect that a trace with an underlying intention will exhibit a clear progression from the initial state towards the goal state, which is the most efficient way to bring about that goal, given the initial state. Note that our agent does not know at this stage whether or not there is an underlying intention to the trace, and even if there is, if it is reached successfully. On the other hand, unintentional traces would not be driven by such efficiency, and would fluctuate towards and away from the initial state, without any clear directionality.

We define a state-distance measure *dist*, which measures the shortest path between two states of the world, given all possible actions that can take us from one state to the other. There are a few requirements for the distance measure. We do not require this distance to obey symmetry ($d(s_1, s_2) = d(s_2, s_1)$). However, it should always be positive, and equal 0 only from a state to itself. In addition, it should obey the triangle inequality. Using any such distance measure, we capture the notion of optimality, in the sense of a shortest path from one state to another.

Thus, from the original state trace we induce a sequence of distance measurements $d_1 = dist(s_1, s_0), ..., d_n = dist(s_n, s_0)$, measuring the *optimal (minimal) distance* between each state in the sequence, and the initial state. In this way, for every state we have an indication of how much the demonstrating agent would have had to invest (in time, number of atomic actions, or any other resource, depending on how the distance is defined), had it been intending to reach that state. We posit that enough information is preserved in this sequence for our observing agent to come to a satisfying decision regarding the presence of an underlying intention.

This distance measure is dependent on the nature of the world being modeled. For example, when dealing with spatial targets, the distance could simply be the Euclidean, and indeed it is, in the second of our experiments. In a discrete state-space, defined by STRIPS notation, we use Bonet and Geffner [1999]'s Heuristic Search Planner (HSP) to generate optimal plans from the initial state to every state in the trace, and the number of action steps in each generated plan is taken to be the distance to the respective state. This distance served us in our first experiment. Other distance measures might incorporate information regarding cost expended or effort invested in the actions.

The behavior of the sequence of distances conveys how efficiently the demonstrating agent performed is actions. If it acted efficiently—taking only optimal action steps that bring it closer to the goal—then the sequence of distances will be monotonically increasing, since every state reached will be more distant from the initial state than the state at which the agent was at one time step before. While if the agent acted randomly, executing various actions that do not necessarily lead anywhere, then the sequence of distances will fluctuate, and will not display any clear progression away from the initial state.

We want to quantify this intuitive reasoning and calculate from the distance

sequence a measure of intention. A naive approach would be to check the monotonicity of the sequence—if the distances of each state from the initial state increase monotonically, then we have a very strong indication of efficiency, which conforms to the rationality principle, and therefore, we can strongly conclude the presence of intention. However, expecting the sequence to strictly increase, or even merely non-decrease, at every point, is too strong a requirement, and would not stand up to the flexibility inherent in real-life motion. Very rarely will human motion display complete monotonicity of this distance sequence, no matter how intentional the actions from which it was induced. We therefore use a different, softer, approach: for every state, we check if the distance from it to the initial state is greater than that of the previous state. We call this a local increase, and we take the proportion of local increases in the sequence to be our intention measure. That is, we look to see at how many of the states along the trace has the distance from the initial state increased, as compared to the previous state, out of the total number of states in the trace. This will give us an idea of how efficient the action sequence is. Of course, if the sequence does happen to be completely monotonic, then a local increase will be found at every point, and so the proportion will equal 1. Yet, for the less-than-perfect sequences, there will still be ample margin to convey their intentionality.

More formally,

$$u = |\{s_i : d_i > d_{i-1}\}_{i=1}^{n}| \tag{3.1}$$

is the number of states in the trace where the distance from the initial state increases, as compared to the distance at the previous state. Taking this number and dividing it by the total number of states in the trace,

$$t = \frac{u}{n} \tag{3.2}$$

gives us a measure of intention for the action sequence.

The higher the resulting t, the more intention is attributed to the action. If a binary answer is preferred, we can determine a cutoff level which serves as a threshold above which we conclude intention is present, and below which we conclude it is not.

For example, in the case of clear intention, we would expect a strictly monotonically increasing sequence of distances; the agent proceeds from the initial state, at each step moving farther and farther away from it, and closer and closer

to the intended goal. At the other end, if the observed agent is not driven by an intention to reach any particular state, we would expect the sequence to fluctuate in a seemingly random fashion, with the agent sometimes moving away from the initial state and sometimes moving back towards it. Of course, this is merely a motivational argument. In Chapters 4 and 5, describing the experiments, we show that this simple intuitive method does indeed produce the expected results.

3.3 Affordance Extraction

Only if in previous stage the presence of intention has been established (indicated by an output of `True`) will the process continue on to the task of determining the actual content of the intention. To this end, we propose employing a variation on the theme of affordances, as described in Section 2.5. The environment and the objects in it can be analyzed and their affordances extracted, and these affordances will play a part in the next stages of the process.

Follow-up studies using Meltzoff's [Meltzoff, 1995] original re-enactment paradigm have shown that the ability to imitate unsuccessful goals is existent at 18 months of age, but not at 12 months [Bellagamba and Tomasello, 1999]. However, recent developments seem to indicate that what differentiates the children in these two age groups is not their intention-reading ability per se, but rather their ability to limit the range of possible outcomes to a small set of goals. Limiting the range of possible outcomes is crucial, since this is what makes the behavior transparent to its goal [Csibra and Gergely, 2006]. Nielsen [2009] has shown that once 12-month-old children become acquainted with the affordances of the objects and their parts, they are then able to deduce the intended goal of the actor manipulating the objects. Yet, when the affordances are not made explicit to the children (as was the case in Bellagamba and Tomasello [1999]'s experiment), they are unable to interpret the intentions of the actor. This is strong evidence to the fact that the ability to extract affordances from objects, based on prior knowledge, is a prerequisite to the ability to read the intentions underlying actions performed on those objects. For this reason we incorporate the Affordance Extractor sub-module into our model.

The module of affordance extraction takes as input only the environment and the objects in it—it does not make use of the observed action sequence. As such,

it could theoretically be executed independently of the previous modules. Nevertheless, we place it within the model at this point, since it would be inefficient to extract affordances before the presence in intention has been ascertained. The output of this module is a set of k affordances, $\{g_j\}_{j=1}^{k}$, in the sense of states which could serve as the intended goal of the acting agent.

The subject of affordances is tangent to the subject of this research, however, it is not directly related to intention. For this reason, we do not propose a model for extracting affordances from objects. A large body of research has accumulated, both theoretical and practical—as described in Section 2.5—which can facilitate the implementation of such a module and its incorporation into the proposed cognitive model of intention recognition. We leave the details to others, and focus on how such an implementation can be made use of in our context—in the next section.

3.4 Success Detection

Given as input an action sequence already determined to be intentional by the Intention Detector module, and a list of affordances from the Affordance Extractor module, the question of whether or not the actor succeeded in achieving its goal can now be answered. This answer can be given in a very straightforward manner, after the previous stages have been completed. Formally, the question and answer can be described as $s_n \in \{g_1, ...g_k\}$, where s_n is the observed terminal state and $\{g_j\}_{j=1}^{k}$ are the affordances extracted from the objects in the environment.

Simply, if the terminal state which the acting agent has brought about by its actions is one of the affordances, we assume that it is the intended goal at which the actions were aimed, and that the agent has successfully achieved it. If, on the other hand, the terminal state is not one of the affordances, we assume the agent failed at realizing its intention. This follows from our premise that the intended goal coincides with one of the extracted affordances.

This stage of Success Detection is of conceptual importance more than practical. If answering the question of whether or not the acting agent was successful is not of interest, then it can be ignored. The implementation itself consists of nothing more than a simple logical test, and moving on to the next stage without

explicitly answering it will cause no loss of efficiency, as will be seen. Yet we state it as a module in itself in order to show how the previous stages lay the groundwork for solving what might otherwise have been a difficult question.

3.5 Intention Prediction

This module is the second of the two main focuses of this dissertation. It is concerned with predicting the intention underlying the observed stream of actions. Its input is the trace of actions along with the list of possible afforded goals, as returned from the Affordance Extractor. Its output is the final output of the whole process of intention recognition, namely, a goal $g \in \{g_j\}_{j=1}^{k}$, which is most likely the intended goal, given the observations.

We next present a formalization of the problem at hand, followed by three possible heuristics which can be employed for this task. Each of these heuristics is based on different information extracted from the action sequence or from the objects. In the next chapter, we describe a two-phased experiment on human subjects, whose results validate the use of the notion of affordances and the ability to correctly choose among them using the presented heuristics.

Based on the findings from the affordance literature (quoted above in Section 2.5), and on our experiments (described below in Chapter 5.2.1), we posit that observation of the objects invokes possible goal states, along with a distribution over them. Recall the notation $g_1, ..., g_k$ for k possible afforded goals, and $p_1, ..., p_k$ for their respective likelihoods, with $p_1 + ... + p_k = 1$. These g_i are the goal-states considered as possible intentions underlying the observed actions.

For the case of s_n coinciding with one of the goals g_i, it would make sense to conclude that the sequence of actions was successful in achieving this goal. If s_n is not one of these goals, we conclude failure, and seek a way of choosing which g_i is the intended goal. This in essence, is the content of the Success Detector module, which, as explained above, can be ignored without loss of functionality or efficiency. It is actually built into the Intention Predictor, and only conceptually distinct from it.

We propose three different heuristics, which build upon each other, for intention prediction. We will show how these heuristics play a role in the way humans determine which goal is the one most likely intended by the acting agent. The

first heuristic takes into account only the objects in the environment, disregarding the observed actions and their effect on the objects. It is defined by the prior probability distribution p_i. Acting according to this heuristic alone would produce the choice of that g_i with the highest p_i.

The second heuristic considers further information, namely, that of the state of the environment brought about by the actions, s_n. The distance function, $dist(s_i, s_j)$, between states, is utilized here. This distance need not necessarily satisfy all the usual requirements of distance functions (such as symmetry), however it must always be positive and equal zero if and only if $s_i = s_j$, and *it is always optimal*.

The distance measure could be the same one utilized in the Intention Detection module (above, Section 3.2). Given this distance function, we compute k values, $d_i = dist(s_n, g_i)$, for each of the k possible goals, g_i. Our second proposed heuristic utilizes this distance sequence, d_i. A reasonable way of acting according to it would be to choose that g_i with the lowest d_i, i.e. the goal closest to the terminal state arrived at. This can be seen as a realization of the mechanism of stimulus enhancement by spatial contiguity, mentioned as one of the clues for predicting intention in Section 2.3.

The third heuristic is motivated by the psychological Principle of Rational Action [Gergely and Csibra, 2003], which states that intentional action functions to bring about future goal-states by the most rational actions available to the actor within the constraints of the situation. In essence, it means that the action should display some form of efficiency. Consider g to be the intended goal, then the sequence of states beginning with $s_0, ..., s_n$ and continuing directly to g should exhibit efficiency. Making use of the complete trace of action available to the perceiver, $s_0, ..., s_n$, we define an intention measure which attributes a value to each of the potential goals, g_j. For each goal g_j, we measure the length of the path $s_0, ..., s_n, g_i$, and the length of the path going directly from s_0 to g, and divide the second by the first: $r_j = \frac{\sum_{i=1}^{n} dist(s_{i-1}, s_i) + dist(s_n, g_j)}{dist(s_0, g_j)}$ These lengths are calculated using the same distance function as above. The resulting ratio relays how long the actual path to g_j *would* be, compared to how long it *could optimally* be. These ratios, r_j, define our third heuristic of choosing the g_j with the highest intention, r_j.

Each of these heuristics could potentially serve to rank the afforded goals,

and choose the highest ranking one as that most likely intended by the acting agent. In the chapter describing the experiments for the Intention Prediction module (Chapter 5) we present an environment for evaluating them, and compare their effectiveness at the task of intention prediction to human performance.

Chapter 4

Experiments for Evaluating the Measure of Intention Detection

In this section we describe the experiments used to evaluate the proposed measure of intention detection. The problem formalization appears in Section 3.2, along with the proposed measure of intention for the task of intention detection. We now go on to describe two experimental setups in which this measure of intention was tested. The first environment is an artificial replication of Meltzoff's experiment, using standard AI planning problem description language (STRIPS). The second environment uses real life data from the online CAVIAR database of surveillance videos.

4.1 Experiment I: Discrete Version of Meltzoff's Experiment

The first environment in which we evaluated the proposed measure of intention consists of a discrete abstraction of Meltzoff [1995]'s experiments. The next section describes how we rendered Meltzoff's experiments into a computational form, using standard AI planning problem description language (STRIPS). This is followed by a results section which shows the performance of the model in this environment.

4.1.1 Experimental Setup

We modeled Meltzoff's experiment environment as an 8-by-8 grid, with several objects and several possible actions which the agent can execute with its hands, such as grasping and moving. We implemented two of the five object-manipulation experiments mentioned by Meltzoff: the dumbbell and the loop-and-prong. For the dumbbell, there is one object in the world, which consists of two separable parts. The dumbbell can be grasped by one or both hands, and can be pulled apart. For the loop-and-prong, there are two objects in the world, one stationary (the prong), and one that can be moved about (the loop). The loop can be grasped by the hand, and released on the prong or anywhere else on the grid. As previously described, we use Bonet and Geffner [1999]'s HSP* to compute the distance measure.

We manually created several traces for the dumbbell and for the loop and prong scenarios, according to the descriptions found in Meltzoff's experiment, to fit the four different experimental groups. In addition, we created a random trace, which does not exhibit any regularity. We added this trace since the children in Meltzoff's Control Manipulation group were sometimes shown a sequence with underlying intention, albeit not the target one. Since we want to test our model on traces that have no underlying intention whatsoever, we artificially created such a random trace.

For the dumbbell scenario, all traces start out with both hands at position (1,1), and the dumbbell is stationary at position (5,5). The traces are verbally described in Table 4.1. A graphic description is given as well for the first trace, in Figure 4.1. For the loop and prong scenario, there is only one active hand on the scene, which in all traces starts out at position (1,1). The loop starts out at position (3,3), and the prong is stationary at position (5,5). The traces for this pair of objects are described in Table 4.2. For each trace we calculated the sequence of distances, using the above mentioned HSP algorithm, and then computed the proportion t.

*HSP is downloadable from http://www.ldc.usb.ve/~bonet/

Trace Name	Trace Description
Demonstration Target	Left and right hands move from initial position towards the dumbbell, grasp it and pull it apart. A visual representation of this trace is given in Figure 4.1(a-n).
Demonstration Intention I	Left and right hands move from initial position to dumbbell, grasp it and pull, with left hand slipping off, leaving the dumbbell intact.
Demonstration Intention II	Same as above, with right hand instead of left slipping off.
Control Baseline	No movement—both hands remain static at initial position.
Control Manipulation	Left and right hands move from initial position to dumbbell, grasp it and remain static in that position for several steps.
Random	Right hand moves towards the dumbbell and grasps it, then releases it and moves away. Then left hand wanders around the grid, then right hand joins left.

Table 4.1: Description of traces for each of the experimental groups in the dumbbell experiment.

4.1.2 Results

Figure 4.2 shows plots of the sequences of distances associated with the dumbbell experiments. The step number in the sequence is depicted in the X axis. The Y axis measures the distance of the respective state from the initial state. Figure 4.3 shows the same for the prong and loop experiments. In Meltzoff's experiments, every child was shown three traces, and only then was handed the objects. There is certainly information in this seeming redundancy; see Meltzoff et al. [1999] who show that when only one trace was shown to the children in the Demonstration Intention group, they were unable to reproduce the goal. However, we do not incorporate the redundant information at this stage in our model (see the discussion in Section 6.2 for more on this). So, while every child was shown three possibly different traces, we calculated our measure of intention separately for each of these traces, which is why we have more than one row in

Trace Name	Trace Description
Demonstration Target	Hand moves from initial position to loop, grasps it and places it on prong.
Demonstration Intention I	Hand moves from initial position to loop, grasps it and places it to the right of the prong (in our interpretation, the loop "misses" the prong).
Demonstration Intention II	Hand moves from initial position to loop, grasps it and places it to the left of the prong.
Control Baseline	No movement—hand remains static at initial position.
Control Manipulation I	Hand moves from initial position to loop, grasps it and moves it along top of prong, from right to left.
Control Manipulation II	Hand moves from initial position to loop, grasps it and moves it along top of prong, from left to right.
Control Manipulation III	Hand moves from initial position to loop, grasps it and places it just below the prong.
Random	Hand moves from initial position to loop,grasps it and then releases, then moves away to wander about the grid.

Table 4.2: Description of traces for each of the experimental groups in the prong and loop experiment.

the table for some of the groups.

For example, the prong and loop procedure failed in two different ways in Meltzoff's Demonstration Intention condition—either with the loop being placed too far to the right of the prong (Demonstration Intention I in Table 4.4), or too far to the left (Demonstration Intention II in Table 4.4). The children in Meltzoff's Demonstration Intention experimental group each saw three demonstrations—first Demonstration Intention I, then Demonstration Intention II, and then once again Demonstration Intention I—while in our replication of the experiment, every such trace was a demonstration in itself.

Table 4.3 shows the calculated measure of intention for each of the traces in the dumbbell experiment, and Table 4.4 shows the same for the prong and loop experiment. In both tables, each row corresponds to a different type of state sequence. The right column shows the measure of intention as computed by the

method described above.

Trace	Measure of Intention
Demonstration Target	1
Demonstration Intention I	0.8333
Demonstration Intention II	0.9166
Random	0.5384
Control Manipulation	0.8333
Control Baseline	0

Table 4.3: Calculated Measure of Intention for STRIPS Implementation of Dumbbell Experiment.

Trace	Measure of Intention
Demonstration Target	1
Demonstration Intention I	1
Demonstration Intention II	1
Random	0.5555
Control Manipulation I	0.7777
Control Manipulation II	0.7777
Control Manipulation III	1
Control Baseline	0

Table 4.4: Calculated Measure of Intention for STRIPS Implementation of Prong and Loop Experiment.

Figure 4.2a shows the distance sequence for the Demonstration Target trace, for the thirteen-state trace graphically depicted in Figure 4.1. The graph is monotonically increasing, since at every state the demonstrating agent moved farther and farther away from the initial state, and closer to the goal state. Since at each of the twelve states following the initial state there was an increase in the distance, the intention measure calculated from this sequence is 12/12, i.e. 1, as seen in the first row of 4.3. This, of course, is the highest possible score, thereby clearly indicating intention, according to our interpretation.

The same can be seen for the Demonstration Target sequence of the loop and prong objects. Figure 4.3a shows the clear progression away from the initial state in a seven-state sequence. This too results in an intention measure of 7/7, i.e. 1, as seen in the first row of Table 4.4.

In the case of Demonstration Intention traces, we also get a high measure of intention. See for example the distance sequences of the Demonstration Inten-

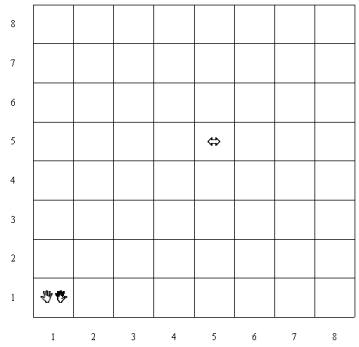
tion traces in the dumbbell experiment in Figures 4.2b and 4.2c. The distance increases along the traces, until the actor stumbles, so to speak, and takes steps that are unproductive in bringing him nearer to the goal that would realize his intention. This stumbling is expressed in the drop towards the end of the distance sequences. The corresponding measures of intention are therefore less than 1, yet still high enough to communicate the presence of intention. In the first Demonstration Intention trace we have the left hand slip off the dumbbell to the left, resulting in a state in which the hand is closer to where it previously was, with respect to the initial state. So there are nine out of eleven steps which increase the distance, resulting in the score shown in the second row of Table 4.3. In Demonstration Intention II, it is the right hand which slips off to the right, bringing it to a state which is yet farther away from the initial state. So there are ten out of eleven steps which increase the distance, as seen in the third row of the table. High measures of intention are also achieved for the two Demonstration Intention traces of the prong and loop experiment shown in Figures 4.3b and 4.3c. In fact, in this case the maximum score of 1 is reached (see the two corresponding rows in Table 4.4), even though the acting agent failed at reaching its goal. Although the agent "stumbled" here too, the stumbling happened in a way which resulted in a state which was farther away from the initial state than the previous state. We see here that our measure of intention is only useful for recognizing the presence of intention, but not for recognizing whether that intention was successfully fulfilled or not.

So far we have seen that action sequences with underlying intention, whether or not successfully realized, receive a high score of intention. What about action sequences which were performed as manipulation, and not aimed at achieving the target action? The case of the Control Baseline trace is simple—since no movement was executed whatsoever, the distance sequence remains a flat zero all along, as seen in Figure 4.2f for the dumbbell experiment and in Figure 4.3h for the loop and prong experiment. The resulting intention scores are therefore zero, as Tables 4.3 and 4.4 show.

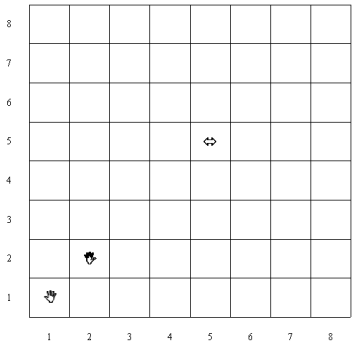
The Control Manipulation traces necessitate a deeper inspection. While our experiments show that the scores they achieved were generally lower than those for the intentional traces, these scores were nevertheless relatively high, and in one case (Control Manipulation III of the loop and prong experiment), maximal.

Indeed, the graph of this trace shows it is monotonically increasing. How can this be explained? Interestingly, Meltzoff's results showed that the children in the Control Manipulation conditions sometimes imitated the actions of the adult, bringing the objects to the same end-state as in the demonstration. This end-state was not the target action chosen for the experiment, yet, obviously, the children were detecting here some other intention worth imitating. So, although the demonstration was a manipulation with respect to the chosen target action, it was interpreted as intentional with respect to the perceived end-state by the children. This is more rigorously controlled and explored by Huang et al. [2002], with the same conclusion—that the children were detecting an underlying intention, even though it was not that which the experimenters had in mind.

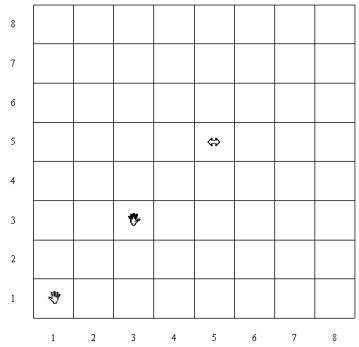
For this reason we designed what we called Random traces—traces with no underlying intention whatsoever, that have the agent move its hands about the state-space in an undirected manner. The distance graphs for these traces fluctuate, as seen in Figures 4.2d and 4.3d, which justly earn them the significantly lower scores appearing in the respective rows of Tables 4.3 and 4.4.



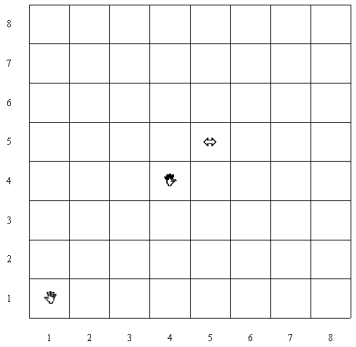
(a) Initial state. Both hands at (1,1), dumbbell at (5,5).



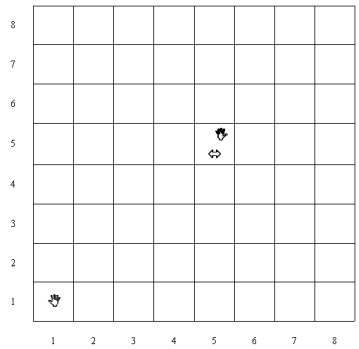
(b) Step one. Right hand moving towards dumbbell.



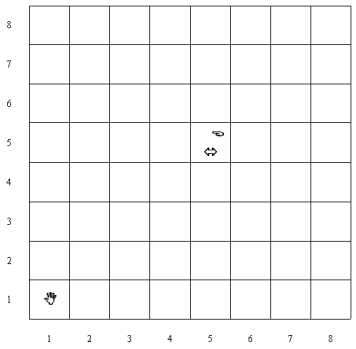
(c) Step two. Right hand continuing towards dumbbell.



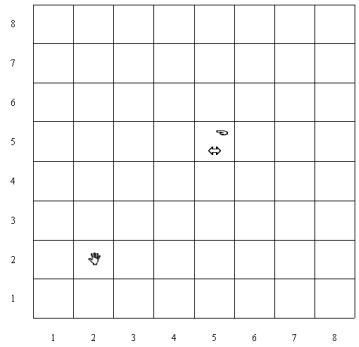
(d) Step three. Right hand continuing towards dumbbell.



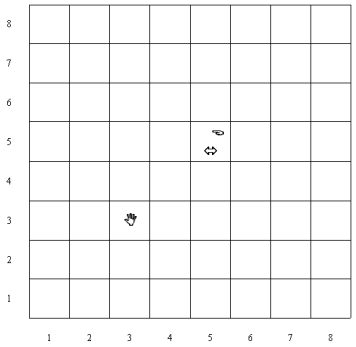
(e) Step four. Right hand at dumbbell.



(f) Step five. Right hand grasping.

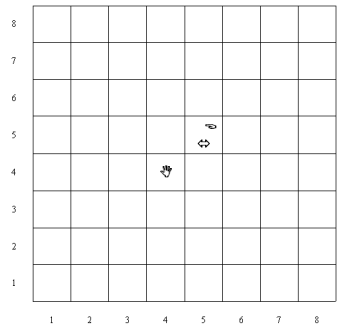


(g) Step six. Left hand moving towards dumbbell.

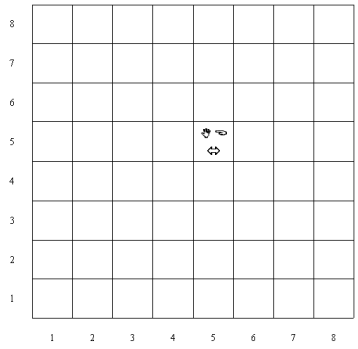


(h) Step seven. Left hand continuing towards dumbbell.

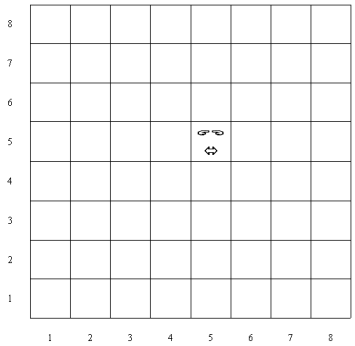
Figure 4.1: Dumbbell Demonstration Target Trace.



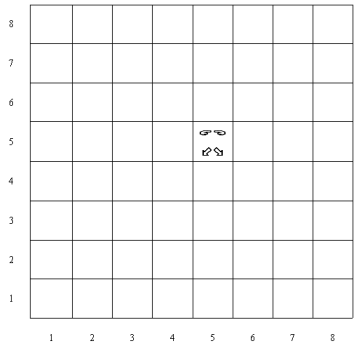
(i) Step eight. Left hand continuing towards dumbbell.



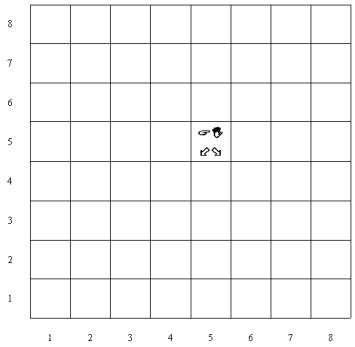
(j) Step nine. Left hand at dumbbell.



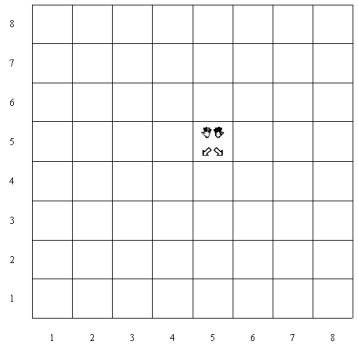
(k) Step ten. Left hand grasping dumbbell.



(l) Step eleven. Pulling apart.

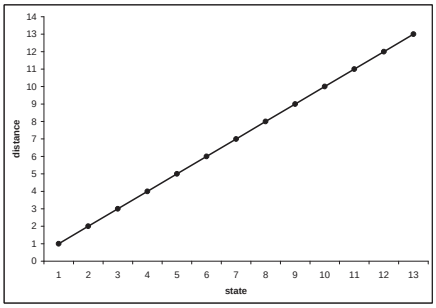


(m) Step twelve. Releasing one hand.

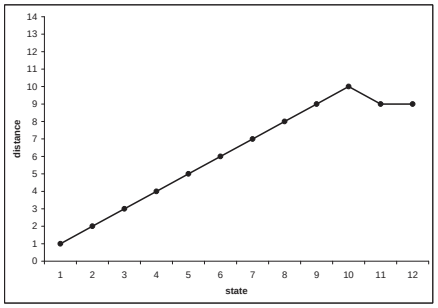


(n) Step thirteen. Releasing other hand.

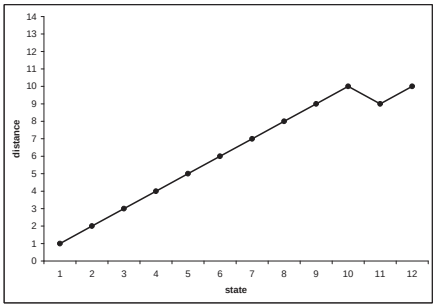
Figure 4.1: Dumbbell Demonstration Target Trace (cont).



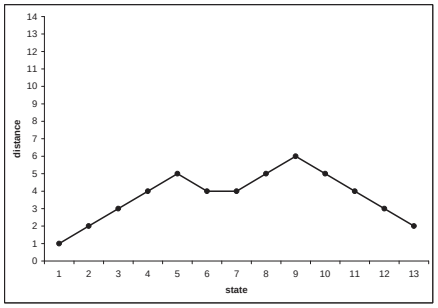
(a) Demonstration Target



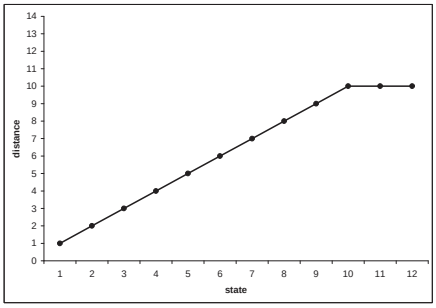
(b) Demonstration Intention I



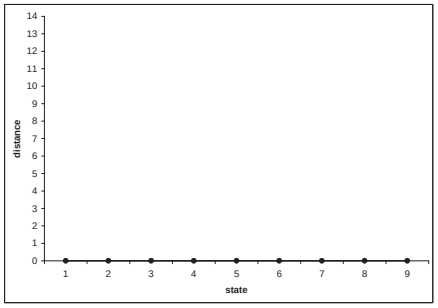
(c) Demonstration Intention II



(d) Random

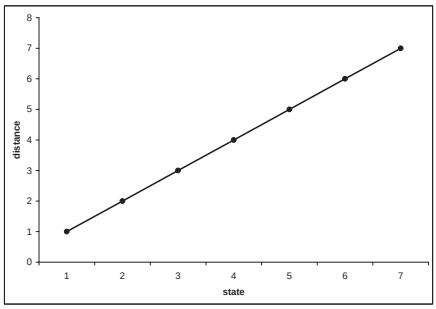


(e) Control Manipulation

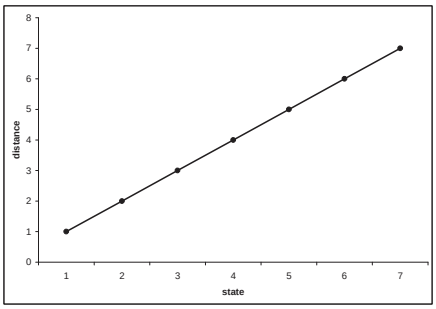


(f) Control Baseline

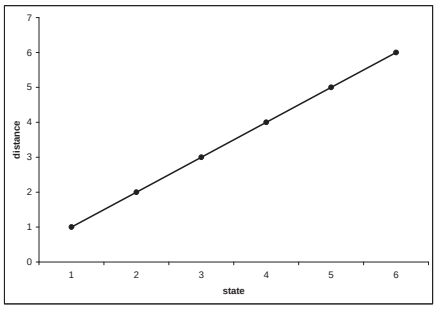
Figure 4.2: Distance as a Function of State in the Dumbbell Experiments.



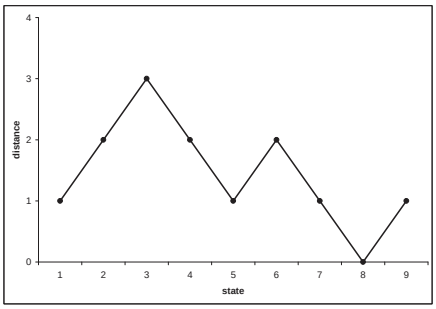
(a) Demonstration Target



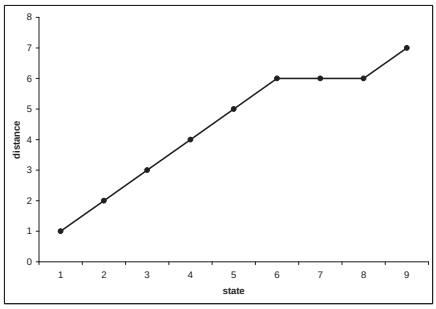
(b) Demonstration Intention I



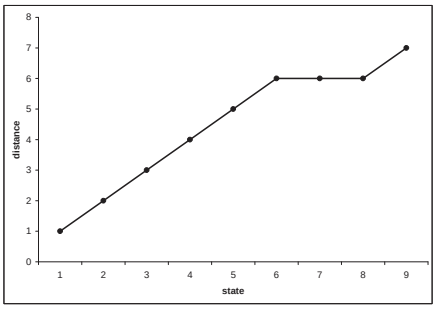
(c) Demonstration Intention II



(d) Random

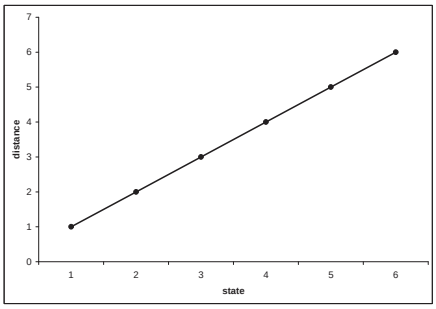


(e) Control Manipulation I

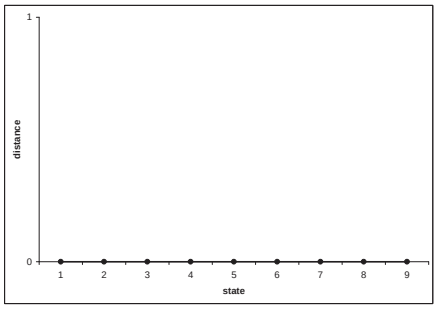


(f) Control Manipulation II

Figure 4.3: Distance as a Function of State in the Prong and Loop Experiments.



(g) Control Manipulation III



(h) Control Baseline

Figure 4.3: Distance as a Function of State in the Prong and Loop Experiments (cont).

4.2 Experiment II: Surveillance Videos

The second environment in which we evaluated the utility of the proposed measure of intention for detecting the presence of intention, uses surveillance videos. These were taken from the CAVIAR database at the University of Edinburgh's School of Informatics[†]. The next section describes the environment, followed by a description of the results, comparing the intention of the observed data according to the proposed measure of intention and according to human judgment. In addition, we inspect the possibility of using the measure of intention for segmenting subgoals.

4.2.1 Experimental Setup

A second set of experiments was carried out in a new domain, in order to compare the judgment of intention according to the proposed measure, to that of human observers. We are interested in how human observers classify real-life human movement, and whether their judgment of intention correlates with that of our model. To test this, we used the CAVIAR video repository of surveillance videos.

4.2.1.1 The Data

The CAVIAR project contains video clips taken with a wide angle camera lens in the entrance lobby of the INRIA Labs at Grenoble, France. In the videos, people are seen walking about and interacting with each other. A typical screen shot from one such video is shown in Figure 4.4. Each video comes with an XML file of the ground truth coordinates of movement for the people seen in the video. We selected a dozen of these movies, and cut from them clips in which single people are seen moving about. Table 4.5 enumerates the clips and the videos in the repository from which they were taken. Some videos had more than one clip extracted from them, in which different characters moved about. In the XML files, these characters are distinguished by unique numbers, named Object IDs. These clips were shown to human subjects, while the ground truth coordinates of the character's movement were extracted from the XML files and fed as input

[†]The EC Funded CAVIAR project/IST 2001 37540, found at URL: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/.

for calculating the intention measure. Clip number 5 was given as an example to the subjects, and therefore does not appear in further analysis.

Figure 4.4: Typical Screen Shot from a CAVIAR Video, with Character Seen Entering From Bottom.

Clip Number	File Name	XML File Name	Object ID
1	Walk1.mpg	wk1gt.xml	1
2	Walk2.mpg	wk2gt.xml	4
3	Walk3.mpg	wk3gt.xml	4
4	Walk3.mpg	wk3gt.xml	2
5	Walk3.mpg	wk3gt.xml	3
6	Browse1.mpg	br1gt.xml	3
7	Browse2.mpg	br2gt.xml	3
8	Browse3.mpg	br3gt.xml	1
9	Browse4.mpg	br4gt.xml	1
10	Browse4.mpg	br4gt.xml	2
11	Browse_WhileWaiting1.mpg	bww1gt.xml	2
12	Browse_WhileWaiting2.mpg	bww2gt.xml	0

Table 4.5: Clip numbers with their corresponding video file name, xml file name and object ID in the CAVIAR repository, from which they were taken.

With respect to intention, the clips we chose show movement ranging from very deliberate (e.g. a person crossing a lobby towards an exit), to not very clear (e.g. a person walking to a paper stand and browsing, then moving leisurely to a different location, etc.). We compared human subjects' judgment of the intention of motions in these videos, to the predictions of our model.

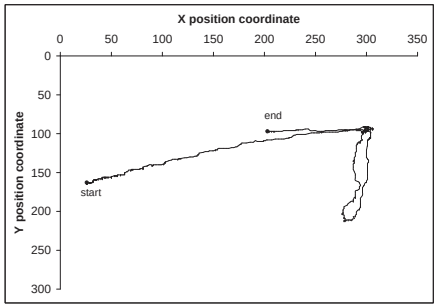
4.2.1.2 Applying the Measure of Intention to the Data

Let us begin by describing how we measure intention using our model. We used the ground truth position data of the selected videos as a basis for our intention measurements. Every frame in the video was taken as a state in the trace, with the planar coordinates of the filmed character describing it. The Euclidean distance was used as the distance measure. As above, for every state we calculated the distance from the initial state, and then checked for how many of those states the distance increased, relative to the previous state.

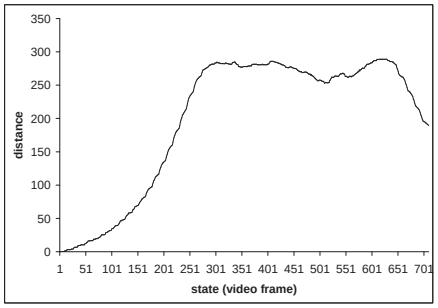
Figure 4.5a plots the path of movement of the observed character, in planar coordinates, for clip number 6, which was taken from video br1gt.mpg of the repository. The character starts moving from the left towards the right, where he spends some time standing in place (since we are only plotting planar coordinates, the amount of time spent at each point is not represented here). From there the character turns downwards, then back upwards, once again spending time at the same spot, and finally moving leftwards, towards the starting point. Figure 4.5b graphs the distances of each state in the path, from the initial state. The X axis marks the video frame number, and the Y axis measures the distance from the initial location of the person in question. Note how for the first 300 frames or so, the graph rises gradually, corresponding to the part of the path where the character moves away from the starting point. Where the character stands in place, the distance graph stays more or less constant. Towards the end of the clip, when the character moves back towards the starting point, the distance drops. The measure of intention for this movement path, as we calculated it, was $t = 0.4$. Using a cutoff value of 0.5, this movement was classified as non-intentional. The interested reader is invited to watch the video and compare it to the graphs presented here.

4.2.1.3 Comparing to Human Judgment

These same video clips were shown to human subjects who were asked to write down their opinion regarding the intention of the viewed character. They were given the option of segmenting the video if they thought the character changed its intention along the trace. Segmentation was enabled at a resolution of seconds.



(a) Path of Movement.



(b) Distances of each state from initial state.

Figure 4.5: Examples from Clip Number 6.

Using Entropy to Measure Divergence of Responses. Here we faced some difficulty in the experiment design. In pilot experiments, it became clear that asking the subjects to directly rank the ”strength of intention” of a video segment leads to meaningless results. For instance, some subjects in pilot experiments chose to give high intention scores to a video segment showing a person seemingly walking around aimlessly. When we asked for an explanation, the answer was that the person in the video clearly intended to pass the time. Such an understanding does not fit the sense of intention with which we are dealing in this study. We thus needed to measure intention indirectly. To do this, subjects were requested to write down a sentence describing the perceived intention of the person in the video, typically beginning with the words ”the person intends to ...”. The idea behind this is that in segments where there is clear intention, a clear answer would emerge (for instance, ”The person intends to exit the room”); in other video segments, the unclear intention would result in more highly varied answers (e.g., some would write ”intends to pass the time”, while others would write ”intends to walk”, etc.). This divergence can be measured by various means; we chose the information entropy function as it is used in statistics to measure dispersion of categorical data.

Using Content Analysis to Extract Categories from the Responses. Before applying this to our data, we first had to standardize the replies, which were given as natural language answers to an open-ended question. A finite number of categories needed to be chosen and assigned to the different descriptions, in a consistent and reliable way. We turned to the social sciences methodology for studying

the content of communication, known as content analysis [Babbie, 2003], and used this as motivation for the analysis described in the following. Two independent coders each analyzed all the input from the subjects. From every description, a verb (e.g. walk, look) and a noun (e.g. location, object) were extracted, reducing the sentence to two words, which together consisted of a unique category. Where the two coders disagreed as to the category to be applied to a given description, a third arbitrator decided between them. The chosen categories were then applied to the data. For every video clip the entropy was calculated per second and then averaged over time, producing a single entropy value for each of the video clips.

4.2.2 Results

4.2.2.1 Measure of Intention Correlates With Human Judgment

Table 4.6 summarizes the resulting entropy values of this analysis, alongside the intention scores as returned by our method. Figure 4.6 plots entropy versus intention of the eleven video clips analyzed. Every point in the graph represents one video clip, analyzed as described above to produce two values. The X axis is the intention measure as calculated by our method, and the Y axis is the entropy value, reached by calculating the divergence of categories across subjects per second, averaged over time.

We calculated the correlation between the entropy and the intention, and found it to be strongly negative at -0.685. The significance of this value was checked using Fisher's r to z transform. We ran a Z test to check the probability of the null hypothesis that the entropy and the intention are uncorrelated, which resulted in P=0.0096. We conclude that the correlation is indeed significant.

This result confirms our conjecture that our method does capture the notion of intention, as judged by humans. This is what we were expecting to see—that the entropy is significantly negatively correlated with the intention. The higher the entropy of movement of a given character, the less clear that character conveyed intention to the observing subjects, and the lower the intention calculated by our method. The inverse is true as well—the lower the entropy of movement of a given character, the clearer its intention was to the human observers, and the higher the intention score achieved by our method.

Clip Number	Intention	Entropy
1	0.644	0.370
2	0.552	0.622
3	0.861	0.160
4	0.636	0.730
6	0.408	0.514
7	0.366	0.483
8	0.431	0.495
9	0.449	0.871
10	0.611	0.521
11	0.481	0.879
12	0.094	0.999

Table 4.6: Measure of Intention and Entropy of Human Judgments for Video Clips.

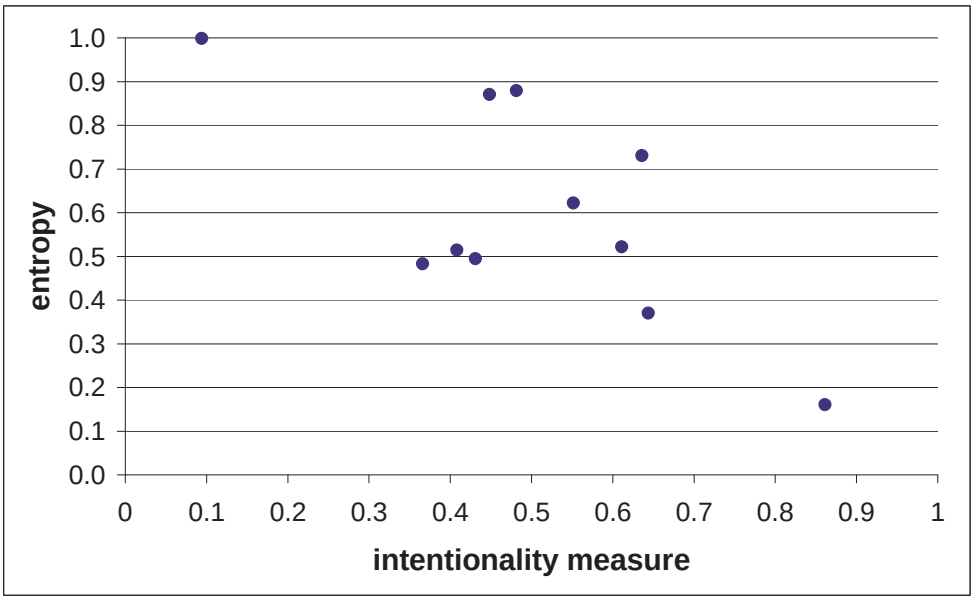


Figure 4.6: Plot of Entropy vs. Intention. Correlation=—0.686.

4.2.2.2 Using Measure of Intention for Segmenting Subgoals

While analyzing the results of the second experiment, the matter of parsing streams of action according to sub-goals arose. Several of the clips we analyzed clearly show changing intentions, e.g. when the character stops in mid-track, turns around 180 degrees, and moves in the opposite direction. If a sequence of actions is expected to have at most one possible goal, when in fact it is composed of several sub-goals, then an observing agent behaving according to our measure of intention would be confused. Take for example the case of a person intending to reach one location, and having accomplished that, moves on to the goal of returning to his original location. If we consider this to be one coherent stream of action, with one goal, which is the resulting end-state, then obviously an agent using this measure would come to the conclusion that there is no underlying intention, since, had the person been intending to be at his original location, he would not have taken the unnecessary and inefficient steps of moving to a different location and then back home. If, on the other hand, it is understood that the stream of action must first be parsed into sub-streams, then each sub-stream can be dealt with separately, by applying the measure of intention to it. Every sub-stream could then be seen as efficient in bringing about its respective sub-goal.

In this experiment, we allowed the participating subjects to write down more than one intention per video clip, in accordance with the way they perceived the intentions changing with time. However, the intention score given to a trace of movement according to our method takes into account the complete trace from beginning to end, without allowing for the possibility of changing intentions along the way. We turn to this possibility now, asking how can these changing intentions be dealt with? Instead of taking only one final intention score, we calculated our measure at every point in the path, and inspected the changes along the resulting graph. We wanted to see if the behavior of the graph of intention, as measured by us, could indicate significant changes in the intention of the observed character. If so, this could prove a useful tool for segmenting sequences of action into subgoals.

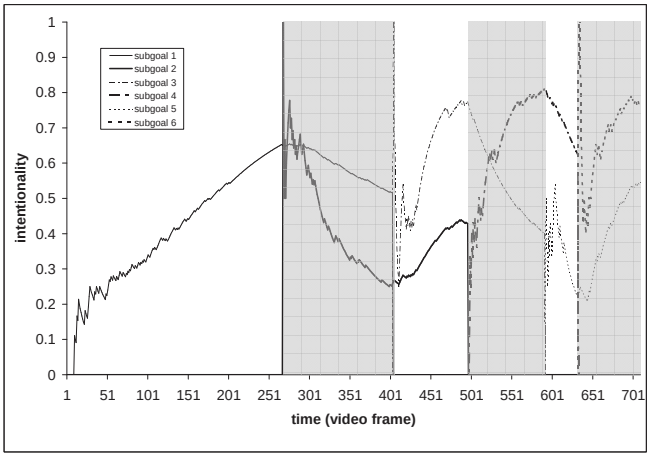
To do this, for each video clip we examined the graph of intention and marked the first clear change of trend in the graph. Reaching an obvious maximum, minimum, or plateau were considered to be clear changes in trend. At the marked point a new subgoal was assumed found, and a new intention score was calcu-

lated, using the previous segment's terminating state as the new segment's initial state. Once again, the first change of trend was marked, and so on, until the end of the intention graph was reached. Given the time frames at which the graph was segmented, the corresponding points along the path of movement were indicated.

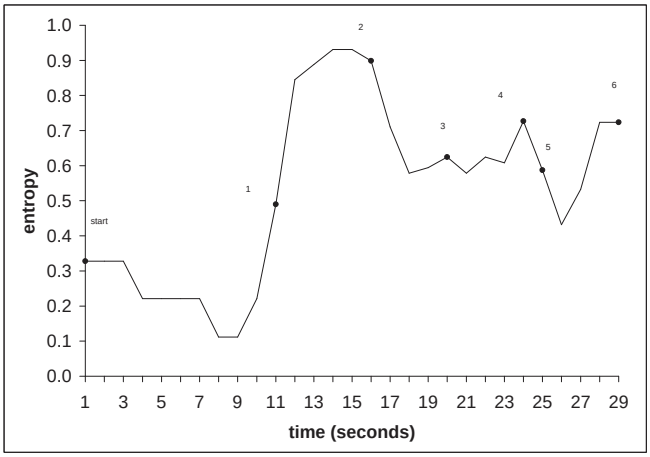
Video clip number 6 is given as an example, in Figure 4.7. The plot area is divided into alternating white and gray strips, corresponding to subgoals found according to the process described above. In the first vertical white area, the plot of intention begins. Where it first peaks significantly, a subgoal is parsed, and the calculation of intention begins again, with the terminal state of the previous segment taken as the initial state for the current segment. In the subsequent vertical gray area, the previous segment's intention plot is continued, so as to demonstrate the significance of the peak, and the second segment's intention plot begins. Where a significant minimum is reached in it, a new vertical white area begins, indicating the new subgoal found. In this area, again, the segment of the previous subgoal continues, so as to demonstrate the minimum found, and the plot of intention for the third subgoal begins. The first subgoal's plot is no longer shown here. And so on—every strip in the plot contains two subgoals' intention plots—the previous and the current (except for the first strip, which contains only the first subgoal).

Figure 4.7c shows where the points found fall along the path. Clearly, the places where subgoals were found to begin mark significant changes of direction or movement—the segment between the "start" point and subgoal 1 have the character moving from left to right, between points 1 and 2 the character is standing in place, between 2 and 3 moving down, 3 and 4 moving up, 4 and 5 standing in place, 5 and 6 moving to the right. This data is summarized textually in Table 4.7.

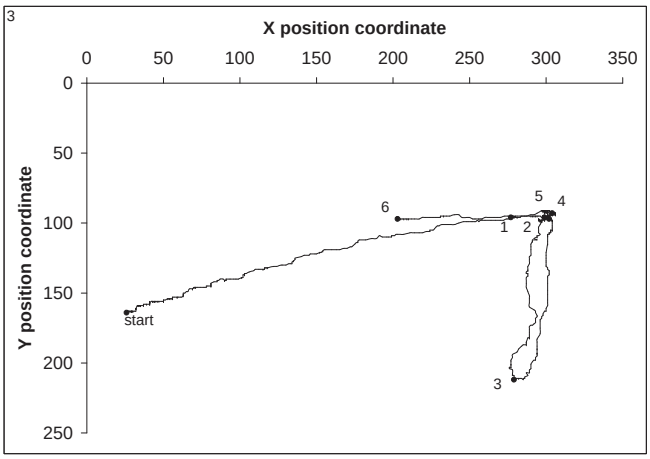
Figure 4.7b shows the plot of entropy as it changes over time, with numbers indicating where subgoals were found. Note that the behavior of the entropy graph is somewhat inverse to the behaviors of the intention graphs of the subgoals—for the first subgoal, the intention graph is increasing, while the entropy graph in that section is decreasing. For the second subgoal, the intention decreases while the entropy increases. The third subgoal also holds this inverse relationship, but the last 3 subgoals do not continue to show such a correspon-



(a) Subgoal Parsing from Intention Graph of Video Clip 6.



(b) Entropy of Video Clip 6.



(c) Path of movement in Video Clip 6.

Figure 4.7: Analysis of Video Clip 6.

Frames	Seconds	Coordinates	Trend	Character	intention
1-265	1-11	(26,164)-(277,96)	increases	walks to ATM	0.653
265-402	11-16	(277,96)-(301,97)	decreases	stands at ATM	0.257
402-495	16-20	(301,97)-(278,212)	increases	walks down	0.774
495-590	20-24	(278,212)-(299,96)	increases	walks up	0.811
590-631	24-25	(299,96)-(304,93)	decreases	stands at ATM	0.220
631-707	25-28	(304,93)-(203,97)	increases	walks up	0.766

Table 4.7: Description of Subgoals Found in Video Clip 6.

dence. Perhaps this is so since those last sections are not very long, and don't contain enough data for the trends to come forth strongly.

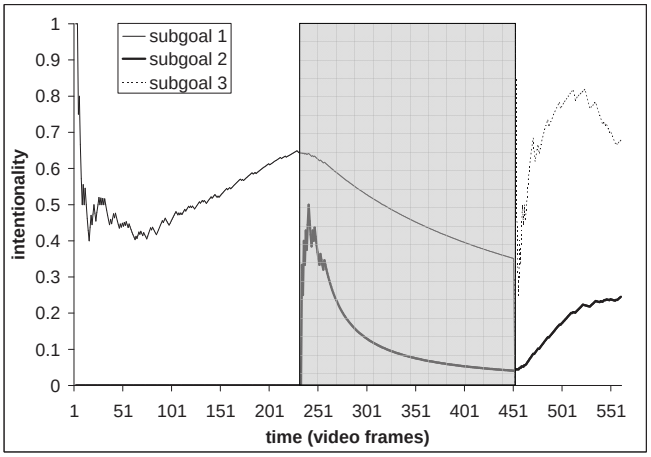
Figure 4.8 depicts the same analysis applied to video clip number 7, serving as another example of the value of the proposed measure of intention for parsing subgoals. A textual summary of the subgoals is given in Table 4.8

Frames	Seconds	Coordinates	Trend	Character	intention
1-231	1-9	(91,70)-(291,98)	increasing	walks to ATM	0.645
231-451	9-18	(291,98)-(306,98)	decreasing	stands at ATM	0.04
451-561	18-22	(306,98)-(287,0)	increasing	leaves ATM	0.679

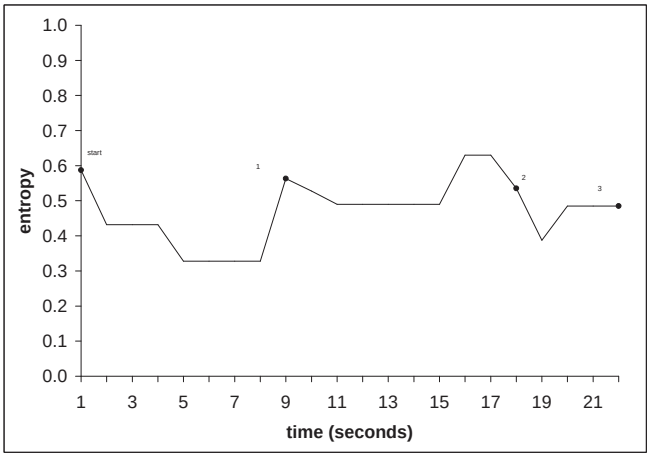
Table 4.8: Description of Subgoals Found in Video Clip 7.

Another example is given in video number 3. This is a simpler example, in which no subgoals were found. Its analysis is shown in Figure 4.9. The character in this video moves in a straightforward manner from the bottom of the screen to the top. Fittingly, the intention score achieved is high, and the entropy is low. The intention graph is smooth—no clear peaks or troughs are present—and so does not indicate any points of changing intentions. The slight change noticed right at the beginning of the path—from moving left to moving up—is obscured by the general noise always present at the beginning of intention graphs, until enough data has accumulated to give a meaningful score.

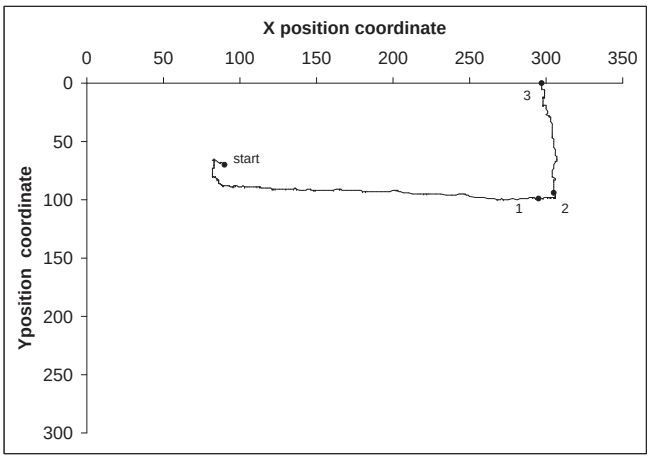
While our results do indicate that the proposed measure of intention can be useful for parsing subgoals, there are some examples in which the segmentation is less than perfect. In some cases, subgoals are found where they don't exist, as in video clip number 1, shown in Figure 4.10. Table 4.9 describes the two



(a) Subgoal Parsing from Intention Graph of Video Clip 7.

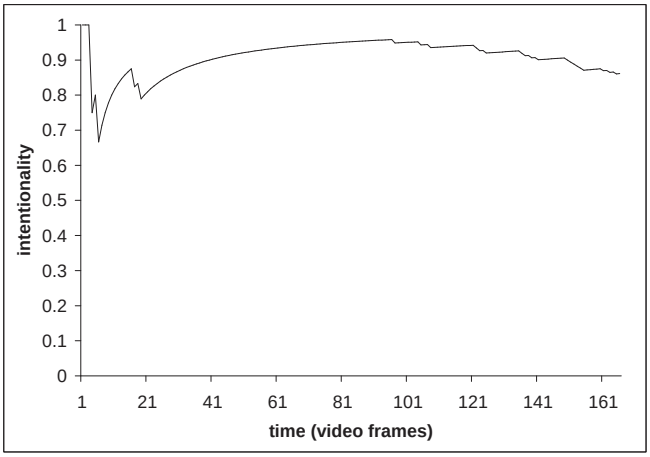


(b) Entropy of Video Clip 7.

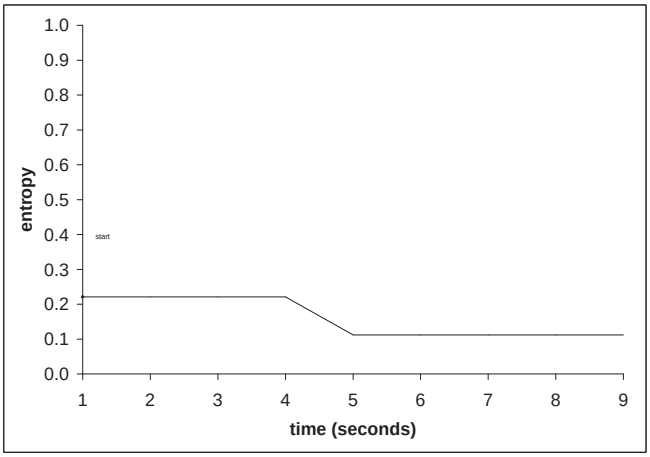


(c) Path of Movement in Video Clip 7.

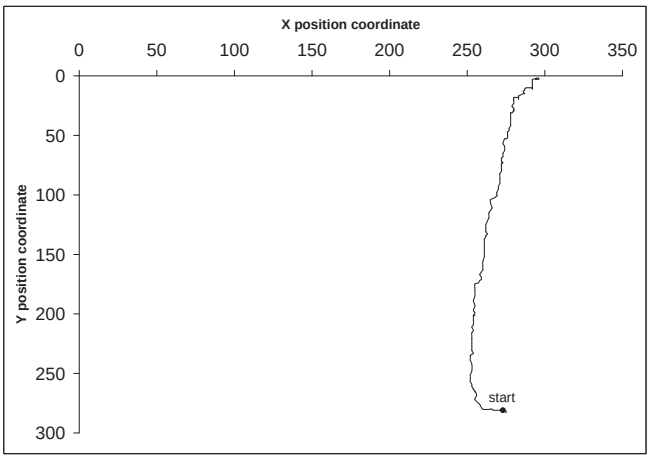
Figure 4.8: Analysis of Video Clip 7.



(a) Subgoal Parsing from Intention Graph of Video Clip 3.



(b) Entropy of Video Clip 3.



(c) Path of Movement in Video Clip 3.

Figure 4.9: Analysis of Video Clip 3.

subgoals found for this clip. In this example, there is an apparent change of curvature in the path at the segmentation point, however it does not seem prominent enough to justify parsing. Indeed, the change of trend in the intention graph is not prominent either, so perhaps using a stricter definition for identifying changes of trend would eliminate such false positive instances.

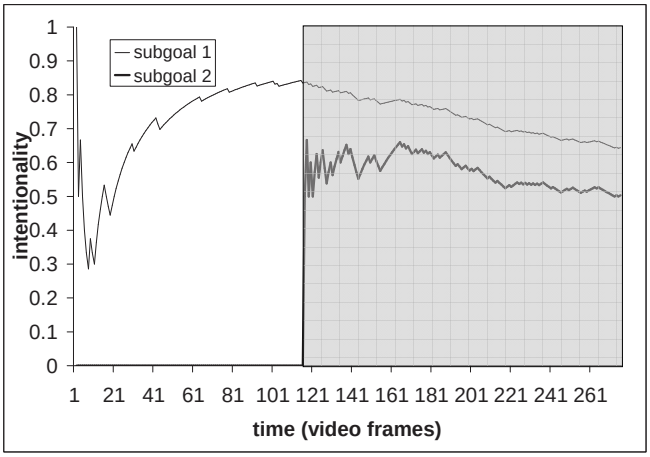
Frames	Seconds	Coordinates	Trend	Character	intention
1-113	1-6	(244,285)-(39,132)	increase	walks up	0.842
113-274	6-11	(139,132)-(82,63)	decrease	walks up	0.503

Table 4.9: Description of Subgoals Found in Video Clip 1.

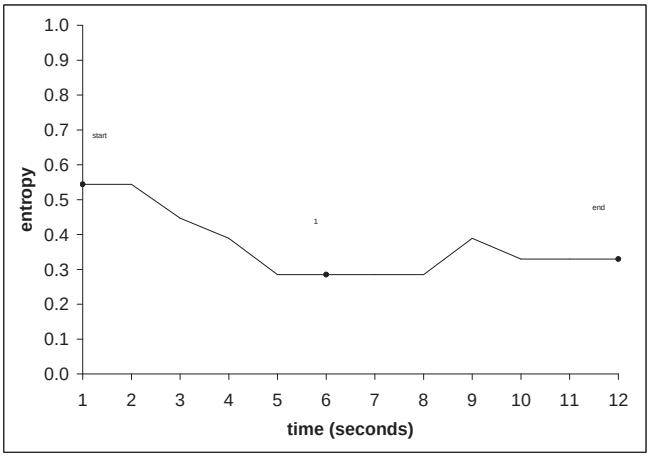
Another example is given in Figure 4.11, this time of the false negative kind, with the description of subgoals in Table 4.10. Using our method, three subgoals were found, while it seemed to us that the third subgoal should have been parsed into an additional subgoal, at the sharp turn the character takes halfway through the subgoal. Perhaps the short length of this section did not contain enough data for such a precise cut. Another possibility is that this is another case where more rigorous criteria for changes of trend in the intention graph might fix the problem. Since overall our method does succeed at segmenting subgoals—as the first few examples show—we did not go into the fine tuning of the parameters. The exact parameters for subgoal parsing need to be found when bringing this method down to practical implementation.

Frames	Seconds	Coordinates	Trend	Character	Intention
1-344	1-15	(29,163)-(177,204)	increase	strolls down	0.474
344-458	15-19	(177,204)-(174,217)	decrease	walks around	0.404
458-509	19-21	(174,217)-(237,283)	plateau	walks down	0.961

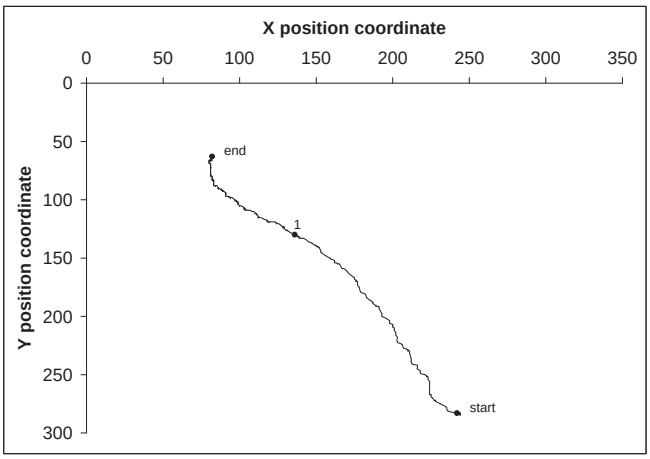
Table 4.10: Description of Subgoals Found in Video Clip 11.



(a) Subgoal Parsing from Intention Graph of Video Clip 1.

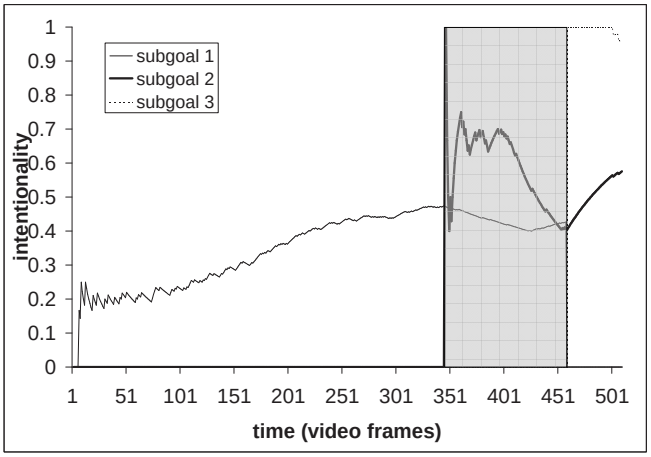


(b) Entropy of Video Clip 1.

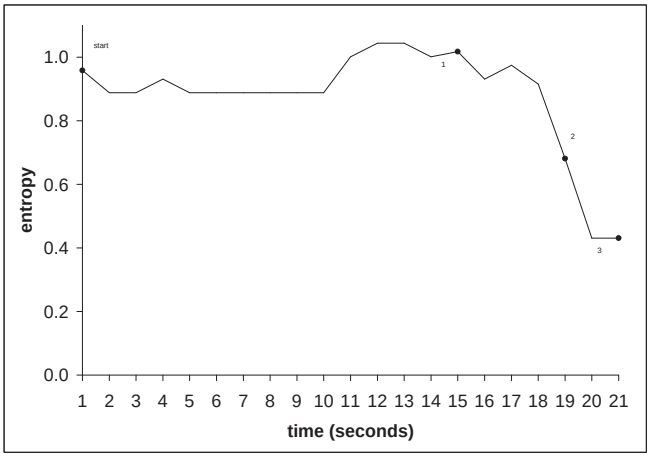


(c) Path of Movement in Video Clip 1.

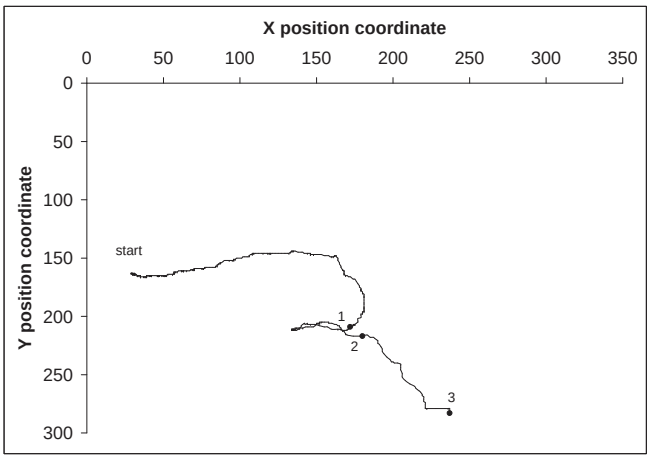
Figure 4.10: Analysis of Video Clip 1.



(a) Subgoal Parsing from Intention Graph of Video Clip 11.



(b) Entropy of Video Clip 11.



(c) Path of Movement in Video Clip 11.

Figure 4.11: Analysis of Video Clip 11.

Chapter 5

Experiments for Evaluating Heuristics of Intention Prediction

We turn now to the task of determining the content of the intention detected in an observed sequence of actions, i.e., predicting the goal state which the actor was intending to bring about by his actions.

According to the theoretical background on affordances reviewed above, we posit that upon perception of objects in the environment, afforded goal states are invoked in the mind of the observer. Our task, therefore, is to extract information from the observed sequence of actions performed on the objects, in order to determine which of the afforded goal states is the one at which the actions are aimed.

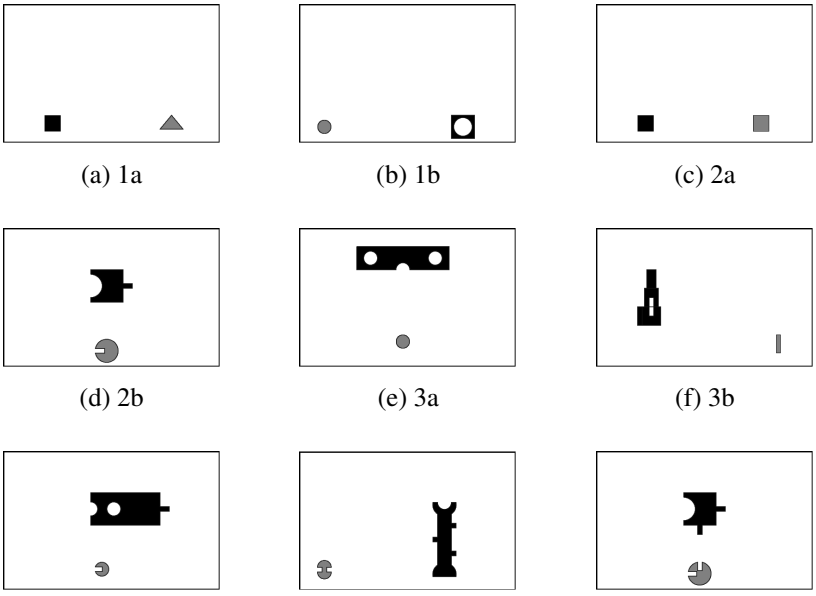
In the following section we describe our experiment, in which human subjects were asked to determine the intention underlying an observed action sequence. We show how the observed process of human intention recognition can be explained according to the three values produced by the heuristics given in Section 3.5—the prior p_j, the distance d_j and the intention measure r_j, computed for each afforded goal state $g_{j=1,\ldots,k}$. We hypothesize that choosing the highest ranking goal according to the intention measure r_j, best approximates the preference demonstrated by the subjects participating in the experiment. The data confirms this, as well as the fact that d_j, and p_j play secondary roles with regard to this task.

5.1 Experimental Setup

As an environment in which to evaluate our model, we chose what could be seen as a two-dimensional version of Meltzoff's setup: scenarios in which two geometric objects exist, one stationary and the other movable. We used several pairs of such objects. Part I of the experiment was meant to determine the various possible afforded goal configurations of each pair of objects, i.e., the g_j's, along with their associated prior probabilities, the p_j's. This is in line with our assertion that upon perceiving the objects, several possible goal-states would be retrieved from the so-called affordance library of the perceiver, along with a distribution over them. Part II of the experiment shows how the priors for these goals interact with the two other values mentioned (distance and intention) in order to determine the intention underlying the observed sequences of actions. We used the Euclidean distance as our distance measure.

The experiment was run as two web applications—one for each part—and the URL addresses were given to approximately 140 computer science undergraduates, who participated in return for credit (mean age: 21.2(3.57), 112 male). The first application consisted of a succession of nine screens, in each of which a pair of objects was presented to the subjects: a black stationary one, and a grey movable one. The subjects were instructed to drag the grey object to whichever configuration seemed "natural" to them, in relation to the black object. The locations chosen by each subject were recorded, for each pair of objects, as were the trajectories of movement leading to those choices. The pairs of objects used are shown in Figure 5.1, with a corresponding identification code for each.

Two weeks after the results of the first part were analyzed, the second application was designed and implemented. It had the subjects view manipulations of the grey object for five of the nine object-pairs used in the first part. For each pair, several paths were constructed, and the grey object was animated along those paths. The subjects were told that the animations they were viewing were from the results of one of the subjects ("student X") on the first part of the experiment—student X had dragged the grey object in each pair to a specific location, but only the first part of the trajectory was being shown. The subjects were instructed to complete the trajectory and drag the grey object to the location where they thought student X had intended to place it. In both applications the order in which the screens were presented was randomized.



(a) 1a (b) 1b (c) 2a

(d) 2b (e) 3a (f) 3b

(g) 3c (h) 4a (i) 5a

Figure 5.1: Object-Pairs and Their Identification Codes.

5.2 Results

The results of the first part of the experiment justify our understanding that non trivial priors exist for possible goals. According to the results of the second part of the experiment, the heuristic based on the intention measure proves most useful for correctly predicting the intended goal. In addition to these two major results, we also suggest using the distance measure or the prior distribution for choosing among afforded goal states for which the intention measure is maximal, i.e. in the case of a tie. The last point of interest arising from the results concerns the generation of new affordances. As this is not the topic of this study, we only briefly touch upon it at the end of the results section.

5.2.1 Existence of Non Trivial Priors for Possible Goals

The null hypothesis for the first part of the experiment would be that, having never before seen the objects presented, the subjects would choose all possible goal configurations with equal probability. The results, however, clearly reveal that non trivial priors do exist for the object-pairs presented. Of course, some object-pairs are more natural than others. For example, pair $1a$ begs to be config-

ured as a house (Figure 5.2), with the grey triangle placed atop the black square, which is presumably why this goal configuration was chosen by 96.49% of the subjects. Other pairs also produced a clear tendency among the subjects to prefer

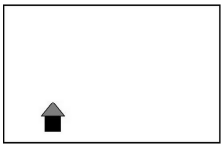
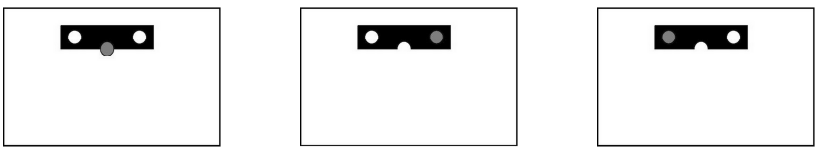


Figure 5.2: Most Frequent State (A) for Object-Pair 1*a*, with Prior 96.49%.

one configuration over another. As an example, consider pair 3*a* (Figure 5.3), for which the subjects chose to place the grey circle in the middle indentation at the bottom of the black object with 69.29% frequency, while they placed the grey circle in the right hole with 10.71% frequency and in the left hole with 12.86% frequency. Such choices could be due to properties such as symmetry and size, however, we are not interested in *why* these preferences emerge, but rather in the fact that they do indeed emerge. Obviously, different pairs of objects afford different configurations, which is why we have taken the liberty to refer to these states as ”affordances”.

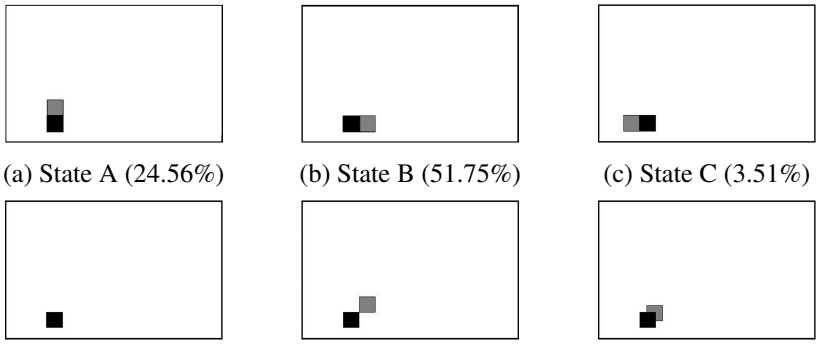
Prior probabilities of states, as determined by the frequencies at which subjects chose the different configurations in the first part of the experiment, are shown in the following figures, for each of the remaining object-pairs. Capital letters denote the states—this lettering was chosen arbitrarily, and is not ordered by frequency. In addition, the lettering for each object-pair is independent—there is no relationship between states of different object-pairs which happen to have the same capital letter. Only states which were chosen by the subjects with frequency above 3% are shown, which is why the sum of frequencies does not always amount to 100%—states with negligible frequency are not shown.



(a) State A (69.29%) (b) State B (10.71%) (c) State C (12.86%)

Figure 5.3: Most Frequent States for Object-Pair 3*a* with Their Priors.

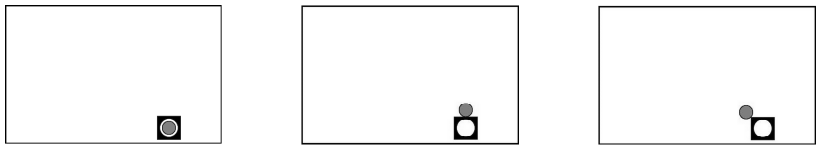
Figure 5.4 shows the empirical priors for object-pair 2*a* (note that in state D—Figure 5.4d—the grey square is placed behind the black square, thus obscured by it). Figure 5.5 shows the priors for object-pair 1*b* (state E—Figure 5.5c—is shown as an example of a configuration which was not chosen at all in this first part of the experiment, and it will be referred to later on, in the second part of the experiment). Figures 5.6, 5.7, 5.8, 5.9 and 5.10 show the priors for affordances of object-pairs 2*b*, 3*b*, 3*c*, 4*a* and 5*a*, respectively.



(a) State A (24.56%) (b) State B (51.75%) (c) State C (3.51%)

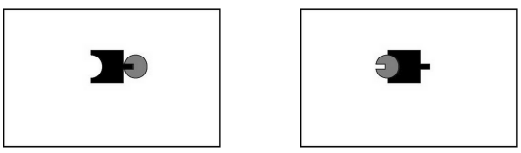
(d) State D (7.89%) (e) State E (3.51%) (f) State F (3.51%)

Figure 5.4: Most Frequent States for Object-Pair 2*a* with their Priors.



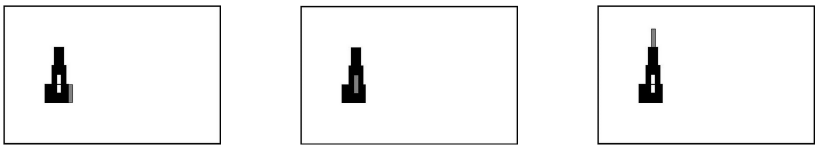
(a) State A (85.82%) (b) State B (11.35%) (c) State E (0.00%)

Figure 5.5: Most Frequent States for Object-Pair 1*b* with their Priors.



(a) State A (68.42%) (b) State B (30.70%)

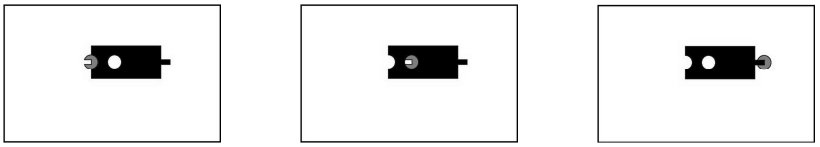
Figure 5.6: Most Frequent States for Object-Pair 2*b* with their Priors.



(a) State A (5.63%) (b) State B (77.46%) (c) State C (11.27%)

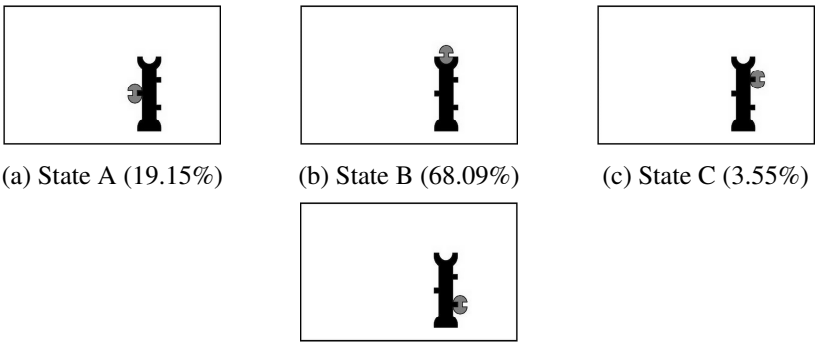
(d) State F (1.41%) (e) State I (0.00%) (f) State J (0.00%)

Figure 5.7: Most Frequent States for Object-Pair 3*b* with Their Priors.



(a) State A (30.22%) (b) State B (17.99%) (c) State C (48.20%)

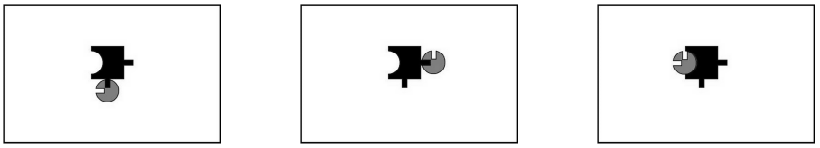
Figure 5.8: Most Frequent States for Object-Pair 3*c* with Their Priors.



(a) State A (19.15%) (b) State B (68.09%) (c) State C (3.55%)

(d) State D (4.96%)

Figure 5.9: Most Frequent States for Object-Pair 4*a* with Their Priors.



(a) State A (41.23%) (b) State B (23.68%) (c) State C (33.33%)

Figure 5.10: Most Frequent States for Object-Pair 5*a* with Their Priors.

5.2.2 Intention Measure for Ranking Goals

Of the three heuristics proposed, the one guided by our intention measure proves to be most informative for inferring the intended goal. Ranking the candidate goals from Part I, g_j, according to their intention measures, r_j, and choosing the highest ranking one, results in the same goal most frequently chosen by the subjects in Part II. In other words, the goal with the highest intention measure coincides with the goal most frequently chosen by the subjects. This observation holds for all five object-pairs and their respective paths of movement demonstrated in the experiment, except for one case, as will be shown in the following.

Before going into the detailed quantitative results, we first present a qualitative summary, in Figure 5.11. This figure shows the success rate of each of the heuristics, at matching the goal state most often chosen by the subjects as the intended one.

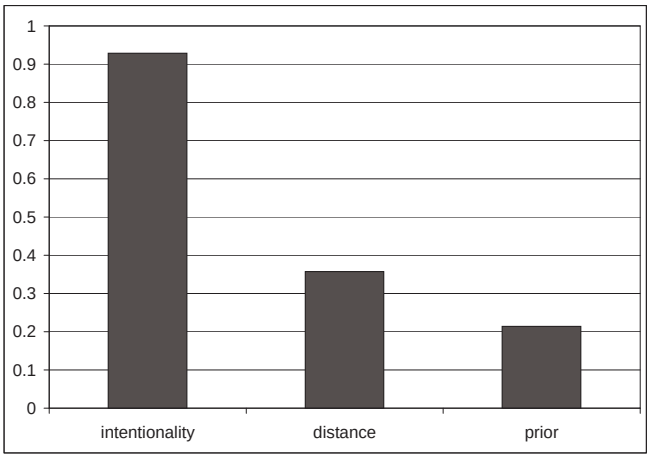


Figure 5.11: Success Rate of Each Heuristic at Predicting the Correct Goal.

The details of these matchings are given in Table 5.1. Every row in the table corresponds to one demonstration of movement—identified by an object-pair and a path. For each such demonstration, the goals with the highest rank are given, according to each measure. The column titled ”Most Frequent” gives the goal state most frequently chosen by the subjects in Part II. This is the goal state we are attempting to guess. The next column, titled ”Maximal Intention”, gives the goal state achieving the highest intention measure. Next, ”Minimal Distance”, gives the goal state which has the shortest distance from the terminal state of the observed path. And last, ”Maximal Prior”, gives the goal state chosen

most frequently by the subjects in Part I (this prior is constant across all paths of a given object-pair). In several instances, more than one goal state achieved the highest value for a given measure. In those cases, all those goal states are given, separated by commas.

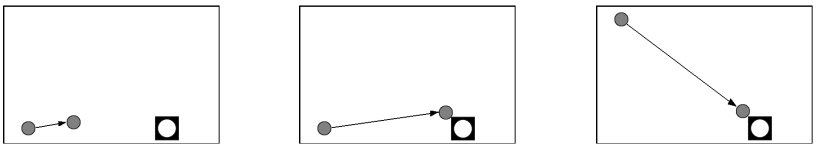
Note that in both parts of the experiment we measured the frequencies of choices of the various resulting goal states—in the first part, given the object-pairs alone, and in the second part, given the object-pairs being manipulated in movement. When referring to the results of Part I, we call these frequencies *priors*. They should not be confused with the frequencies of choice from Part II, which are the results of the observed behavior which we are attempting to match.

(Object-Pair, Path)	Most Frequent	Maximal Intention	Minimal Distance	Maximal Prior
(1*b*, I)	B	B	A	A
(1*b*, II)	B	B	B	A
(1*b*, III)	A	A	B	A
(3*a*, I)	C	C	A	A
(3*a*, II)	C	C	C	A
(3*a*, III)	A	A,B	A,C	A
(3*b*, I)	J	A	C	B
(3*b*, II)	A	A	C	B
(3*b*, III)	B	B,C	C	B
(3*c*, I)	B	B	C	C
(3*c*, II)	B	B	C	C
(3*c*, III)	A	A,B,C	A	C
(4*a*, I)	A	A	B	B
(4*a*, II)	A	A	A	B

Table 5.1: Most Frequently Chosen Goal State vs. Choice According to Heuristics per Object-Pair and Path.

Note how column "Maximal Intention" matches column "Most Frequent" in all but one of the total 14 demonstrations (object-pair 3*b*, Path I), while column "Minimal Distance" does not match in nine of them. "Maximal Prior" matches in only three of the 14 demonstrations. This analysis summarizes the findings and justifies our conclusion that, of the three heuristics proposed, the intention measure is best at predicting the intended goal. We next go into the details of the results, pointing out various aspects of the findings along the way.

Object-Pair 1*b*. For object-pair 1*b*, three paths were shown to the subjects (Figure 5.12). Paths I and II share a common initial state, with Path II continuing on past the terminal state of Path I. Paths II and III share a common terminal state, and differ with regards to their initial state. The three afforded states most frequently chosen by the subjects in Part II were *A*, *B* and *E* (refer to the above-mentioned Figure 5.5), and the frequencies according to which they were chosen, for each path, are given in Table 5.2.



(a) Path I (b) Path II (c) Path III

Figure 5.12: Paths for Object-Pair 1*b*.

state\path	I	II	III
A	16.67	8.33	90.91
B	65.15	78.79	1.52
E	9.85	9.85	3.79

Table 5.2: Frequencies of Choices for Object-Pair 1*b*.

We now compare these empirical results for this object-pair to the prediction based on ranking according to the intention measure. Note that we calculate the intention measure only for states *A* and *B*, since these are the only states which achieved significant positive priors in the first part of the experiment (85.82% and 11.35% respectively, as shown in Figure 5.5). These values of the intention measure are shown in Table 5.3. The results show that for each path, the state scoring the highest intention is also that which was most often chosen by the subjects. Noticeably, by manipulating the trajectory, we were able to cause the subjects to infer a goal which had a relatively low prior probability.

It is interesting to further compare paths I and II: the results show that the longer Path II left less room for ambiguity in the subjects' decision between states *A* and *B*, so that although state *A* was not chosen with highest frequency for either path, its frequency of choice for Path I was higher than for Path II.

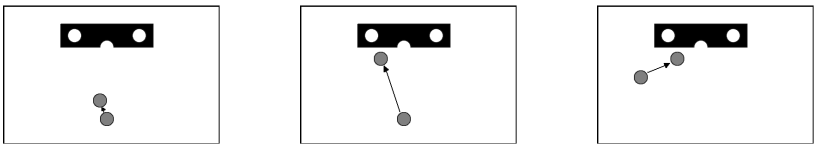
state\path	I	II	III
A	0.996	0.949	0.999
B	1.000	1.000	0.973

Table 5.3: Measure of Intention for Object-Pair 1*b*.

The measure of intention also reflects this—state *A* received a lower value of intention in Path II than in Path I.

Another point worth noting is that both states received high measures of intention for all three paths, and the differences between these values, while significant, are not great. This does not reflect the substantial gaps between their respective values of frequency of choice. For example, in Path III, the measure of intention of state *A* is greater than that of state *B* by 0.026%, while the frequency of choice of state *A* is greater than that of state *B* by 89.39%. Thus, while ranking according to this intention measure preserves the order of frequency of choice, the relative weights of the values do not correspond. However, since for the task at hand we are only interested in choosing the highest ranking afforded state, we need not be concerned about normalization.

Object-Pair 3*a*. Object-pair 3*a* supports these results as well. Here too, three paths were shown to the subjects (Figure 5.13). Path II begins as Path I does, and



(a) Path I (b) Path II (c) Path III

Figure 5.13: Paths for Object-Pair 3*a*.

continues further. Paths II and III end at the same position, but begin at different ones. Table 5.4 presents the empirical results for this object-pair—only the three most frequently chosen states are shown, since the others achieved negligible frequencies. The states themselves are depicted in Figure 5.3.

Table 5.5 gives the calculated measure of intention for each of the three paths and the three most frequently afforded states (from the first part of the experiment). Ranking the possible intended states according to this measure, we arrive

state\path	I	II	III
A	14.18	3.73	79.85
B	0.00	0.00	16.42
C	78.36	92.54	0.75

Table 5.4: Frequencies of Choices for Object-Pair 3*a*.

at results quite close to those of our subjects'. The only difference is in Path III, where states *A* and *B* both achieve the maximal intention score of 1. We later show what information can be used to break such a tie.

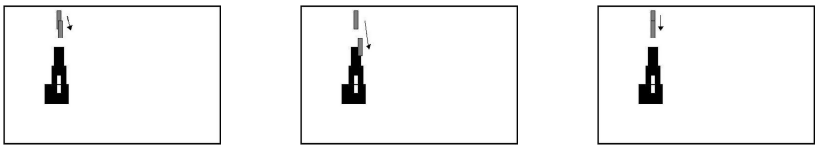
state\path	I	II	III
A	0.98	0.79	1.00
B	0.94	0.72	1.00
C	1.00	1.00	0.75

Table 5.5: Measure of Intention for Object-Pair 3*a*.

Object-Pair 3*b*. Results for object-pair 3*b* are shown next. Figure 5.14 depicts the three different paths shown to the subjects. Here, the movable object in all three paths starts out at the same position. Path II begins as Path I, and continues a bit farther, while Path III moves in a slightly different direction from the start. Table 5.6 shows the subjects' choice of goal states for each of the paths. Table 5.7 shows the measure of intention for each of the goal states which achieved significant priors (above 3%) in the first part.

Notice that three new goal states appear at this stage, in Table 5.6—goals which were not chosen with significant frequency in the first part of the experiment (or not at all), yet in the second part they were. In Paths II and III, this does not affect our prediction according to the measure of intention, since the goal states achieving the highest rank according to this measure turn out to be one of the original three which achieved high priors (*A*, *B*, *C*). However, in Path I, the original three goal states, *A*, *B* and *C* are each chosen by the subjects in this second part of the experiment with frequency below 20%. Only goal state *J*, which in the first part of the experiment was *not chosen by any of the subjects*, received the most "votes" here—25.18%. This is the only case in which our measure of

intention fails to predict the correct goal state. We will return to this issue when discussing dealing with new affordances.



(a) Path I (b) Path II (c) Path III

Figure 5.14: Paths for Object-Pair 3*b*.

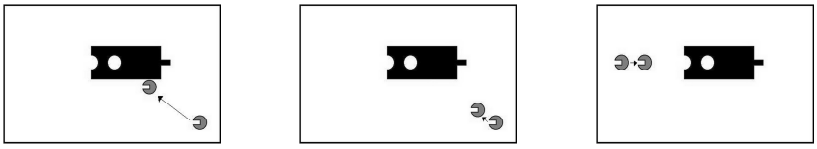
state\path	I	II	III
A	13.67	41.73	0.00
B	18.71	7.19	59.71
C	14.39	0.00	36.69
F	7.19	28.06	0.00
I	17.27	0.72	0.72
J	25.18	18.71	0.00

Table 5.6: Frequencies of Choices for Object-Pair 3*b*.

state\path	I	II	III
A	0.999997339	0.999988829	0.998315324
B	0.998137457	0.99167065	1
C	0.988936353	0.484244163	1

Table 5.7: Measure of Intention for Object-Pair 3*b*.

Object-Pair 3*c*. The fourth of the object-pairs presented to the subjects was 3*c*. The three paths for this pair are given in Figure 5.15. Here, Path II is a short "version" of Path I, while Path III shares nothing in common with them. The frequencies of the subjects' choices are given in Table 5.8, and the measures of intention in Table 5.9. For Paths I and II, the highest ranking goal state according to the measure of intention matches the one most frequently chosen by the subjects. However, for Path III, all three candidate goal states achieved the maximal value of intention. As mentioned above, we will discuss strategies for disambiguating between such tied goal states in the following.



(a) Path I (b) Path II (c) Path III

Figure 5.15: Paths for Object-Pair 3*c*.

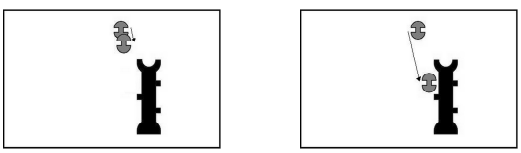
state\path	I	II	III
A	8.09	11.03	72.79
B	83.82	66.18	8.82
C	2.94	16.91	15.44

Table 5.8: Frequencies of Choices for Object-Pair 3*c*.

state\path	I	II	III
A	0.994128317	0.998505671	1
B	0.999995551	0.99998616	1
C	0.718324618	0.948277727	1

Table 5.9: Measure of Intention for Object-Pair 3*c*.

Object-Pair 4*a*. Object-pair 4*a* was the last of the five object-pairs used in this part of the experiment. The two paths for this pair are given in Figure 5.16, and the resulting frequencies for the four most chosen goal states are in Table 5.10. For both paths, state *A* was most often chosen even though its prior is significantly lower than that of state *B*. Calculated measures of intention are given in Table 5.11, and once again, the highest ranking goal state matches that which was most often chosen by the subjects.



(a) Path I (b) Path II

Figure 5.16: Paths for Object-Pair 4*a*.

state\path	I	II
A	69.29	85.00
B	19.29	2.14
C	1.43	2.14
D	2.14	1.43

Table 5.10: Frequencies of Choices for Object-Pair 4*a*.

state\path	I	II
A	0.999991597	0.999999761
B	0.918639653	0.463140117
C	0.972263947	0.791587525
D	0.992876392	0.950940113

Table 5.11: Measure of Intention for Object-Pair 4*a*.

5.2.3 Breaking Ties

The above analysis has shown that choosing the goal state with the highest intention measure will almost always correctly predict the intended goal state in a way which matches human predictions. However, in three cases (Path III of object-pair 3*a*, Path III of object-pair 3*b*, and Path III of object-pair 3*c*), more than one goal state achieved the highest value of intention, according to our measure. In all these cases, one of the tied goal states coincides with the goal state most frequently chosen as the intended one by the subjects. We will now discuss possible ways of decreeing which of the tied goal states should be chosen as the intended one.

While the distance measure and the prior values of the goal states proved to be inferior to the intention measure at the task of predicting intention, we propose that they can play a secondary role, for breaking ties. Given tied goal states, we can rank them according to their distance measures and according to their priors. Which of the two is better at breaking the ties and decreeing the intention in accordance with the subjects' choices?

Table 5.12 shows the highest ranking goal states for these three cases, where for the "Minimal Distance" and "Maximal Prior" column, the ranking was only between those goal states ranked equally maximally according to "Maximal in-

tention".

(Object-Pair, Path)	Most Frequent	Maximal intention	Minimal Distance	Maximal Prior
(3*a*, III)	A	A,B	A	A
(3*b*, III)	B	B,C	C	B
(3*c*, III)	A	A,B,C	A	C

Table 5.12: Most Frequently Chosen Goal State vs. Choice According to Heuristics per Object-Pair and Path, for Tied Goal States.

Inspection of this table does not resolve the issue. For the tied goal states of the case of object-pair 3*a* (*A* and *B*), both the distance measure and the prior prefer goal state *A*, which is what the subjects most often preferred. For the tied goal states of the case of object-pair 3*b* (*B* and *C*), the distance measure wrongly ranks *C* over *B*, while the prior correctly ranks *B* first. The inverse is true of the tied goal states of the case of object-pair 3*c* (*A*, *B* and *C*): the distance measure correctly decrees goal state *A* as the intended one, while the prior wrongly prefers *C* over *A*.

Nevertheless, it seems to us that the distance measure should be used for breaking ties. This, for the simple reason that it contains more information than the prior does—it takes into account the terminal state of the observed trajectory of motion, while the prior relies only on the affordances inherent in the objects themselves, regardless of the intentional manipulation performed on them. In addition, referring back to Figure 5.11, note that overall, distance was a better predictor of intention than prior.

Numerical details of the calculation of distance measure for each object-pair and each path, are given in the following Tables 5.13, 5.14, 5.15, 5.16, and 5.17. Note that the numbers given are not the absolute distance of the terminal state from the goal state, but rather that distance, divided by the total length of the path. This is for normalization purposes, and does not affect the relative ranking of the goal states.

5.2.4 Dynamic Generation of New Affordances

The one case in which our measure of intention failed at predicting the intended goal state occurred in the first path of object-pair 3*b*. Since the measure of in-

state\path	I	II	III
A	0.671867777	0.161047671	0.133559308
B	0.673362408	0.123362611	0.101528092
E	0.627397208	0	0

Table 5.13: Distances for Object-Pair 1*b*.

state\path	I	II	III
A	0.73	0.29	0.38
B	0.79	0.48	0.59
C	0.78	0.28	0.38

Table 5.14: Distances for Object-Pair 3*a*.

state\path	I	II	III
A	0.875117165	0.625354683	0.876706556
B	0.855813708	0.57024362	0.857142857
C	0.5	0.26550759	0.5

Table 5.15: Distances for Object-Pair 3*b*.

state\path	I	II	III
A	0.505994833	0.821498	0.62962963
B	0.409780754	0.787669299	0.72972973
C	0.340439936	0.686763288	0.836065574

Table 5.16: Distances for Object-Pair 3*c*.

state\path	I	II
A	0.792008403	0.201331258
B	0.668779763	0.358788311
C	0.795203728	0.359743257
D	0.844938759	0.429732538

Table 5.17: Distances for Object-Pair 4*a*.

tention is only calculated for those goal states which received a significant prior (above 3%) in the first part of the experiment, goal state J, the one voted most likely to be the intended goal, was not even considered. Had it been considered, its value of intention would have competed with that of goal state A, and then ties would have had to be broken, as discussed above. However, when using the intention measure within our framework for predicting the intended goal, we can only take into account afforded goal states—as determined by their priors.

This situation hints at the preliminary stage of acquiring affordances. While for the purposes of this study we assume a library of affordances already exists, along with a prior distribution over them, obviously, this assumption is not entirely correct. New affordances can be dynamically generated based on the observation sequence, and the perceived intention plays a role in their generation.

To see this, note that the failure of correctly predicting goal J in the first path of object-pair $3b$ demonstrates a new affordance being ”born”—although the goal state J was not chosen by any of the subjects as a possible configuration for object-pair $3b$ during the first part of the experiment, when presented with a display of intentional movement which did not seem to be aimed at any of the high-prior goal states (A, B or C), a new goal state somehow afforded itself to the observers.

The same can be seen in the case of object-pair $1b$. There, state E achieved a prior of zero during the first part of the experiment, yet, in the second part of the experiment it was chosen with significant frequency (9.85% in each of Paths I and II, and 3.79% in Path III). However, in this case, this phenomenon of a new afforded state being ”born” did not affect the performance of the prediction process, as compared to the results of the second part of the experiment, since, while this new goal state was chosen relatively often, it was not often enough to overcome the frequency of choice of the intended goal state—one which had achieved a high prior in the first part of the experiment.

In accordance with this, a complete model of the cognitive ability of goal prediction would have to take into account the process of affordance generation, and not rely only on those affordances already present in the repertoire of the observer. As crucial as this is for completing the picture, since it is an entirely different area of study, worthy of its own research, we do not go into it here. We only point out that even when leaving out this important ability of affordance

generation, the process we described was able to correctly predict the intended goal with close to 93% accuracy.

Chapter 6

Discussion and Future Work

Several points pertaining to the above-presented results deserve further considerations. Some of them clarify and justify the developed model, and some of them are left for future research. We conclude with a short summary of what we hope to be the contribution of this work.

6.1 Different Measures of Intention

Two related, but different, problems were addressed in this work. The first was the problem of determining whether an observed sequence of actions was being performed with an intention in mind. To this end, the measure t was proposed. The second problem concerned the intended goal at which the sequence of actions was being directed. For this, the measure r was proposed. The question begs to be asked: could not one unified measure be devised, so as to solve both problems? After all, both measure claim to capture a sense of intention.

In order to answer this question, we present a summary of the two problems and their respective measures of intention, side by side, in Table 6.1. This presentation will assist in the understanding of the critical differences between the two problems, and will explain the necessity of using two different measures of intention for solving them.

For the solution of the first problem, a high enough value of t (given some threshold) indicates intention. In essence, what this t measures, is the proportion of states which are more distant from the initial state, compared to the previous

	First Problem	Second Problem
Problem:	Determine whether the observed sequence is intentional or not	Choose from among several possible afforded goals, the one goal most likely intended by the performing agent
Input Arguments:	$s_0, \ldots, s_n$	$s_0, \ldots, s_n, g_{j=1,\ldots,k}$
Intention Measure:	$t = \frac{\lvert\{s_i : d_i > d_{i-1}\}_{i=1}^{n}\rvert}{n}$	$r_{j=1,\ldots,k} = \frac{dist(s_0, g_j)}{\sum_{i=1}^{n} dist(s_{i-1}, s_i) + dist(s_n, g_j)}$

Table 6.1: Summary of the Two Problems and Their Respective Measures of Intention.

state. That is, it captures the notion of rationality in the sense of consistency in moving away from the initial state. In the case of intention, it is expected that the actions will move the agent consistently farther and farther away from the initial state, towards its goal (unknown to the observer at this stage).

For the solution of the second problem, the goal g_j with the highest measure r_j will be chosen as the goal intended by the agent (with possible fine tuning according to other heuristics). This measure conveys how efficiently the observed sequence can reach the goal state, when the path is extended to the goal (denominator of r_j), compared to the most efficient way of reaching the goal from the initial state (numerator of r_j).

With this summarizing comparison at hand, several reasons for choosing measure t for the first problem, and measure r for the second, can be proposed, along with an explanation of why using one instead of the other will not produce satisfactory results.

6.1.1 Different Input Parameters

For the first problem, we do not know whether or not there is a goal state, let alone what that goal state is. Therefore, we cannot ask how efficiently the goal state can be reached (which is what the second measure deals with). The terminal state, at this stage, cannot be used as a possible goal state, since the sequence might be intentional, yet fail at bringing about the intended goal, in which case

measuring the efficiency of reaching the failed terminal state would be misleading. So, instead of looking forwards and asking whether the agent is proceeding efficiently towards some state (which is what we do for the second problem), we look backwards and ask whether the agent is consistently proceeding away from the initial state. This explains why the second measure should not be used for the first problem.

In essence, this difference arises from the simple fact that the two problems deal with different input. While the first problem takes only the path of observed motion as input, the second problem considers the possible afforded goals as well. Of course, it could be suggested to take these afforded goals into account already at the first stage, however, choosing among several possible goals before knowing yet whether or not any such goal in fact exists, would be conceptually wrong. This is so since in the case of unintentional action, returning as an answer a goal claimed to be the intended goal, would obviously be misleading.

6.1.2 Absolute vs. Relative Values

In addition, for the first problem, we have one observed sequence of actions, which must be analyzed with regard to itself only. This, as opposed to the case in the second problem, where we have several sequences—the observed sequence extended to each of the possible goal states. These sequences can be compared to each other, as far as their efficiency in bringing about their goals. In other words, in the first problem one absolute value is produced (the intention of the observed path of motion), while in the second problem several relative values are produced (one for each afforded goal state).

Consider for example the schematic depiction in Figure 6.1 of an observed path of motion, with three possible goals and their respective extensions of the path to them. According to the first measure, all three paths (from s_0 through s_n to each of g_1, g_2 and g_3) will receive the maximal value of 1, since at every point along the path the distance from s_0 increases, for all the goals. For the first problem, this would be considered a satisfactory result, since, indeed, all three paths display intention. Yet, for differentiating between the goals with regard to their likelihood of being the intended one, the first measure is not strong enough. However, according to the second measure, goal g_2 will be scored highest, since it is reached by the most efficient way from the initial state, as opposed to the

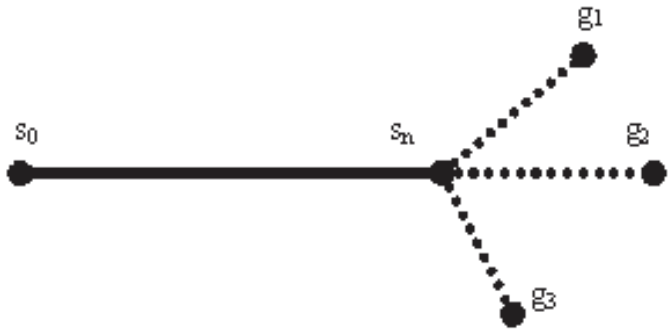


Figure 6.1: Schematic Path of Motion with Three Possible Goals.

two other goals, for which more efficient paths exist. This explains why the first measure cannot be used for the second problem.

6.1.3 Enabling Parsing of Subgoals

An added benefit of using the first measure for the first problem is the ability to parse the sequence of actions into subgoals, according to changes in the trend of the intention graph, as discussed above, in 4.2.2.2. This is enabled due to the fact that this measure incorporates information about the time—by dividing by the number of states in the sequence, which can be thought of as the number of time steps, given that one action is performed at each time step.

Consider for example the case of an agent starting out at s_0, moving towards location s_i, lingering there for a while, and then continuing to another location s_n. This is presented schematically in Figure 6.2. Applying the second measure here for the first problem, would indicate an intentional sequence (with the goal

Figure 6.2: Schematic Path of Motion with Agent Lingering at s_i.

of reaching s_n). Yet, applying the first measure would produce a graph of intention which rises towards s_i, then drops as time continues yet the distance from the initial state remains constant, then once again increases as the agent moves towards s_n. It is these changes of trend which would enable parsing this sequence into two subgoals, s_i and s_n.

It is inherent to the specification of the first problem that time be taken into account, while for the second problem this is not necessary—the prefix of the path (s_0 to s_n) is already determined to be intentional by the first stage, and remains fixed for each of the possible goals—to which the paths are extended in the optimal way, without wasting time, which means that time is irrelevant here.

6.2 The Role of Repetition

In Meltzoff [1995]'s original experiment, children were shown to be able to predict the intended goal in two conditions: when the goal was successfully achieved (Demonstration Target), and when the goal was attempted but failed (Demonstration Intention). According to these results, perceived intention is enough for predicting the goal. Follow-up studies by Meltzoff et al. [1999] have shown that one failed attempt demonstration was not enough to produce imitation by the observing children, as opposed to one successful demonstration, which was sufficient. According to this account, when dealing with failed goals, repetition is necessary for the process of intention recognition.

The question remains as to which of the two stages in the recognition process makes use of the information carried by repetition. When the failed intention was displayed only once, did the children fail to imitate since they did not detect intention (first stage of intention detection), or did they detect intention yet failed to imitate since they were not able to infer the intention from the insufficient observations (second stage of intention prediction)?

For each of these possibilities, the repetition can be explained as playing a specific role. If the repetition is relevant for the first stage of intention detection, it can be seen as playing an enhancement role, strengthening the measure of intention more and more upon each observation, finally driving it over the required threshold for determining the presence of intention. Such mechanisms are familiar in neurophysiology—single neurons, and consequently the output of neural networks, have thresholds which cause the system to fire only when the stimulus is strong enough. This also fits with the intentionality theory of Heider [1958], which points at persistence as one of the main characteristics of intentional action. Our measure for intention detection returns a value in the continuous range of [0,1], which can be binarized by a threshold, as described in Section 3.2.

If, on the other hand, the repetition is relevant for the second stage of intention prediction, it fits into the role that equifinality plays in inferring intended goals. The principle of equifinality, as mentioned in Section 2.2.2, states that when attempting to realize an intended goal, the actor will vary his actions, depending on the environmental constraints. Indeed, in Meltzoff [1995]'s original experiment, the three failed attempts were usually not exact copies of each other, but rather differed in such aspects as the initial state or the path taken to reach

the goal. As such, each demonstration could serve as an additional clue in the quest of intention prediction, while only one clue might not be sufficient.

Other follow-up studies have been performed in an attempt to clarify various aspects of the process of intention recognition, among them, the role of repetition. Huang et al. [2002] came up with a seemingly contradicting finding. In their study, they showed children a stimulus enhancement demonstration, in which the adult performed a sequence of actions which called the observer's attention to those parts of the objects relevant to the goal, by bringing them close to each other. This has been proposed as one of the possible clues for predicting intention ("spatial contiguity", Section 2.4). According to the description of the action sequences, to the viewer, the Stimulus Enhancement condition would look the same as the Demonstration Intention condition. Yet, in Huang et al. [2002]'s experiment the results showed that one Stimulus Enhancement demonstration was enough for correctly predicting intention, as opposed to Meltzoff [1995]'s one failed attempt demonstration, which did produce imitation by the observing children. How can this discrepancy be explained?

Huang (personal communication) suggests the it might not be the repetition itself which plays a role here, but rather the exposure time. While all three Demonstration Intention sequences took 20 seconds to display altogether, a single one took only a third of that. On the other hand, a single Stimulus Enhancement demonstration took the full 20 seconds to display (including the initial state and restoration to the initial state at the end of the demonstration). Therefore, the failure of the children to reproduce the target act after one failed attempt in Meltzoff's experiment might be due to insufficient exposure time. If this is so, it can be concluded that repetition does not truly play a role in the stage of intention prediction.

In our experiments, in the second stage of intention prediction, we did not make available the information carried by repetition. Rather, the viewers were expected to predict the intention given only one observed sequence of actions. According to the above, for the second stage this could be justified by the fact that repetition is not relevant for intention prediction. Indeed, the results support this.

For the first stage of intention detection, there were two different experiments. In the second experiment using surveillance videos, the viewers were

allowed to play the video clips as often as they wanted, and this was not controlled or monitored. However, in the first experiment which simulated Meltzoff [1995]'s, the output was given as a single value, which was high enough to cross the predetermined threshold after only one "viewing". In order to incorporate the mechanism of repetition, the threshold could simply be made higher, and an accumulating value could be used, such that only three repeated attempts would drive the value over the threshold.

6.3 Determining the Point of Failure

According to our model, once failure has been determined (at the stage of Success Detection, Section 3.4), the process of Intention Prediction kicks in. For this, the observed sequence of actions is extended to each of the possible afforded goals, and each of these is compared to the optimal sequence, from the initial state to the respective goal. This was described in detail above (Section 3.5). It is worth noting that the process can be refined if the action sequence is not extended from the (failed) end-state, but rather from the point at which failure commenced.

How can the point at which failure commenced be identified? Once again, the Principle of Rational Action, as it is captured by our measure for intention detection, can be utilized. The measure of intention can be calculated for every state along the trace of observed action, and the resulting behavior of the resulting graph can be analyzed. A noticeable point at which the graph significantly dips, towards the end of the trace (assuming the action was halted close to where the failing began), conveys a meaningful drop in intention, and can be taken as a breakpoint at which failure commenced. Calculating the measure of intention detection through this breakpoint, instead of through the observed end-state, would result in a more accurate hypothesis regarding the intended goal.

6.4 False Beliefs and Environmental Constraints

In this work we assumed there were no environmental or psychological constraints which had to be taken into account. Environmental constraints could be, for example, physical obstacles. Dealing with these can easily be incorporated

into our model: the distance function used by the measures of intentionality must simply be adapted so that it captures the information regarding obstacles. Thus, for example, when using the Euclidean distance, instead of measuring the direct distance between two points, the distance would be measured by a path which circumvents the obstacle in the most direct way possible.

By psychological constraints we are referring to the problem of false beliefs. As mentioned in Section 2.2.2, the Principle of Rational Action on which our measure of intentionality were based, stems from Gergely and Csibra [2003]'s teleological stance. This stance would not necessarily be able to deal with interpretation of actions which is based on false beliefs. It would be interesting to attempt to expand our model to include such cases, and observe if and how the model would then be able to handle them.

6.5 Summary

In this work we have presented a cognitive model of human intention recognition. Its main contribution is meant to be, firstly, in the explanation of the process as a whole and the interaction between the modules composing it. We have tried to justify this with reference to the large body of research which has accumulated on the topic of intention in the field of psychology.

Secondly, we elaborated on two of the core modules, those of intention detection and intention prediction, describing a way to translate psychological principles, such as the Principle of Rational Action, affordances, and stimulus enhancement by spatial contiguity, into measures and concepts which can be computationally implemented. These translations were evaluated in comparison to human judgment of intention, proving their validity and utility at solving the task at hand.

To summarize, the contributions of this dissertation are:

- A proposal of a comprehensive model relating all the necessary components which play a part in the process of intention recognition.

- Introduction of measures of intention which are used for detecting the presence of intention in a sequence of observed actions, and predicting their intended outcome.

- Devising experimental methods for testing these measures of intention, and comparing their usefulness at the task at hand to human performance.

This research can be taken forward on several fronts. The model as is can be implemented in a software of hardware agent, along with other cognitive and social skills, and its performance evaluated. At the same time, the model can be expanded to deal with false beliefs and pretense, as well as static and dynamic environmental constraints. These are only a few of the different directions future research can follow, as elaborated above.

While there is still some way to go in order to render the ideas presented here into a full working implementation, we believe this work greatly advances the current understanding of the process of intention recognition. As such, we hope it will be of interest and of use to researchers in the multidisciplinary communities dealing with intention recognition.

Bibliography

E.L. Altschuler, A. Vankov, V. Wang, V.S. Ramachandran, and Pineda J.A. Person see, person do: Human cortical electrophysiological correlates of monkey see monkey do cell. *Society for Neuroscience Abstracts*, 1997. [cited at p. 15]

M.A. Arbib, A. Billard, M. Iacoboni, and E. Oztop. Synthetic brain imaging: Grasping, mirror neurons and imitation. *Neural Networks*, 13:975–997, 2000. [cited at p. 15]

D. Avrahami-Zilberbrand and G. Kaminka. Towards dynamic tracking of multi-agents teams: An initial report. In *Proceedings of Workshop on Plan, Activity, and Intent Recognition*, 2007. [cited at p. 17]

E.R. Babbie. *The Practice of Social Research*. Wadsworth Publishing, 2003. [cited at p. 55]

M. Baldoni, G. Boella, and L. van der Torre. Modelling the interaction between objects: Roles as affordances. In J. Lang, F. Lin, and J. Wang, editors, *Knowledge Science, Engineering and Management*, volume 4092 of *Lecture Notes in Computer Science*, pages 42–54. Springer Berlin / Heidelberg, 2006. [cited at p. 23]

M.P. Banchetti-Robino. Ibn sina and husserl on intention and intentionality. *Philosophy East and West*, 54(1):71–82, 2004. [cited at p. 10]

S. Baron-Cohen. *Mindblindness: an Essay on Autism and Theory of Mind*. MIT Press, 1995. [cited at p. 8]

H. Bekkering, A. Wohlschlager, and M. Gattis. Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology*, 53(1):153–164, 2000. [cited at p. 14, 15]

F. Bellagamba and M. Tomasello. Reenacting intended acts: Comparing 12- and 18-month-olds. *Infant Behavior and Development*, 1999. [cited at p. 33]

I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 1987. [cited at p. 26]

A. Billard and M.J. Mataric. Learning human arm movements by imitation:: Evaluation of a biologically inspired connectionist architecture. *Robotics and Autonomous Systems*, 37:145–160, 2001. [cited at p. 18]

S.J. Blakemore and J. Decety. From the perception of cction to the understanding on intention. *Nature Reviews, Neuroscience*, 2001. [cited at p. 10, 12]

N. Blaylock and J. Allen. Hierarchical instantiated goal recognition. In *AAAI Workshop on Modeling Others from Observations (MOO-2006)*, 2006. [cited at p. 16]

B. Bonet and H. Geffner. Planning as heuristic search: new results. In S. Biundo and M. Fox, editors, *Procceedings of the 5th European conference on planning*, Durham, UK, 1999. Springer: Lecture Notes on Computer Science. [cited at p. 31, 40]

A.E. Booth and S. Waxman. Object names and object functions serve as cues to categorization for infants. *Developmental Psychology*, 38(6):948–957, 2002. [cited at p. 24]

A.C. Brandone. *The Development of Intention Understanding in the First Year of Life: An Exploration of Infants' Understanding of Successful vs. Failed Inentional Actions*. PhD thesis, University of Michigan, 2010. [cited at p. 18]

A.C. Brandone and H.M. Wellman. You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science*, 20(1):85–91, 2009. [cited at p. 18]

C. Breazeal. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59:119–155, 2003. [cited at p. 18]

C. Breazeal and B. Scassellati. Robots that imitate humans. *Trends in Cognitive Sciences*, 6(11):481–487, 2002. [cited at p. 18]

C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, 11:31–62, 2005. [cited at p. 15]

S. Calinon and A. Billard. Learning of gestures by imitation in a humanoid robot. In K. Dautenhahn and C. Nehaniv, editors, *Imitation and Social Learning in Robots, Humans and Animals: Behavioral, Social and Communicative Dimensions*. Cambridge University Press, 2007. [cited at p. 18]

A.J. Carona, R.F. Carona, and S.E. Antell. Infant understanding of containment: An affordance perceived or a relationship conceived? *Developmental Psychology*, 24(5): 620–627, 1988. [cited at p. 23]

M. Carpenter, N. Akhtar, and M. Tomasello. Fourteen- through 18-month old infants differentially imitate intentional and accidental actions. *Infant behavior development*, 1998. [cited at p. 12]

M. Casasola and L.B. Cohen. Infant categorization of containment, support and tight-fit spatial relationships. *Developmental Science*, 5(2):247–264, 2002. [cited at p. 23]

E. Charniak and R.P. Goldman. A bayesian model of plan recognition. *Artificial Intelligence*, 64:53–79, 1993. [cited at p. 16]

E. Chiarello, M. Casasola, and L.B. Cohen. Six-month-old infants' categorization of containment spatial relations. *Child Development*, 72(3):679–693, 2003. [cited at p. 23]

P. Cisek. Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society, Biological Science*, 362(1485): 1585–1599, 2007. [cited at p. 22, 25, 26]

S. Cochin, C. Barthelemy, B. Lejeune, S. Roux, and Martineau. J. Perception of motion and qeeg activity in human adults. *Electroencephalography Clinical Neurophysiology*, 107:287–295, 1998. [cited at p. 15]

S. Cochin, C. Barthelemy, S. Roux, and Martineau. J. Observation and execution of movement: Similarities demonstrated by quantified electroencephalograpy. *European Journal of Neuroscience*, 11:1839–1842, 1999. [cited at p. 15]

L. Craighero, L. Fadiga, C.A. Umilta, and G. Rizzolatti. Evidence for visuomotor priming effect. *NeuroReport*, 8(1):347–349, 1996. [cited at p. 24]

G. Csibra and G. Gergely. The teleological origins of mentalistic action explanations: a developmental hypothesis. *Developmental science*, 1998. [cited at p. 11]

G. Csibra and G. Gergely. Social learning and social cognition: The case for pedagogy. In M.H. Johnson and Y.M. Munakata, editors, *Processes of Change in Brain and Cognitive Development: Attention and Performance*, pages 249–274. Oxford University Press, 2006. [cited at p. 33]

R.H. Cuijpers, H.T. van Schie, M. Koppen, W. Erlhagen, and H. Bekkering. Goals and means in actions observation: A computational approach. *Neural Networks*, 19:311–322, 2006. [cited at p. 19]

M. Dapretto, M.S. Davies, J.H. Pfeifer, A.A. Scott, M. Sigman, S.Y. Bookheimer, and M. Iacoboni. Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, 9:28–30, 2005. [cited at p. 16]

M. Davies and T. Stone. *Mental Simulation: Evaluations and Applications*. Blackwell Publishers, 1995. [cited at p. 14]

D.C. Dennet. *The Intentional Stance*. MIT press, 1989. [cited at p. 10]

M.R. Dogar, M Cakmak, E. Ugur, and E. Sahin. From primitive actions to goal-directed behavior using a formalization of affordances for robot control and learning. Technical report, Department of Computer Engineering Middle East Technical University, Turkey, 2007. [cited at p. 26]

E. Erdemir, C.B. Frankel, S. Thornton, B. Ulutas, and K. Kawamura. A robot rehearses internally and learns an affordance relation. In *Proceedings of IEEE 7th International Conference on Development and Learning*, 2008. [cited at p. 26]

L. Fadiga, L. Fogassi, G. Pavesi, and Rizzolatti. G. Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, 73:2609–2611, 1995. [cited at p. 15]

P. Fitzpatrick, G. Metta, L. Natale, S. Rao, and G. Sandini. Learning about objects through action—initial steps towards artificial cognition. In *Proceedings of 2003 IEEE International Conference on Robotics and Automation*, 2003. [cited at p. 26]

L. Fogassi, P.F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, and G. Rizzolatti. Parietal lobe: From action organization to intention understanding. *Science*, 308:662–667, 2005. [cited at p. 16, 25]

P.H. Foo, G.W. Ng, K.H. Ng, and R. Yang. Application of intent inference for surveillance and conformance monitoring to aid human cognition. In *Proceedings of the 10th International Conference on Information Fusion*, 2007. [cited at p. 8]

V. Gallese and A. Goldman. Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Science*, 2(12):493–501, 1998. [cited at p. 16]

V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. Action recognition in the premotor cortex. *Brain*, 119(2):593–609, 1996. [cited at p. 16]

W.G. Gaver. Technology affordances. In *CHI '91 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology*, 1991. [cited at p. 22]

C. Geib and R.P. Goldman. Partial observability and probabilistic plan/goal recognition. In *Proceedings of the IJCAI-05 Workshop on Modeling Others From Observations*, 2005. [cited at p. 16]

G. Gergely and G. Csibra. Teleological reasoning in infancy: the naive theory of rational action. *TRENDS in Cognitive Science*, 7(7), 2003. [cited at p. ii, 10, 11, 30, 36, 93]

G. Gergely, Z. Nasady, G. Csibra, and S. Biro. Taking the intentional stance at 12 months of age. *Cognition*, 56:165–193, 1995. [cited at p. 8]

G. Gergely, Bekkering H., and I. Kiraly. Rational imitation in preverbal infants. *Nature*, 415:755, 2002. [cited at p. 14]

J.J. Gibson. *The Theory of Affordances*, pages 67–82. Lawrence Erlbaum, Hillsdale, NJ, 1977. [cited at p. 22]

R. Gordon. Folk psychology as simulation. *Mind and Language*, 1(2):158–171, 1986. [cited at p. 14]

J. Grezes and J. Decety. Does visual perception of object afford action? evidence from a neuroimaging study. *Neuropsychologia*, 40(2):212–222, 2002. [cited at p. 25]

J. Grezes, M. Tucker, J. Armony, R. Ellis, and R.E. Passingham. Objects automatically potentiate action: an fmri study of implicit processing. *European Journal of Neuroscience*, 17(12):2735–2740, 2003. [cited at p. 22, 25]

J.J. Guajardo and A.L. Woodward. Is agency skin deep? surface attributes influence infants' sensitivity to goal-directed action. *Infancy*, 6(3):361–384, 2004. [cited at p. 10]

E. Hanna and A.N. Meltzoff. Peer imitation by toddlers in laboratory, home, day-care contexts: Implications for social learning and memory. *Developmental Psychology*, 29:701–710, 1993. [cited at p. 13]

R. Hari, N. Forss, S. Avikainen, S. Kirveskari, S. Salenius, and G. Rizzolatti. Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proceedings of the National Academy of Sciences of the USA*, 95(25):15061–15065, 1998. [cited at p. 15]

S. Hart. An intrinsic reward for affordance exploration. In *Proceedings of IEEE 8th International Conference on Development and Learning*, 2009. [cited at p. 26]

K. Harui, N. Oka, and Y. Yamada. Distinguishing intentional actions from accidental actions. In *Proceedings of the 4th IEEE international conference on development and learning*, 2005. [cited at p. 9]

J. Heal. Mind, reason and imagination. In *Understanding Other Minds from the Inside*, pages 28–44. Cambridge University Press, 2003. [cited at p. 14]

F. Heider. *The psychology of interpersonal relationships*. Wiley, 1958. [cited at p. 12, 90]

T. Hofer, P. Hauf, and G. Aschersleben. Infant's perception of goal-directed actions performed by a mechanical device. *Infant Behavior and Development*, 28(4):466–480, 2005. [cited at p. 29]

J. Hong. Goal recognition through goal graph analysis. *Journal of Artificial Intelligence Research*, 15(1):1–30, 2001. [cited at p. 16]

S. Hongeng and J. Wyatt. Learning causality and intentional actions. In E. Rome, J. Hertzberg, and G. Dorffner, editors, *Towards affordance-based robot control*. Springer, 2008. [cited at p. 17]

J.S. Horst, L.M. Oakes, and K.L. Madole. What does it look like and what can it do? category structure influences how infants categorize. *Child Development*, 76(3):614–631, 2005. [cited at p. 24]

C.T. Huang, C. Heyes, and T. Charman. Infants behavioral reenactment of failed attempts: Exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology*, 38(5):840–855, 2002. [cited at p. 21, 45, 91]

S. Itakura, H. Ishida, T. Kanda, Y. Shimada, H. Ishiguro, and K. Lee. How to build an intentional android: Infants' imitation of a robot's goal-directed actions. *Infancy*, 13 (5), September 2008. [cited at p. 10, 29]

I. Kiraly, B. Jovanovic, W. Prinz, G. Aschersleben, and G. Gergely. The early origins of goal attribution in infancy. *Conciousness and cognition*, 12(4):752–769, 2003. [cited at p. 11]

N. Lesh and O. Etzioni. A sound and fast goal recognizer. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1704–1710, 1995. [cited at p. 16]

M. Lopes, F.S. Melo, and L. Montesano. Affordance-based imitation learning in robots. In *Proceedings of 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007. [cited at p. 18, 26]

K.L. Madole and L.B. Cohen. The role of object parts in infants attention to form function correlations. *Developmental Psychology*, 31(4):637–648, 1995. [cited at p. 24]

K.L. Madole, L.M. Oakes, and L.B. Cohen. Developmental changes in infants attention to function and formfunction correlations. *Cognitive Development*, 8(2):189–209, 1993. [cited at p. 24]

A.N. Meltzoff. Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*, 24:470–476, 1988. [cited at p. 13]

A.N. Meltzoff. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31(5), 1995. [cited at p. i, ii, 2, 10, 12, 13, 14, 19, 20, 29, 33, 39, 90, 91, 92]

A.N. Meltzoff. Imitation of televised models by infants. *Child Development*, 59:1221–1229, 1998. [cited at p. 13]

A.N. Meltzoff. Imitation as a mechanism of social cognition: origins of empathy, theory of mind, and representation of action. In U. Goswami, editor, *Blackwells Handbook of Childhood Cognitive Development*. Blackwell, 2002. [cited at p. 11]

A.N. Meltzoff. The "like-me" framework for recognizing and becoming an intentional agent. *Acta Psychologica*, 124(1):26–43, 2007. [cited at p. 15]

A.N. Meltzoff and J. Decety. What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London*, 2003. [cited at p. 15, 19]

A.N. Meltzoff and A. Gopnik. The role of imitation in understanding persons and developing a theory of mind. In H. Baron-Cohen, S. Tager-Flusberg and D.J. Cohen, editors, *Understanding Other Minds*, pages 335–366. Oxford University Press, 1993. [cited at p. 15]

A.N. Meltzoff and M.K. Moore. Imitation in newborn infants: Exploring the range of gestures imitated and underlying mechanisms. *Developmental Psychology*, 25:954–962, 1989. [cited at p. 13]

A.N. Meltzoff and M.K. Moore. Early imitation within a functional framework: The importance of person identity, movement and development. *Infant Behavior and Development*, 15:479–505, 1992. [cited at p. 15]

A.N. Meltzoff and M.K. Moore. Imitation, memory and the representation of persons. *Infant Behavior and Development*, 17:83–99, 1994. [cited at p. 15]

A.N. Meltzoff and M.K. Moore. Infants' understanding of people and things: From body imitation to folk psychology. In J.L. Bermdez, A. Maarcel, and N. Eilan, editors, *The Body and the Self*, pages 43–69. The MIT Press, 1995. [cited at p. 15]

A.N. Meltzoff and M.K. Moore. Explaining facial imitation: A theoretical model. *Early Development and Parenting*, 6:179–192, 1997. [cited at p. 13]

A.N. Meltzoff, A. Gopnik, and B.M. Repacholi. Toddlers' understanding of intentions, desires and emotions: exploration of the dark ages. In P.D. Zelazo, J.W. Astington, and D.R. Olson, editors, *Developing theories on intention: social understanding and self control*. Lawrence Erlbaum Associates, 1999. [cited at p. 19, 41, 90]

A.N. Metlzoff and M.K. Moore. Newborn infants imitate adult facial gestures. *Child Development*, 54:702–709, 1983. [cited at p. 13]

J.V. Miro, V. Osswald, M. Patel, and G. Dissanayake. Robotic assistance with attitude: A mobility agent for motor function rehabilitation and ambulation support. In *Proceedings of the IEEE International Conference on Rehabilitation Robotics*, 2009. [cited at p. 8]

R.R. Murphy. Case studies of applying gibsons ecological approach to mobile robots. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(1), 1999. [cited at p. 26]

C.L. Nehaniv and K. Dautenhahn, editors. *Imitation and Social Learning in Robots, Humans, and Animals: Behavioural, Social and Communicative Dimensions*. Cambridge University Press, New York, 2007. [cited at p. 19]

M. Nielsen. 12-month-olds produce others intended but unfulfilled acts. *Infancy*, 14(3): 377–389, 2009. [cited at p. 33]

D. Oztop, E.and Wolpert and M. Kawato. Mental state inference using visual control parameters. *Cognitive Brain Research*, 22:129–151, 2005. [cited at p. 19]

E. Oztop, M. Kawato, and M. Arbib. Mirror neurons and imitation: A computationally guided review. *Neural Networks*, 19(3):254–271, 2006. [cited at p. 19]

A. Pinz. Object categorization. *Foundations and Trends in Computer Graphics and Vision*, 1(4):255–353, 2005. [cited at p. 26]

D.G. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1, 1978. [cited at p. 14]

P.C. Quinn. The categorization of above and below spatial relations by young infants. *Child Development*, 65(1):58–69, 1994. [cited at p. 23]

M Ramirez and H. Geffner. Probabilistic plan recognition using off-the-dhelf classical planners. In *Proceedings of the 10th AAAI Conference on Artificial Intelligence*, 2010. [cited at p. 17]

R.P.N. Rao, A.P. Shon, and A.N. Meltzoff. A bayesian model of imitation in infants and robots. In C.L. Nehaniv and K. Dautenhahn, editors, *Imitation and social learning in robots, humans, and animals: behavioural, social and communicative dimensions*. Cambridge University Press, 2007. [cited at p. 19]

B. Ridge, D. Skocaj, and A. Leonardis. Unsupervised learning of basic object affordances from object properties. In *Computer Vision Winter Workshop*, 2009. [cited at p. 26]

E. Rivlin, S.J. Dickinson, and A. Rosenfeld. Object recognition by functional parts. In *Image Understanding Workshop*, 1994. [cited at p. 26]

G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131–141, 1996. [cited at p. 16]

G. Rizzolatti, L. Fogassi, and V. Gallese. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews in Neuroscience*, 2001. [cited at p. 16]

E. Sahin, M. Cakmak, M.R. Dogar, Ugur E., and G. Ucoluk. To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior December*, 15(4):447–472, 2007. [cited at p. 22]

S. Schaal, A.J. Ijspeert, and A. Billard. Computational approaches to motor learning by imitation. *Philosophical Transaction of the Royal Society of London*, pages 537–547, 2003. [cited at p. 18]

M.M. Sitskoorn and A.W. Smitsman. Infants' perception of dynamic relations between objects: Passing through or support? *Developmental Psychology*, 31(3):437–447, 1995. [cited at p. 23]

M.C. Smith. Cognizing the behavior stream: the recognition of intentional action. *Child development*, 1978. [cited at p. 10]

R. St. Amant. User interface affordances in a planning representation. *Human Computer Interaction*, 14(3):317–354, 1999. [cited at p. 22]

A. Stoytchev. Toward learning the binding affordances of objects: A behavior-grounded approach. In *Proceedings of 2005 AAAI Symposium on Developmental Robotics*, 2005. [cited at p. 26]

M. Tucker and R. Ellis. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3):830–846, 1998. [cited at p. 24]

J.S. Watson. The elementary nature of purposive behavior: Evolving minimal neural structures that display intrinsic intentionality. *Evolutionary Psychology*, 3:24–48, 2005. [cited at p. 11, 30]

J.H.G. Williams, A. Whiten, T. Suddendorf, and D.I. Perrett. Imitation, mirror neurons and autism. *Neuroscience and Biobehavioral Reviews*, 25:287–295, 2001. [cited at p. 16]

A.L. Woodward. Infants selectively encode the goal object of an actors reach. *Cognition*, 69:1–34, 1998. [cited at p. 8]

A.L. Woodward, J.A. Sommerville, and J.J. Guajardo. How infants make sense of intentional action. In B.F. Malle and L.J. Moses, editors, *Intentions and Intentionality: Foundations of Social Cognition*, pages 149–169. MIT Press, 2001. [cited at p. 29]

L. Ye, W. Cardwell, and L.S. Mark. Perceiving multiple affordances for objects. *Ecological Psychology*, 21(3):185–217, 2009. [cited at p. 26]