Towards Computational Models of Intention Detection and Intention Prediction

Elisheva Bonchek-Dokow[a,*], Gal A. Kaminka[b,a]

[a]*Brain Science Research Center*
Bar Ilan University, Israel
[b]*MAVERICK Group*
Computer Science Department
Bar Ilan University, Israel

Abstract

Intention recognition is one of the core components of mindreading, an important process in social cognition. Human beings, from age of 18 months, have been shown to be able to extrapolate intentions from observed actions, even when the performer failed at achieving the goal. Existing accounts of intention recognition emphasize the use of an intent (plan) library, which is matched against observed actions for recognition. These therefore cannot account for recognition of failed sequences of actions, nor novel actions. In this paper, we begin to tackle these open questions by examining computational models for components of human intention recognition, which emphasize the ability of humans to detect and identify intentions in a sequence of observed actions, based solely on the rationality of movement (its efficiency). We provide a high-level overview of intention recognition as a whole, and then elaborate on two components of the 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 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. We evaluate the model, and these two components, in context of existing literature, and in a number of experiments with more than 140 human subjects. For intention detection, our model

[*] Corresponding author

Email addresses: `le7.bonchek.dokow@gmail.com` (Elisheva Bonchek-Dokow), `galk@cs.biu.ac.il` (Gal A. Kaminka)

was able to attribute high levels of intention to those traces perceived by humans as intentional, and vice versa. For intention prediction as well, our model performed in a way that closely matched that of humans. The work highlights the intimate relationship between the ability to generate plans, and the ability to recognize intentions.

Keywords: Intention Recognition, Intention Prediction, Cognitive Modeling

1. Introduction

Intention recognition is one of the core processes of *mindreading*, an important component in social cognition. Intention recognition involves identifying the goal of an observed sequence of actions, performed by some acting agent. It is a process by which an agent can gain access to the goals of another, and predict its future actions and trajectories. While it is not sufficient, by itself, for full mental state attribution (e.g., it does not ascribe beliefs to the observed agent), it is of critical importance in social interaction, and is of obvious evolutionary benefit. Indeed, human beings, from age of 18 months, have been shown to be able to extrapolate intentions from observed actions, even when the performer failed at achieving the goal (Meltzoff, 1995).

Existing accounts of intention recognition in artificial intelligence (*plan recognition*) and machine vision (*activity recognition*) emphasize the use of an intent (plan, activity) library, which is matched against observed actions for recognition. These therefore cannot account for recognition of failed sequences actions, nor novel actions. Moreover, these accounts ignore cognitive science literature, which shows that the process involves a number of component processes: recognizing the agent as capable of possessing intentions, recognizing that the observed action sequence is intentional, hypothesizing the intent of the sequence (even if the sequence results in a failing), and more (we discuss this in Section 2).

In this paper, we focus on modeling two of these components, *intention detection* and *intention prediction.* By *intention detection* we mean the ability to discern whether 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.

In particular, we focus on the use of rationality (efficiency) of an observed action trajectory or plan as a possible basis for intention recognition. We argue that while there are several different ways in which humans may carry out intention detection and prediction (discussed in detail in Section 2), it is often possible to determine a level of intentionality of an observed sequence of actions, based solely on the observed actions, and the ability to plan (optimally). We thus highlight a role for planning within recognition.

Following Section 2, which motivates our work in context of existing literature, we begin in Section 3 with a brief description of an abstract intention recognition model, in which two of the components are intention detection and intention prediction (we describe this abstract model to put the component processes in context of the larger area of research). We then provide a detailed account of the computational models underlying these processes from our perspective, which focuses on rationality. In Sections 4 and 5 we evaluate the hypothesized models for *intention detection* and *intention recognition* processes, respectively.

In particular, in Section 4, the intention detection model was evaluated in a discrete-state recreation of key experiments in humans (Meltzoff, 1995), and in detecting intentionality in activity recognition videos. In both settings for evaluating the first component of intention detection, the results confirm that our model closely matches human performance. Traces of action that were deemed by human observers as highly intentional, were ranked similarly by our model, while traces of action that were judged by humans as less intentional, achieved lower grades of intention by our model as well. Thus, the predictions of the model were successfully compared to those of human subjects. In addition, our findings show that our model proves useful for detecting sub-goals as well.

In a final set of experiments (Section 5), the intention prediction component was evaluated with data from human subjects, manipulating two-dimensional objects in a computer-based experiment. These experiments show equally promising results. Our model was able to predict the correct intention of various action traces with high accuracy, using our suggested heuristic. Two other heuristics are evaluated as well, and show significantly inferior prediction ability.

Finally, in Section 6, we discuss the significance of the results, highlighting several aspects such as the use of different measures of intentionality and the role of the intention detection method we propose, along-side and complementary to other methods in intention recognition. We also suggest possible directions for future research.

2. Background and Related Work

First introduced by Premack & Woodruff (1978), *theory of mind* (also: *folk psychology*, *mentalizing*, and *mindreading*) 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. This ability enables a psychological attribution of causality to human acts, rather than the physical causality generally attributed to inanimate objects (Meltzoff, 1995).

Different accounts are given by psychologists for the mechanism underlying this ability. One of them, known as *simulation theory* (Gordon, 1986; Davies & Stone, 1995; Heal, 2003), has gained popularity and credibility lately, in part due to the discovery of mirror neurons (Gallese et al., 1996; Gallese & Goldman, 1998; Fogassi et al., 2005; Dapretto et al., 2005). In the words of Breazeal et al. (2005), simulation theory posits that by simulating another person's actions and the stimuli the other is experiencing using their own behavioral and stimulus processing mechanisms, humans can make predictions about the behaviors and mental states of the other 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 head of the other.

Thus theory of mind is intimately related to imitation, in subtle ways (Meltzoff & Moore, 1992, 1994, 1995; Meltzoff & Decety, 2003; Meltzoff & 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 her body, along with its inputs and outputs, is similar to those of the adults she sees around her, then she can simulate their behavior within her 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 paper is motivated by one specific line of investigations on the relation between theory of mind and imitation, that begins with an experiment by Meltzoff (1995). The experiment makes use of infants' tendency to imitate, to explore their mindreading capabilities, and specifically their ability to recognize the intent behind an observed sequence of actions. We discuss this experiment in Section 2.1. We then step back to discuss various perspectives on intentional actions, and define a scope for the definition used in this paper (Section 2.2). Next, we survey literature pertaining to the two main contributions in this work, *intention detection* (Section 2.3), and *intention prediction* (Section 2.4). We close with a discussion of *affordances*, a notion which we (and others) build on heavily in intention prediction (Section 2.5).

2.1. Meltzoff's Experiment

We elaborate here on a description of Meltzoff's (1995) 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; in general, 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. This control group did not show significant success at reproducing the target act.

- 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. This control group, too, 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? Meltzoff's (1995) experiment 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. Indeed, 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).

Motivated by this insight, this paper addresses *intention detection*, separately from *intention prediction*. Intention detection is the problem of *detecting* whether the observed sequence is intentional (i.e., has a specific goal). We discuss previous work related to it in Section 2.3. Intention prediction deals with the challenge of identifying (predicting) the specific goal of a sequence of actions. We discuss previous work related to it in Section 2.4.

However, before discussing these two main areas of contribution, we qualify the scope of our work. Meltzoff's experiments focused on intentions conveyed in sequences of actions. We follow in this, and restrict ourselves to such observations. Section 2.2 below discusses of how such intentional actions fit within a larger understanding of intentions.

2.2. Intentional Actions

Meltzoff's experiments examined human responses to stimuli involving an observed sequence of actions, which were designed to achieve (or fail to achieve) some specific goal. The stimuli differed not only in their outcome (success or failure to achieve the goal), but also in the sequence of actions leading towards it.

Indeed, throughout this work, the term *intentional action* refers to actions which are performed by an agent with the purpose of bringing about some desired final state. Three keywords should be emphasized here: *action*, *purpose*, and *final state*. Each such keyword distinguishes our definition of intentional action from alternative 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.

In particular, we use the term *actions* as causing a change in the world, observable to the agent that is attempting to recognize intentions. The term *final state* refers to the world state at the end of the observed sequence. The term *purpose* relates to the desired final state, whether achieved or not.

Thus *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 intention-less. 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.

In contrast, Harui et al. (2005) utilize a different definition. They too aim to distinguish between intentional and unintentional action. However, 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.

Thus observing a person reaching for a cup of coffee, and causing it to spill (possibly uttering "ooops!"), we would classify the action sequence as intentional (possessing intention), but failing to achieve it. Harui et al. would classify it as unintentional.

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 & 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 & 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.

There is also philosophical opinion, which views all conscious action as intentional (e.g., 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, rather than some end-state of the motion.

Consider, for example, a dancer: there is no end-state which the dancer is aiming to bring about, rather, the motion itself—the particular sequence of actions—is the goal. Likewise, waving good-bye is another example. Here we do not care about the ending position of the hand at the end of the wave, but rather for the repeated left and right motion of an open hand. The same motion with a closed fist, would not signify the same intent, nor would a forward-backward motion with an open hand. In the context of this work, we exclude such notion of intentional actions, limiting ourselves to motions or plans which are carried out to bring about a certain goal end-state.

Indeed, this paper is restricted to handling detecting and predicting intentions relating to achievement goals (some desired final state), rather than detecting and predicting maintenance goals, where actions are taken to maintain some world state over time, rather than to bring it about once. Thus for instance, observing one person following another at a fixed distance, the techniques we describe here may or may not detect intention in the follower, depending the actions of the person being followed. If the latter moves purposefully, then so may the follower, and thus our intention detection technique may detect an intention in the movement. However, our intention prediction techniques will certainly not be able to identify "maintain distance" as the intention of the follower.

Finally, 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, but in the state of mind of an agent. As such, they require different tools and mechanisms, and are out of the scope of this work.

2.3. Detecting Intentionality

Within the scope of intentional actions, as described above, we posit that the actions themselves can be inspected for underlying intention.

In particular, we build heavily on the *Principle of Rational Action* (Gergely & Csibra, 2003; Watson, 2005). 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. We hypothesize that detecting such rationality in the choice of actions is a sufficient condition for declaring the observed actions *intentional*.

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.

There are many other factors, aside from the Principle of Rational Action, which play a role in determining intentionality. These can be used as a basis for detecting intentionality in a sequence of actions (the scope of this paper), and also in other forms of intentional actions not addressed here.

A complete model of human intention recognition would of course take into account all other factors as well.

Some of the other factors in determining intentionality include: 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 & Decety, 2001), persistence in repeating actions over and over, or reaching the same final state again and again in multiple demonstrations (Meltzoff et al., 1999; Huang et al., 2002), duration of observing the action (Marhasev et al., 2009; Huang et al., 2002), and expending of effort (Heider, 1958). In our proposed model we focus on the Principle of Rational Action and ignore for now other features, as we attempt to isolate sufficient conditions for detecting intentionality within the scope discussed above (sequences of actions for achieving a goal). We are inspired in this pursuit by research that 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 in some cases.

And yet understanding the role of different factors in detecting intentionality is very challenging, as they affect each other. For instance, several psychological experiments, by Meltzoff and colleagues (Meltzoff, 1995; Meltzoff et al., 1999) and others (Huang et al., 2002), have made use of repetition in their demonstrations of intention. That is, the sequences of action were shown not once, but several times, in order to facilitate the correct guessing of the actor's intention. These studies suggest that perhaps, in certain situations (e.g. when the actions result in failure), repetition is at least useful, for guessing the correct intention.

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. However, it has been suggested (Huang, personal communication) that it is not the repetition per se which plays a role here, but rather the longer exposure time to the stimulus, which the repetition allows for. Another possibility is that each repetition serves to strengthen the certainty of the inferred intention, so as to drive the value over some required decision threshold.

In our experiments we did not make use of repetition, nor did we incorporate it into our model. We look forward to continuing research in experimental psychology which will clarify how and where repetition is made use of in determining and predicting intention. These findings can be used to correctly include repetition in our model.

2.4. Predicting Intention

As opposed to intention detection (i.e. determining whether or not the observed actions were performed intentionally), intention prediction (i.e. identifying 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 computer science and engineering.

Cognitive Modeling and Developmental Psychology. Huang et al. (2002) 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. In order to make use of this clue, infants—and artificial agents with the same social abilities—must be able to identify what actions can be performed with objects, i.e., the objects' affordances. We make significant use of affordances in our model of intention prediction, and discuss affordances in Section 2.5. We examine this clue of stimulus enhancement, and show that while it is a useful one, it is by no means the only one, nor the most significant.

Research in psychology attempts to pinpoint the age at which intention understanding matures. By correctly placing it within the context of other developing skills, 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 & 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 enables correctly identifying and implementing the building blocks of artificial cognitive systems with intention understanding abilities. Nehaniv & Dautenhahn (2007), Oztop & Kawato (2005), Meltzoff & Decety (2003) and Meltzoff et al. (1999) are all examples of this approach.

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, two challenges raised by Meltzoff's (1995) experiments are rarely, if at all, addressed in plan recognition literature: 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?

We note that most recent plan recognition works focus on using probability distributions over possible explanations for an observed sequence of actions (Charniak & Goldman, 1993; Geib & Goldman, 2005). Using consistency rules (Lesh & Etzioni, 1995; Hong, 2001) and learning (Blaylock & Allen, 2006; Wang et al., 2012), earlier goal recognition systems return a likelihood-ranked set of goals consistent with the observed sequence. We too evaluate the use of a probability distribution over possible goals. However, as we show, people utilize additional information (aside from a-priori likelihood and distance) in making their inference. Avrahami-Zilberbrand & Kaminka (2007) discuss additional ways, such as a bias towards hypotheses that signify threat.

Another method which utilizes probability distributions over possible goals is that of Kelley et al. (2008). Their method uses Hidden Markov Models and is implemented on a robot, dealing with recognition of activity, such as "following", "meeting", "passing by". We target a different category of intentions, namely, those which can be formulated as a state of the environment which serves as a goal.

Kelley et al. (2012) further develop their system for intention recognition, based on contextual information, and employing affordances. Our approach differs from theirs in several aspects, first and foremost in that we separate intentionality detection from prediction. That aside, their work makes use of relevant contextual information (i.e., the context to the observed actions), while we concentrate on the information available in the action stream alone (though possibly, as the actions imply relations between objects). In principle, the two approaches are complementary. See also the discussion on affordances in Section 2.5.

Another system which recognizes intentional actions has been implemented by Hongeng & Wyatt (2008) on a robot. Their work differs from ours in several respects. First and foremost, they emphasize the visual input analysis, which is outside the scope of our work. Second, they aim to identify *actions*, such as grasp, reach, push, and not *states*, i.e., desired end-states of the world, which is what we do. Towards the end of their article, 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 is in ours. Finally, they do not compare the performance of their system to that of humans, as we do here.

An approach similar to ours has been suggested by Ramirez & 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. However, while they aim specifically to solve the automated planning problem, we work within the context of human intention recognition. Additionally, we address intention detection, which is outside the scope of their work. Finally, since we are interested in how humans perform the task, we conduct experiments in which we compare the performance of our model to that of humans.

Robot Imitation. Another field in which there is work relevant to ours, is that of robot imitation, where one of the key challenges is recognizing the goal to be imitated, in our words, "intention prediction" (for a review, see Breazeal & Scassellati, 2002).

Recent work in this area emphasizes the use of affordances (see Section 2.5 below), which we make use of as well. For instance, Lopes et al. (2007) 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.

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 St. Amant (1999) and Sahin et al. (2007).

Since Gibson's (1977) introduction of affordances, as ecological properties of the environment which depend on the perceiver, the concept has evolved into various forms. The term "affordance" is thus 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, putting the notion into the context of intention prediction.

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. For example, Gaver (1991) defines affordances as "potentials for action". The same is true of Cisek (2007), who 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 in the artificial intelligence planning literature 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.

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 & Smitsman, 1995), "containment" (Carona et al., 1988; Chiarello et al., 2003), "above" and "below" (Quinn, 1994) and "tight-fit" (Casasola & Cohen, 2002).

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 & Smitsman, 1995; Carona et al., 1988; Chiarello et al., 2003; Quinn, 1994; Casasola & 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 & 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 & 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. Our model makes use of such an affordance library.

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 & 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 (e.g., Grezes & Decety (2002); Grezes et al. (2003)).

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.

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. The model we propose shows how the principle of rationality is used to choose between affordances invoked by an observed sequence of actions, and we indeed demonstrate that rationality overrides a-priori likelihoods.

3. Intention Recognition Processes

Intention recognition involves—in addition to other processes—both detection of the presence of intention, and prediction of the intention. These are two separate core processes. We believe them to be conceptually and practically distinct: 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 (see Section 2.3). We propose, in accordance with the findings from the work of Meltzoff (1995) (see Section 2.1) that the determining factor in the decision to imitate or not to imitate the observed acting adult, 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.

In this section, we first describe an abstract view of intention recognition as a whole, with some of its major components evident from previous work (Section 3.1). We then describe each of the components (Sections 3.2–3.5).

The main contributions of this paper focus in particular around the use of rationality in detecting intention (Section 3.3) and predicting it (Section 3.5). The others are discussed briefly, to provide context.

3.1. Overview

We propose an abstract model of intention recognition (schematically described in Figure 1) to put the work described in this paper in context. The model 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 observed 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, theoretical justification will be given for the modules and the connections between them, and in the next sections, empirical evidence will be provided for those two modules which are at the core of the model: Intention Detection and Intention Prediction.

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 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. Section 3.2 discusses this in detail.

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. Section 3.3 elaborates on this.

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 the notion of affordances, as discussed above. Recall that 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 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.

3.2. Intentional Being Detection

Experiments with children have shown that when the same actions are performed by a human and by a mechanical arm, the observers tend to attribute intention only to the human, and not to the mechanical being (Meltzoff, 1995). Another set of studies (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.

The above serves as conceptual justification to our position that a preliminary condition which actions must fulfill in order to have intention attributed to them, is that they be performed by an intentional being. In addition, there is also the practical consideration: if 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 at all capable of intention.

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.

3.3. 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

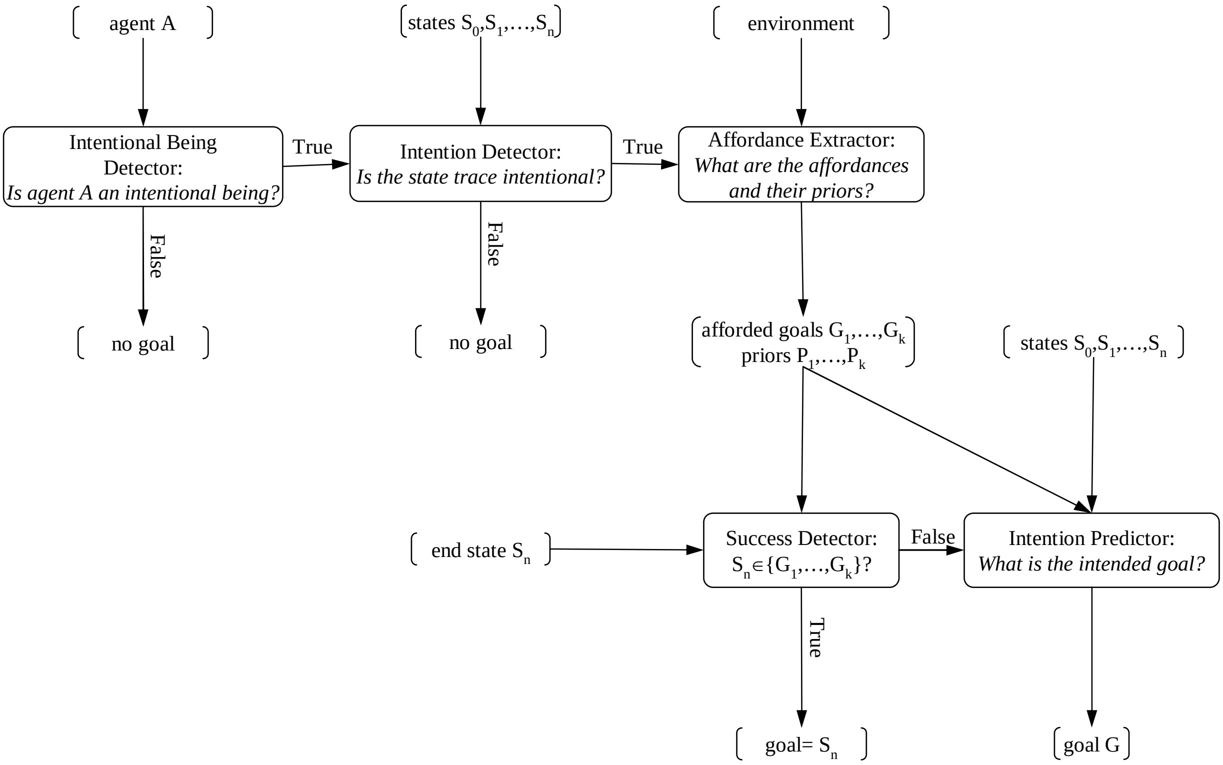

agent A
states $S_0,S_1,\ldots,S_n$
environment
Intentional Being Detector:
Is agent A an intentional being?
True
Intention Detector:
Is the state trace intentional?
True
Affordance Extractor:
What are the affordances and their priors?
False
False
no goal
no goal
afforded goals $G_1,\ldots,G_k$
priors $P_1,\ldots,P_k$
states $S_0,S_1,\ldots,S_n$
end state S_n
Success Detector:
$S_n \in \{G_1,\ldots,G_k\}$?
False
Intention Predictor:
What is the intended goal?
True
goal= S_n
goal G

Figure 1: Scheme of Proposed Model.

answer is binary: `True` if the action sequence is deemed intentional, `False` otherwise.

How can this question be answered? As described above (Section 2), we hypothesize that a sufficient condition of intentional action, under some conditions, is that it is efficient, in the sense of the Principle of Rational Action (Gergely & Csibra, 2003; Watson, 2005). 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, i.e., an efficient way to bring about that goal, given the initial state. 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 propose that the rationality principle serves as a sufficient condition in many settings. 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 intentionality.*

We denote the observation trace by $s_0, ..., s_n$, 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.

We define a state-distance measure *dist*, which measures the optimal sequence of actions between two states of the world, given all possible actions that can transform one state to the other. It should naturally be positive for two states that are different, and equal to 0 when measuring from a state to itself. We do not require this distance to obey symmetry ($d(s_1, s_2) = d(s_2, s_1)$).

Note that this definition includes the Euclidean distance in space (under the assumption that spatial motion is a possible action, but instant star trek-like transportation is not), but is not limited to it. For instance, one can measure the shortest sequence of otherwise-equal actions to load a delivery truck and send it on its way, or to build a stack of blocks in a particular order (two examples that sometimes appear in artificial intelligence planning literature). The requirement is that *dist* capture the notion of optimality.

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.

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,

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

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,

$$t = \frac{u}{n} \tag{2}$$

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. In Section 4 we experiment with this measure evaluating sequences of actions in two different domains. We show that this simple intuitive method does indeed produce promising results, when compared to human assessment of the same sequences.

3.4. Affordance Extraction

If in the previous stage the presence of intention has been established (indicated by an output of `True`) the process continues 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 (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 & 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 & 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 & Tomasello's (1999) 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 of 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 build on this research, and focus on how such an module can be made use of in our context.

3.5. Intention Prediction

This module is the second of the two main topics of this paper. 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. The first two heuristics are operationalizations of clues suggested by others, in previous work. Our model is based on a third heuristic of efficiency and rationality. The experiments contrast this third heuristic with the previous two, and show that it turns out to give better results.

Based on the findings from the affordance literature (quoted above in Section 2.5), 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 contrast three different heuristics 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. This corresponds to work by Cisek (2007), who suggested an affordance competition hypothesis for action selection.

The second heuristic considers further information, namely, that of the state of the environment brought about by the actions, s_n. A distance function, $dist(s_i, s_j)$, between states, is utilized here. The distance measure could be the same one utilized in the Intention Detection module (above, Section 3.3), or not, as long as it fulfills the same requirements, mentioned above, and *it is always optimal.*

Given such a 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 by Huang et al. (2002).

The third heuristic is novel, a contribution of our work. It is motivated by the psychological Principle of Rational Action (Gergely & Csibra, 2003). 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 plan $s_0, ..., s_n, g_j$, and the length of the plan going optimally from s_0 to g_j, and divide the second by the first:

$$r_j = \frac{dist(s_0, g_j)}{\sum_{i=1}^{n} dist(s_{i-1}, s_i) + dist(s_n, g_j)}$$

These lengths are calculated using the same distance function as above. The resulting ratio relays how long the actual plan from s_0 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 section describing the experiments for the Intention Prediction module (Section 5) we evaluate the effectiveness of these heuristics at the task of intention prediction, compared to human performance.

3.6. 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. Under the strong assumption that the affordance extraction module is complete (i.e., generates all possible affordances), the implementation itself consists of nothing more than a simple logical test. However, given an incomplete set of affordances, the decision on success or failure can be complex, involving for instance recognition of facial and vocal expressions, common sense reasoning (background knowledge), recognition of gestures, etc. Expanding on this is outside the scope of this paper.

4. Experiments in Intention Detection

In this section we describe the experiments used to evaluate the proposed measure of intention detection, discussed in Section 3.3. 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[1]). The

[1] PDDL files which support STRIPS notation, compatible with the software employed, were used.

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's (1995) experiments. First, we describe how we rendered Meltzoff's experiments into a computational form, using standard AI planning problem description language (STRIPS) (4.1.1). This is followed by a results section which shows the performance of the model in this environment (4.1.2).

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.

To compute the distance measure *dist*, we use Bonet & Geffner's (1999) HSP[2] (Heuristic Search Planner): given two states of the world, HSP finds the optimal sequence of operators (pickup, put down, move hand) leading from one given state to another. The number of actions in the optimal plan is taken to be the distance between the two given states. We note that this distance measure is not Euclidean, as it considers actions that do not move in space (e.g., grasp, release), and in any case motions are only allowed between grid cells, not arbitrary angles. We also note that the measure gives the same weight to all actions (that is, releasing has the same weight as moving one cell).

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

[2] HSP is downloadable from http://www.ldc.usb.ve/˜bonet/

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 1. A graphic description is given as well for the first trace, in Figure 2. 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 2. For each trace we calculated the sequence of distances, using the above mentioned HSP algorithm, and then computed the proportion t.

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 2(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.

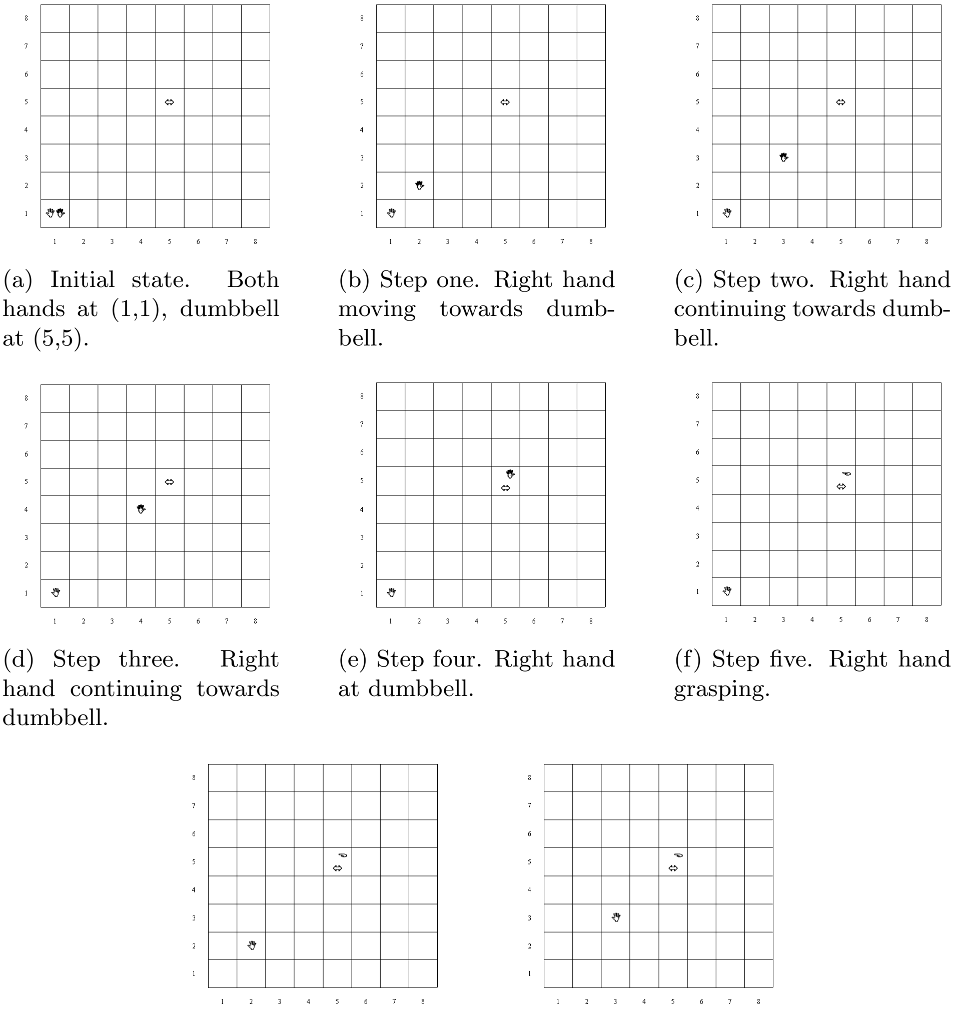
Table 1: Description of traces for each of the experimental groups in the dumbbell experiment.

4.1.2. Results

Figure 3 shows plots of the sequences of distances associated with the dumbbell experiments. The step number in the sequence is depicted in the

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



(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.

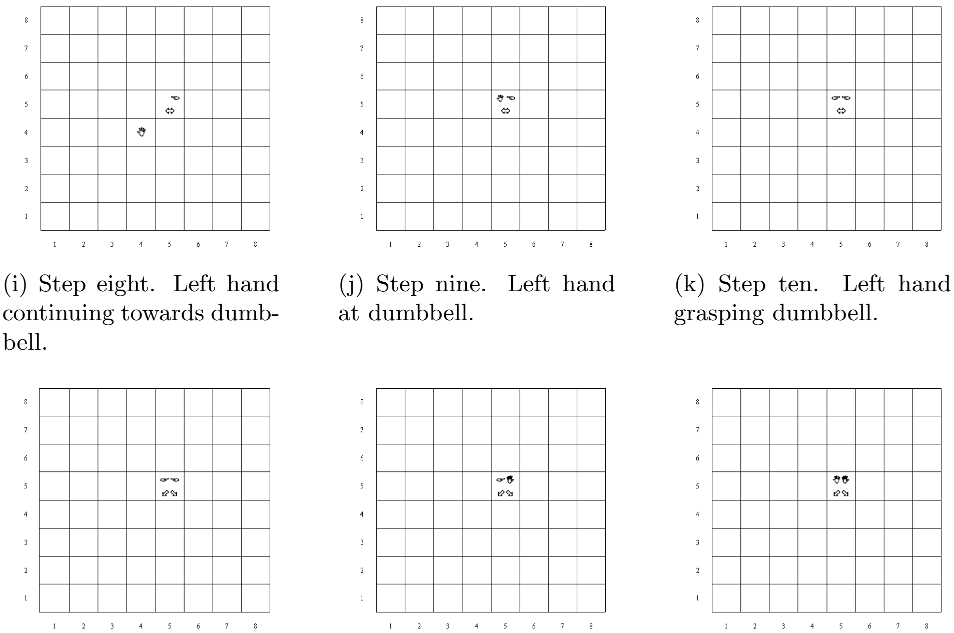
(e) Step four. Right hand at dumbbell.

(f) Step five. Right hand grasping.

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

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

Figure 2: 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.

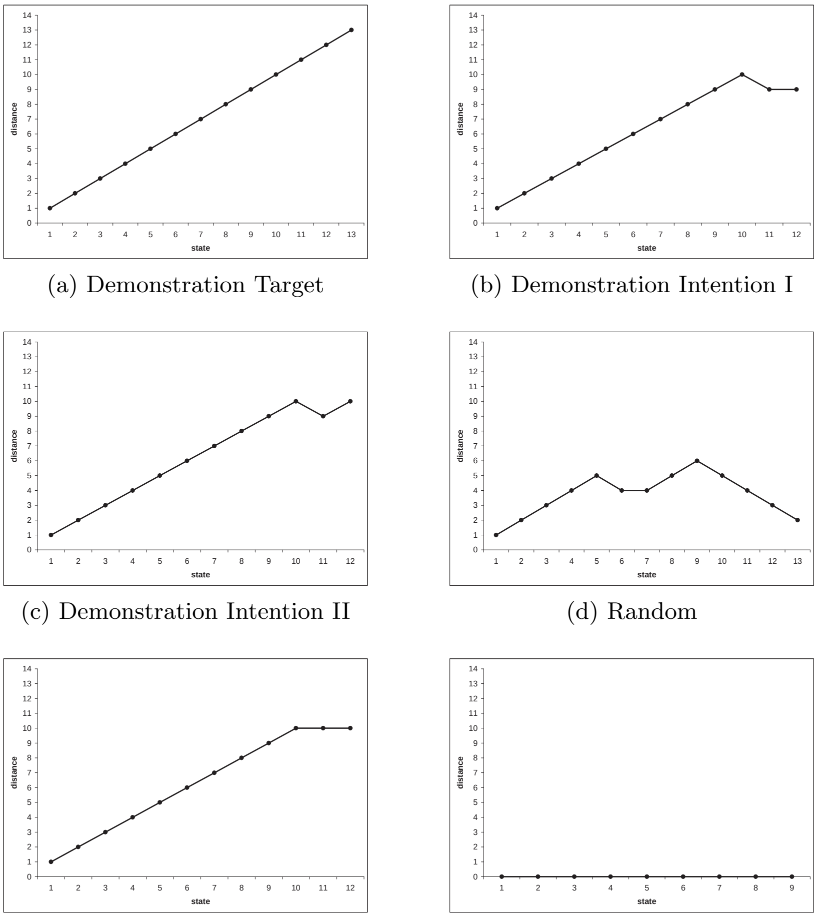
(m) Step twelve. Releasing one hand.

(n) Step thirteen. Releasing other hand.

Figure 2: Dumbbell Demonstration Target Trace (cont).

X axis. The Y axis measures the distance of the respective state from the initial state. Figure 4 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; Meltzoff et al. (1999) 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 on the unclear role of repetition in Section 2.3 for more on this). So, while every child was shown three possibly different traces, we calculated the measure of intention separately for each of these traces, which is why we have more than one row in 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), or too far to the left (Demonstration Intention II in Table 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.



(a) Demonstration Target

(b) Demonstration Intention I

(c) Demonstration Intention II

(d) Random

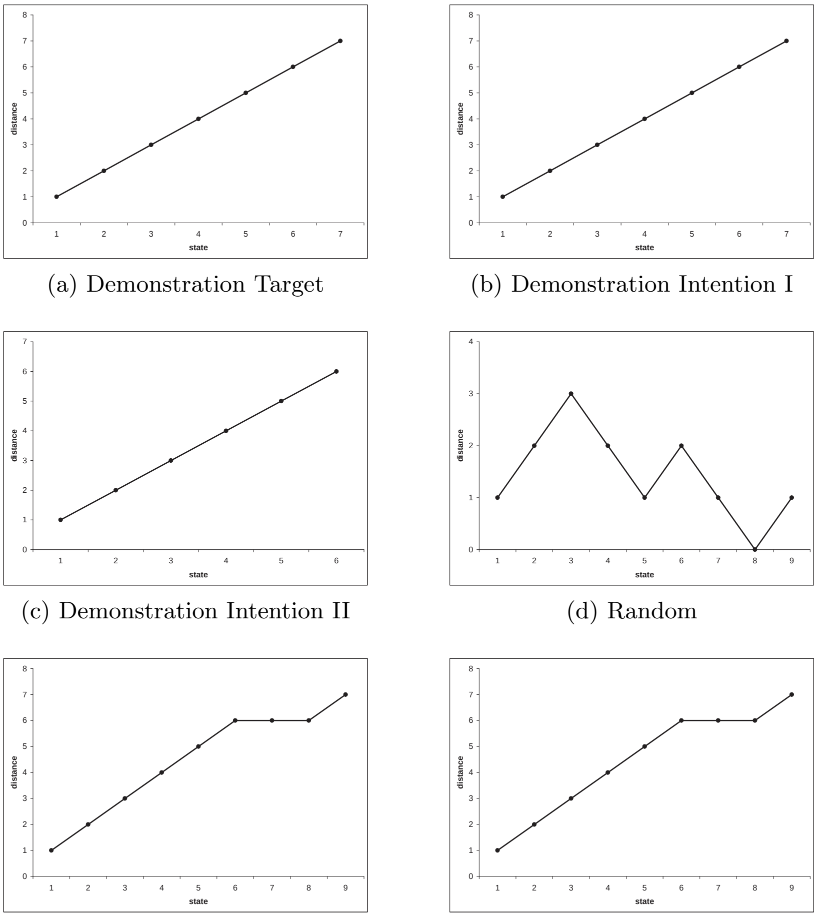
(e) Control Manipulation

(f) Control Baseline

Figure 3: Distance as a Function of State in the Dumbbell Experiments.

Table 3 shows the calculated measure of intention for each of the traces in the dumbbell experiment, and Table 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.

Figure 3a shows the distance sequence for the Demonstration Target trace, for the thirteen-state trace graphically depicted in Figure 2. The



(a) Demonstration Target

(b) Demonstration Intention I

(c) Demonstration Intention II

(d) Random

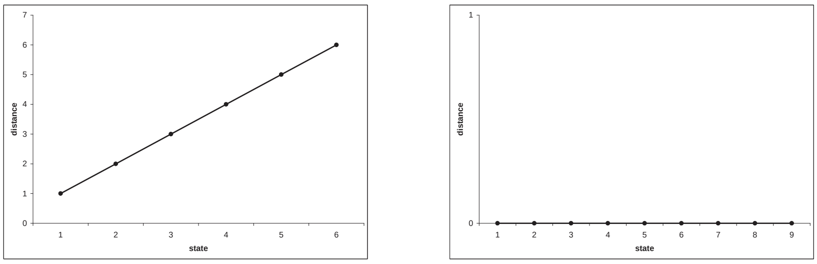
(e) Control Manipulation I

(f) Control Manipulation II

Figure 4: Distance as a Function of State in the Prong and Loop Experiments.

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 3: Calculated Measure of Intention for STRIPS Implementation of Dumbbell Experiment.



(g) Control Manipulation III

(h) Control Baseline

Figure 4: Distance as a Function of State in the Prong and Loop Experiments (cont).

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

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 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 4a 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.

In the case of Demonstration Intention traces, we also get a high measure of intention. See for example the distance sequences of the Demonstration Intention traces in the dumbbell experiment in Figures 3b and 3c. 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 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 4b and 4c. In fact, in this case the maximum score of 1 is reached (see the two corresponding rows in Table 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 3f for the dumbbell experiment and in Figure 4h for the loop and prong experiment. The resulting intention scores are therefore zero, as Tables 3 and 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 3d and 4d, which justly earn them the significantly lower scores appearing in the respective rows of Tables 3 and 4.

4.2. Experiment II: Surveillance Videos, Comparison with Human Subjects

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[3]. Section 4.2.1 describes the environment, followed by a description of the results (Section 4.2.2), 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

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

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

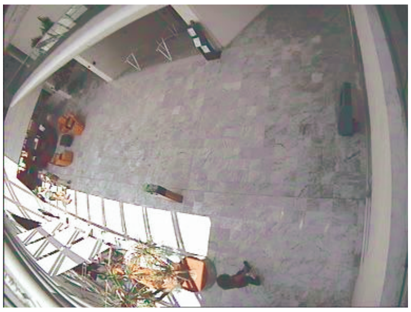


Figure 5: Typical Screen Shot from a CAVIAR Video, with Character Seen Entering From Bottom.

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

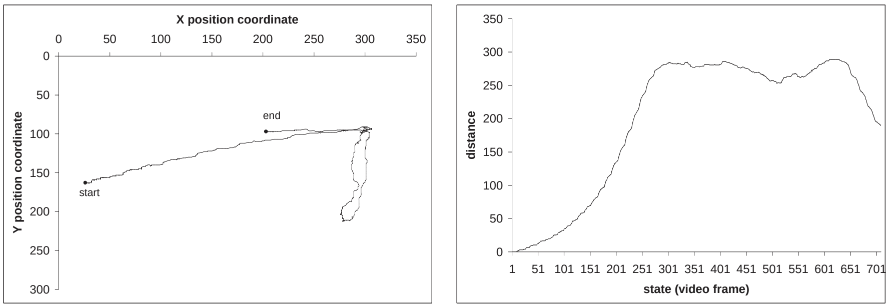
Table 5: Clip numbers with their corresponding video file name, xml file name and object ID in the CAVIAR repository, from which they were taken.

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.

Applying the Measure of Intention to the Data. Let us begin by describing how we measure intention in this domain. 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 it approximates the optimal motion in space[4]. 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 6a 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 6b 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] It is only an approximation of the optimal motion, as it ignores obstacles that are present along the path, which the human in the video necessarily avoids.



(a) Path of Movement.

(b) Distances of each state from initial state.

Figure 6: Examples from Clip Number 6.

Comparing to Human Judgment. These same video clips were shown to 12 human subjects who were asked to write down their opinion regarding the intention of the viewed character. Each human subject saw all 12 clips. 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.

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. To do this, 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 for the analysis described in the following. Two independent coders each analyzed all the input from the 12 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

Measure of Intention Correlates With Human Judgment. Table 6 summarizes the resulting entropy values of this analysis, alongside the intention scores as returned by our method. Figure 7 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.

A negative correlation between entropy and our measure implies a *positive* correlation between human judgment and our measure. Smaller values of entropy signify more agreement between subjects, and thus clearer perceived intention.

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, and 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 a given video segment, the less clear that character conveyed intention to the observing subjects, the lower the intention measure calculated by our method. The inverse is true as well—the lower the entropy, the clearer its intention was to the human observers, and the higher the intention score achieved by our method.

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

Table 6: Measure of Intention and Entropy of Human Judgments for Video Clips.

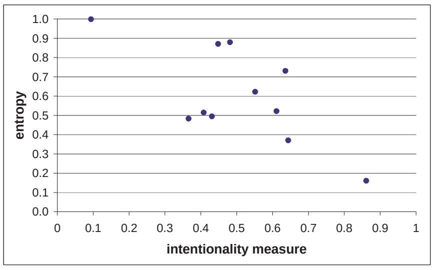


Figure 7: Plot of Entropy vs. Intention. Correlation= -0.686.

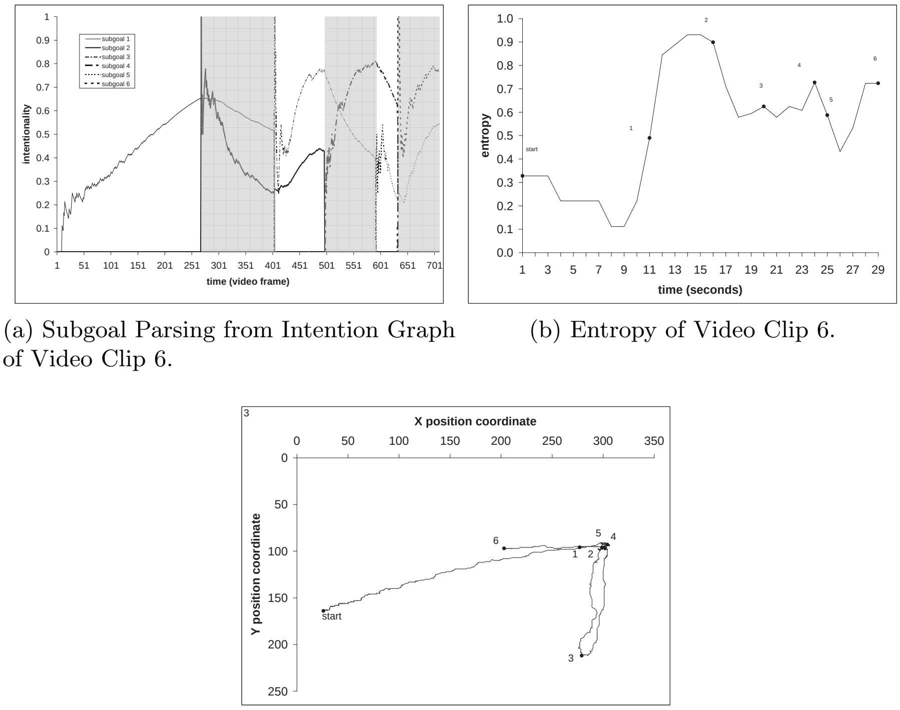
Using Intention Detection for Segmenting Subgoals. While analyzing the results of the second experiment, the matter of parsing streams of action according to sub-goals arose. Several of the clips we analyzed clearly show changing intentions, e.g. when the character stops in mid-track, turns around 180 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 calculated, 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 8. 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).



(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 8: Analysis of Video Clip 6.

Figure 8c 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 7.

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 7: Description of Subgoals Found in Video Clip 6.

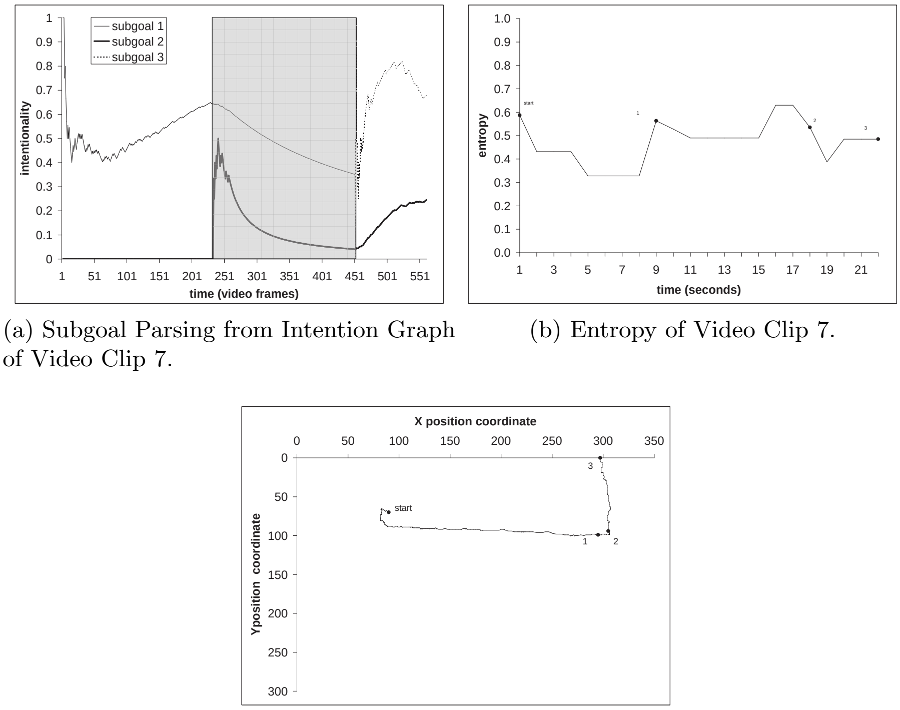
Figure 8b shows the plot of entropy as it changes over time, with numbers indicating where subgoals were found. Note that the behavior of the entropy graph is somewhat inverse to the behaviors of the intention graphs of the subgoals—for the first subgoal, the intention graph is increasing, while the 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 correspondence. 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 9 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 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 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 10. The character in this video moves in a straightforward manner from the bottom



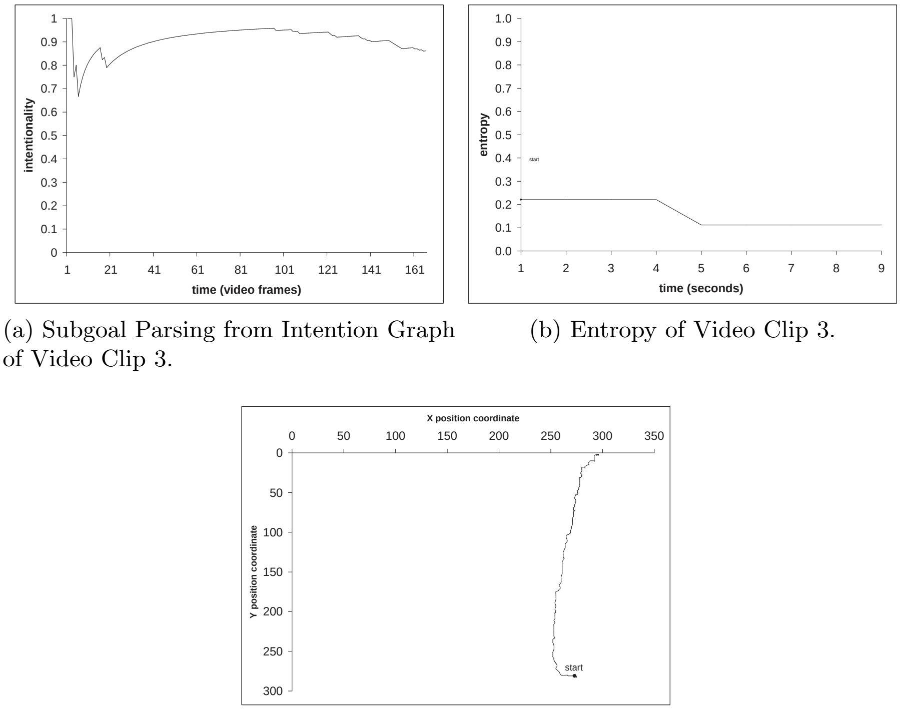
(a) Subgoal Parsing from Intention Graph of Video Clip 7.

(b) Entropy of Video Clip 7.

(c) Path of Movement in Video Clip 7.

Figure 9: Analysis of Video Clip 7.

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.



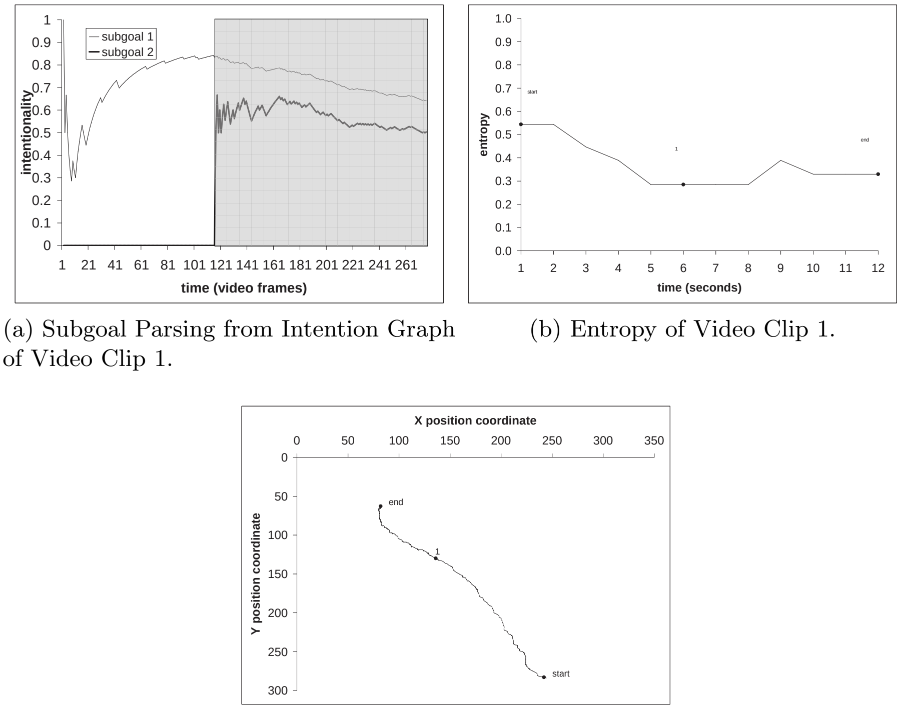
(a) Subgoal Parsing from Intention Graph of Video Clip 3.

(b) Entropy of Video Clip 3.

(c) Path of Movement in Video Clip 3.

Figure 10: Analysis of Video Clip 3.

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 11. Table 9 describes the two 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.



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

(b) Entropy of Video Clip 1.

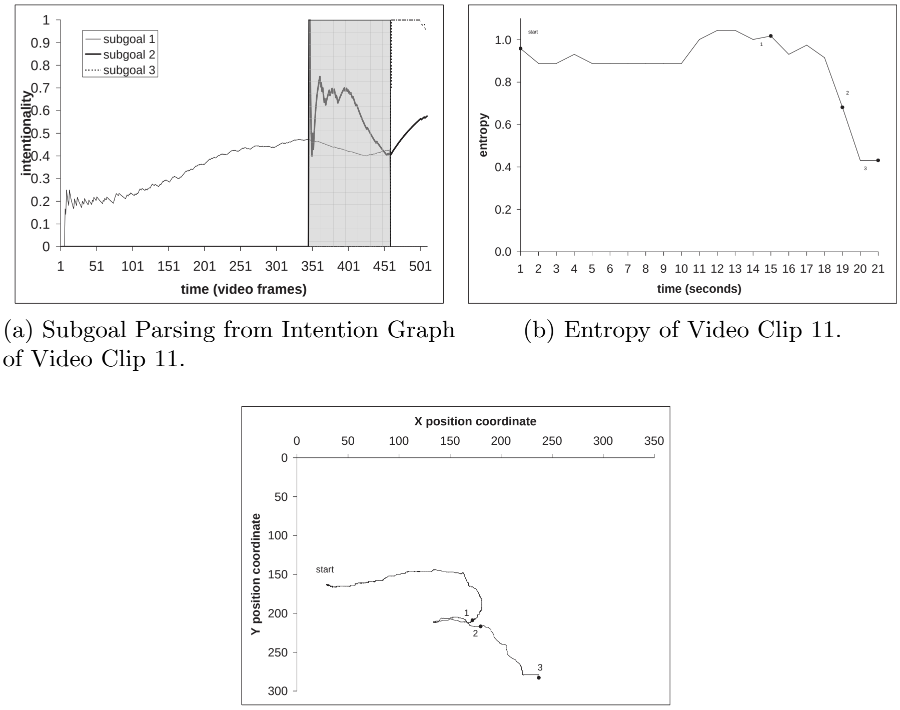
(c) Path of Movement in Video Clip 1.

Figure 11: Analysis of Video Clip 1.

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 9: Description of Subgoals Found in Video Clip 1.

Another example is given in Figure 12, this time of the false negative kind, with the description of subgoals in Table 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 segment 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.



(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 12: Analysis of Video Clip 11.

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 10: Description of Subgoals Found in Video Clip 11.

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 discussed in Section 3.5—the prior p_j (an operationalization of the clue arising from the work of Cisek (2007)), the distance d_j (an operationalization of the clue arising from the work of Huang et al. (2002)) and the intention measure r_j, computed for each afforded goal state $g_{j=1,\ldots,k}$ (the heuristic suggested in this paper). 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. It also shows that d_j and p_j can play secondary roles with regard to this task.

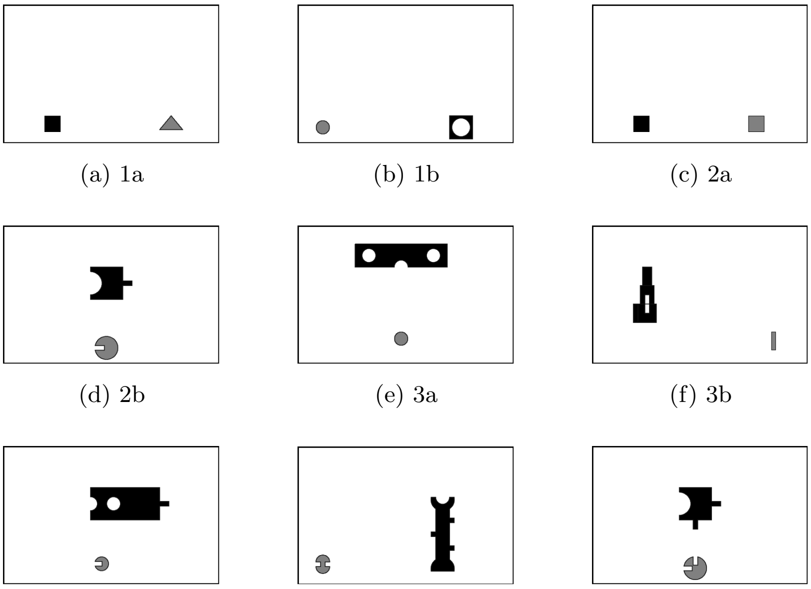
5.1. Experimental Setup

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 gray movable one. The subjects were instructed to drag the gray 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 13, with a corresponding identification code for each.



(a) 1a

(b) 1b

(c) 2a

(d) 2b

(e) 3a

(f) 3b

(g) 3c

(h) 4a

(i) 5a

Figure 13: Object-Pairs and Their Identification Codes.

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 gray object for five of the nine object-pairs used in the first part. For each pair, several paths were constructed, and the gray 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 gray 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 gray 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.

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. Part I: 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 1*a* begs to be configured as a house (Figure 14), with the gray 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 one configuration over another. As an example, consider pair 3*a* (Figure 15), for which the subjects chose to place the gray circle in the middle indentation at the bottom of the black object with 69.29% frequency, while they placed the gray circle in the right hole with 10.71% frequency and in the left hole with 12.86% frequency. Such choices

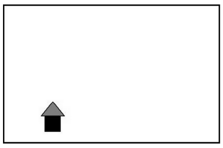
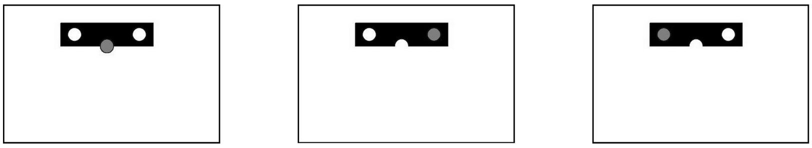


Figure 14: Most Frequent State (A) for Object-Pair 1*a*, with Prior 96.49%.

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.



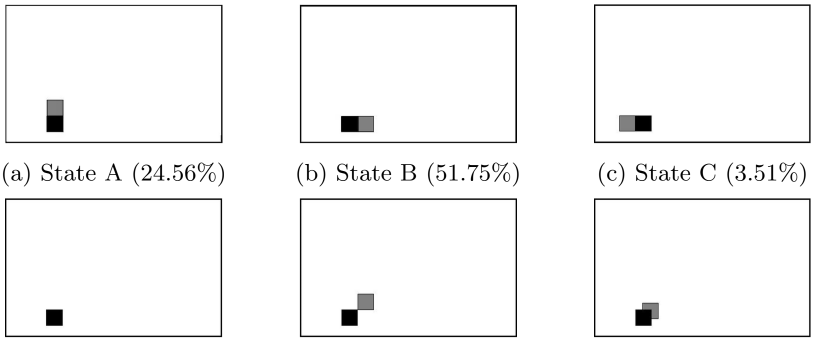
(a) State A (69.29%) (b) State B (10.71%) (c) State C (12.86%)

Figure 15: Most Frequent States for Object-Pair 3a with Their Priors.

Figure 16 shows the empirical priors for object-pair 2a (note that in state D—Figure 16d—the gray square is placed behind the black square, thus obscured by it). Figure 17 shows the priors for object-pair 1b. Figures 18, 19, 20, 21 and 22 show the priors for affordances of object-pairs 2b, 3b, 3c, 4a and 5a, respectively. Figure 17c shows a configuration which was not chosen at all in this first part of the experiment, but was in the second part of the experiment, discussed later.

5.2.2. Part II: 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



(a) State A (24.56%)

(b) State B (51.75%)

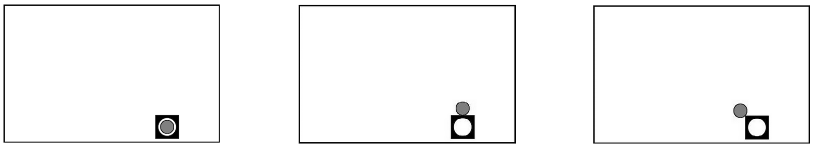
(c) State C (3.51%)

(d) State D (7.89%)

(e) State E (3.51%)

(f) State F (3.51%)

Figure 16: Most Frequent States for Object-Pair 2*a* with their Priors.

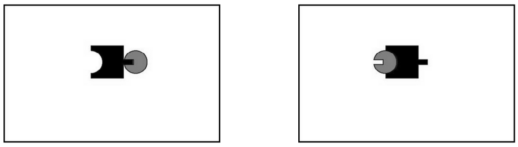


(a) State A (85.82%)

(b) State B (11.35%)

(c) State E (0.00%) (see text)

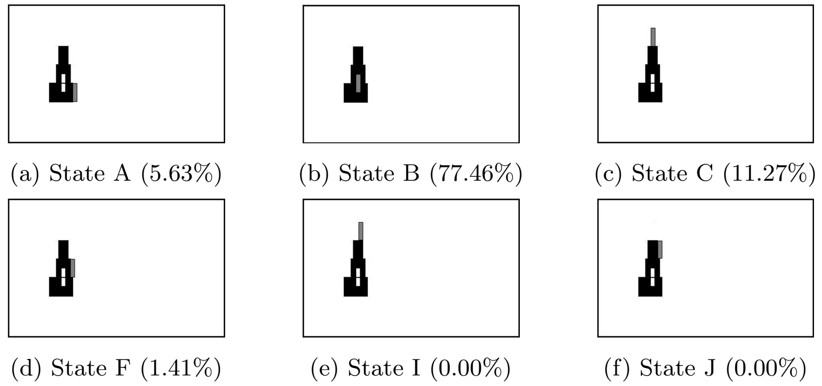
Figure 17: Most Frequent States for Object-Pair 1*b* with their Priors.



(a) State A (68.42%)

(b) State B (30.70%)

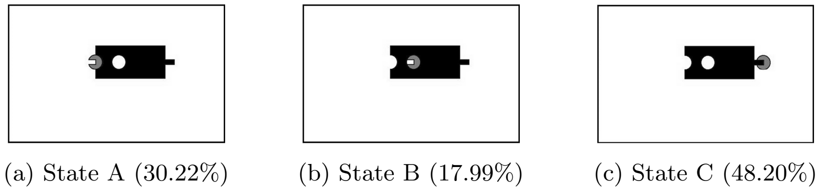
Figure 18: Most Frequent States for Object-Pair 2*b* with their Priors.



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

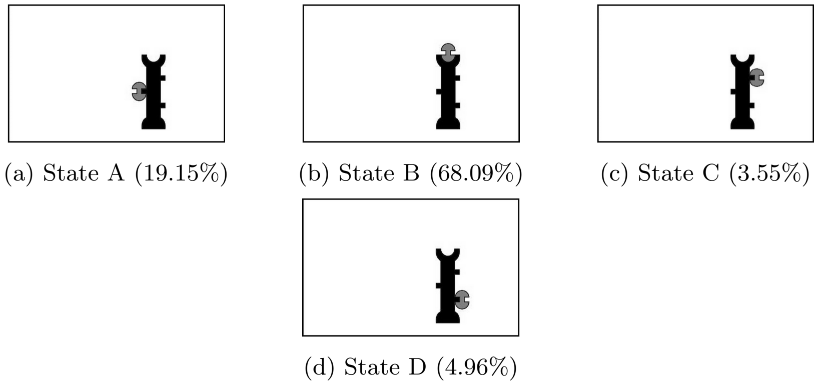
(d) State F (1.41%) (e) State I (0.00%) (f) State J (0.00%)

Figure 19: 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%)

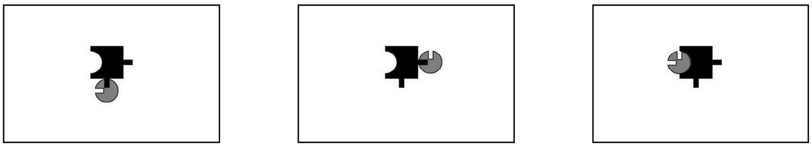
Figure 20: Most Frequent States for Object-Pair 3*c* with Their Priors.



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

(d) State D (4.96%)

Figure 21: 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 22: Most Frequent States for Object-Pair $5a$ with Their Priors.

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 23. 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.

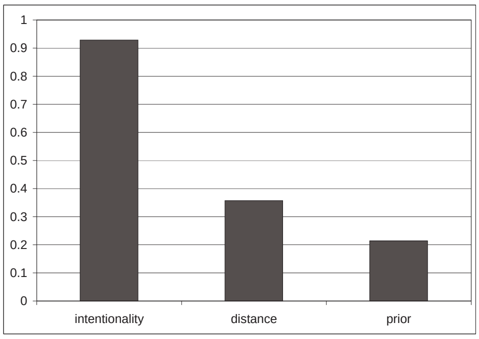


Figure 23: Success Rate of Each Heuristic at Predicting the Correct Goal.

The details of these matchings are given in Table 11. 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 11: 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 24). 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 17), and the frequencies according to which they were chosen, for each path, are given in Table 12.

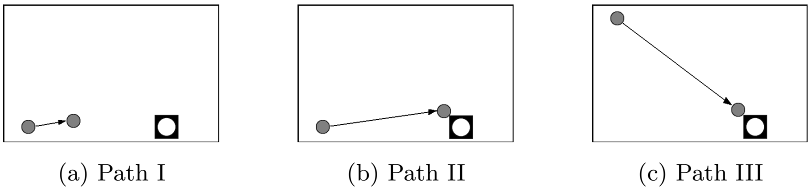

(a) Path I (b) Path II (c) Path III

Figure 24: Paths for Object-Pair 1b.

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 12: Frequencies of Choices for Object-Pair 1b.

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 17). These values of the intention measure are shown in Table 13. 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.

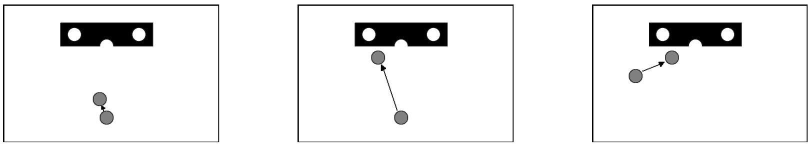
state\path	I	II	III
A	0.996	0.949	0.999
B	1.000	1.000	0.973

Table 13: Measure of Intention for Object-Pair 1b.

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. 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 3a. Object-pair 3a supports these results as well. Here too, three paths were shown to the subjects (Figure 25). Path II begins as Path I does, and continues further. Paths II and III end at the same position, but begin at different ones. Table 14 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 15.



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

Figure 25: Paths for Object-Pair 3a.

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 14: Frequencies of Choices for Object-Pair 3a.

Table 15 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 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.

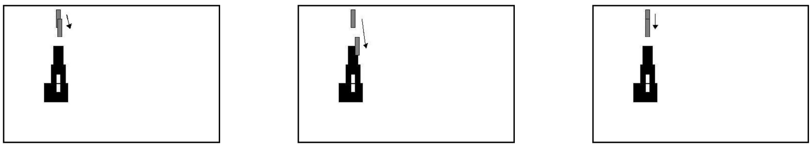
state\path	I	II	III
A	0.98	0.79	1.00
B	0.94	0.72	1.00
C	1.00	1.00	0.75

Table 15: Measure of Intention for Object-Pair 3a.

Object-Pair 3b. Results for object-pair 3b are shown next. Figure 26 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 16 shows the subjects' choice of goal states for each of the paths. Table 17 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 16—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.

Object-Pair 3c. The fourth of the object-pairs presented to the subjects was 3c. The three paths for this pair are given in Figure 27. 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 18, and the measures of intention in Table 19. 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



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

Figure 26: Paths for Object-Pair 3*b*.

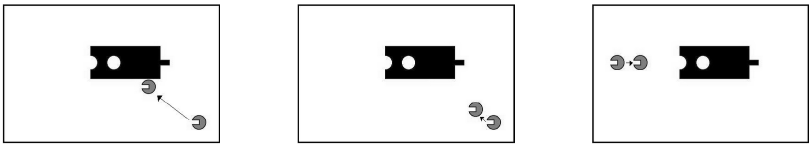
state\path	I	II	III
A	13.67	41.73	0.00
B	18.71	7.19	59.71
C	14.39	0.00	36.69
F	7.19	28.06	0.00
I	17.27	0.72	0.72
J	25.18	18.71	0.00

Table 16: 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 17: Measure of Intention for Object-Pair 3*b*.

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 27: Paths for Object-Pair 3c.

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 18: Frequencies of Choices for Object-Pair 3c.

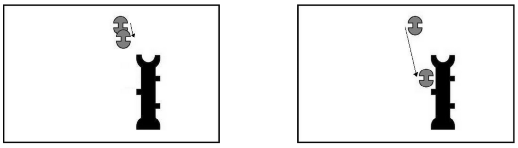
state\path	I	II	III
A	0.994128317	0.998505671	1
B	0.999995551	0.99998616	1
C	0.718324618	0.948277727	1

Table 19: Measure of Intention for Object-Pair 3c.

Object-Pair 4a. Object-pair 4a 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 28, and the resulting frequencies for the four most chosen goal states are in Table 20. 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 21, and once again, the highest ranking goal state matches that which was most often chosen by the subjects.

5.2.3. Breaking Ties: The Role of Priors and Distance.

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



(a) Path I

(b) Path II

Figure 28: Paths for Object-Pair $4a$.

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 20: Frequencies of Choices for Object-Pair $4a$.

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 21: Measure of Intention for Object-Pair $4a$.

cases (Path III of object-pair $3a$, Path III of object-pair $3b$, and Path III of object-pair $3c$), 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 22 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 intention".

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

Table 22: 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 $3a$ (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 $3b$ (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 $3c$ (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 23, 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 23, 24, 25, 26, and 27. 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.

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 23: 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 24: 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 25: Distances for Object-Pair 3*b*.

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 intention 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

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 26: 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 27: Distances for Object-Pair 4*a*.

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 3*b* 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 3*b* 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 1*b*. 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.

6. Discussion and Future Work

Several points pertaining to the above-presented results deserve further considerations. Some of them require further clarification within the field of psychology, and as such are left for future research. We conclude with a short summary of the contributions of this work.

6.1. Plan Generation and Intention Recognition, and Ties to Mirroring

A key insight we ourselves draw from this work is on the intimate relationship between the ability to plan towards a goal, and the ability to recognize intentions. Optimal planning is inspired by the principle of rationality. And this paper shows that at least for some forms of intentional actions, so is intention recognition. Thus conceptually, the capabilities are linked.

The link between plan generation and intent recognition is in fact very strong in this work. To detect intentionality using the principle of rational action, this paper suggests using a heuristic measure which computes the ratio between the length of the observed sequence (in plan steps, generalizable to plan step costs), and an optimal plan for achieving the same end-state of the sequence. This definition makes it a clear requirement for the observing agent to be able to generate a hypothesized optimal plan, using the actions available to the observed agent. This raises two separate issues.

First, much depends here on the planning capabilities of the observer. For instance, if we had observed someone drawing a circle in continuous motion (e.g., using a marker on a whiteboard), we may not be able to detect the motion's intentionality with ease, if we had known the drawer had the ability to draw a circle in a single action (e.g., via a stamp in the shape of a circle). Or observing someone erasing a whiteboard, we could not use the Euclidean distance measure as the basis for detecting intentionality, as in Section 4.2. This is because the optimal plan for erasing a whiteboard (against which the observed sequence is compared), is not a plan that moves in a straight line (which the Euclidean distance can quickly summarize), but rather an optimal coverage plan, that moves optimally through the entire space of the whiteboard[5]. When compared to such a plan, a systematic back-and-forth motion of an eraser would be closer to optimality than random erasing motion (touch here, touch there, until the entire whiteboard is clean.) And as a result, recognizing intentionality in the first (systematic) case would be successful using the measure we proposed.

Second, the definition emphasizes that the observing agent must not only be able to recognize the actions in the observed sequence, it must also be able to use them (and others) to compute a hypothetical optimal plan, to compare against. In other words, knowledge of these actions must be stored in a way that is accessible both for planning as well as for recognition. This begs the question of a possible relation between intention recognition as described here, and action recognition, e.g., via mirror neurons and mirroring processes. We hypothesize as to that relation elsewhere (Kaminka, 2013).

6.2. Different Measures of Intention

Two related problems were addressed in this work. For each problem, a different measure of intention was proposed. The question begs to be asked: could not one unified measure be devised, so as to solve both problems? After all, both measures claim to capture a sense of intention.

Yet, there are inherent difference between the problems. For example, for determining the existence of intention (first problem), at each point in the trace we look back at the observed sequence of movement so far. While for determining the content of the intention (second problem), we look forward, from the last observed state, to possible goals. In addition, artificial examples can be devised, for which the first measure fails on the second problem, and vice versa.

A deeper inspection of the two problems and how they relate to each other is called for. We hope that future research might reveal a unified measure for both problems.

6.3. Determining the Point of Failure

According to our model, once failure has been determined (at the stage of Success Detection, Section 3.6), the process of Intention Prediction kicks in.

[5] Incidentally, this is an area in which we have some experience, see Hazon & Kaminka (2008).

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, the Principle of Rational Action on which our measure of intentionality were based, stems from Gergely & 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 paper are:

- A proposal of an abstract model relating all the necessary components which play a part in the process of intention recognition for intentional actions whose purpose is to achieve a goal (as discussed in Section 2.2).
- 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. We intend to use the insights from this work to fill in the details in the abstract model described above. We would additionally need to add details as to how affordances are extracted, and how optimal plans are generated. Once fully implemented, such a model could be applied to a complete intention recognition task of any one of the CAVIAR activities: from detecting presence of intention to detecting the underlying intentions in a sequence, including sub-goals.

At the same time, the model can be expanded to deal with false beliefs and pretense, as well as static and dynamic environmental constraints, and to incorporate additional methods of intention detection, as discussed in Section 2.2, and other categories of intention, as mentioned in Section 2.3. In addition, other distance measures need to be devised and evaluated, for this model to be implemented in different environments. To do this, a language needs to be developed, with which to describe affordances in various environments. Given this, a complete cycle of detection and prediction can be executed–something which we could not do in our experiments on the CAVIAR repository, since the only information available was spatial coordinates of movement.

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, as a component in mindreading.

Acknowledgments

We thank the reviewers for useful comments. This research was enabled through funding by the Paula Rich Foundation, and supported in part by grants from IMOD, the Israeli Science Foundation (ISF), and AFRL/EOARD. Thanks also to Shlomit Kaminka and K. Ushi for their valuable input and support.

References

Avrahami-Zilberbrand, D., & Kaminka, G. (2007). Towards dynamic tracking of multi-agents teams: An initial report. In *Proceedings of Workshop on Plan, Activity, and Intent Recognition*.

Babbie, E. (2003). *The Practice of Social Research*. Wadsworth Publishing.

Baldoni, M., Boella, G., & van der Torre, L. (2006). Modelling the interaction between objects: Roles as affordances. In J. Lang, F. Lin, & J. Wang (Eds.), *Knowledge Science, Engineering and Management* (pp. 42–54). Springer Berlin / Heidelberg volume 4092 of *Lecture Notes in Computer Science*.

Banchetti-Robino, M. (2004). Ibn sina and husserl on intention and intentionality. *Philosophy East and West*, *54*, 71–82.

Bellagamba, F., & Tomasello, M. (1999). Reenacting intended acts: Comparing 12- and 18-month-olds. *Infant Behavior and Development*, .

Blakemore, S., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews, Neuroscience*, .

Blaylock, N., & Allen, J. (2006). Hierarchical instantiated goal recognition. In *AAAI Workshop on Modeling Others from Observations (MOO-2006)*.

Bonet, B., & Geffner, H. (1999). Planning as heuristic search: new results. In S. Biundo, & M. Fox (Eds.), *Procceedings of the 5th European conference on planning*. Durham, UK: Springer: Lecture Notes on Computer Science.

Booth, A., & Waxman, S. (2002). Object names and object functions serve as cues to categorization for infants. *Developmental Psychology*, *38*, 948–957.

Brandone, A. (2010). *The Development of Intention Understanding in the First Year of Life: An Exploration of Infants' Understanding of Successful vs. Failed Inentional Actions*. Ph.D. thesis University of Michigan.

Brandone, A., & Wellman, H. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science*, *20*, 85–91.

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

Breazeal, C., & Scassellati, B. (2002). Robots that imitate humans. *Trends in Cognitive Sciences*, *6*, 481–487.

Carona, A., Carona, R., & Antell, S. (1988). Infant understanding of containment: An affordance perceived or a relationship conceived? *Developmental Psychology*, *24*, 620–627.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month old infants differentially imitate intentional and accidental actions. *Infant behavior development*, .

Casasola, M., & Cohen, L. (2002). Infant categorization of containment, support and tight-fit spatial relationships. *Developmental Science*, *5*, 247–264.

Charniak, E., & Goldman, R. (1993). A bayesian model of plan recognition. *Artificial Intelligence*, *64*, 53–79.

Chiarello, E., Casasola, M., & Cohen, L. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development*, *72*, 679–693.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society, Biological Science*, *362*, 1585–1599.

Craighero, L., Fadiga, L., Umilta, C., & Rizzolatti, G. (1996). Evidence for visuomotor priming effect. *NeuroReport*, *8*, 347–349.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: a developmental hypothesis. *Developmental science*, *1*, 255–259.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In M. Johnson, & Y. Munakata (Eds.), *Processes of Change in Brain and Cognitive Development: Attention and Performance* (pp. 249–274). Oxford University Press.

Dapretto, M., Davies, M., Pfeifer, J., Scott, A., Sigman, M., Bookheimer, S., & Iacoboni, M. (2005). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*, 28–30.

Davies, M., & Stone, T. (1995). *Mental Simulation: Evaluations and Applications*. Blackwell Publishers.

Fogassi, L., Ferrari, P., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, *308*, 662–667.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Science*, *2*, 493–501.

Gaver, W. (1991). Technology affordances. In *CHI '91 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology*.

Geib, C., & Goldman, R. (2005). Partial observability and probabilistic plan/goal recognition. In *Proceedings of the IJCAI-05 Workshop on Modeling Others From Observations*.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naive theory of rational action. *TRENDS in Cognitive Science*, *7*.

Gergely, G., Nasady, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Gibson, J. (1977). The theory of affordances. In *Perceiving, Acting, and Knowing: Toward an Ecological Psychology* (pp. 67–82). Hillsdale, NJ: Lawrence Erlbaum.

Gordon, R. (1986). Folk psychology as simulation. *Mind and Language*, *1*, 158–171.

Grezes, J., & Decety, J. (2002). Does visual perception of object afford action? evidence from a neuroimaging study. *Neuropsychologia*, *40*, 212–222.

Grezes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. (2003). Objects automatically potentiate action: an fmri study of implicit processing. *European Journal of Neuroscience*, *17*, 2735–2740.

Harui, K., Oka, N., & Yamada, Y. (2005). Distinguishing intentional actions from accidental actions. In *Proceedings of the 4th IEEE international conference on development and learning*.

Hazon, N., & Kaminka, G. (2008). On redundancy, efficiency, and robustness in coverage for multiple robots. *Robotics and Autonomous Systems*, *56*, 1102–1114.

Heal, J. (2003). Mind, reason and imagination. In *Understanding Other Minds from the Inside* (pp. 28–44). Cambridge University Press.

Heider, F. (1958). *The psychology of interpersonal relationships*. Wiley.

Hofer, T., Hauf, P., & Aschersleben, G. (2005). Infant's perception of goal-directed actions performed by a mechanical device. *Infant Behavior and Development*, *28*, 466–480.

Hong, J. (2001). Goal recognition through goal graph analysis. *Journal of Artificial Intelligence Research*, *15*, 1–30.

Hongeng, S., & Wyatt, J. (2008). Learning causality and intentional actions. In E. Rome, J. Hertzberg, & G. Dorffner (Eds.), *Towards affordance-based robot control*. Springer.

Horst, J., Oakes, L., & Madole, K. (2005). What does it look like and what can it do? category structure influences how infants categorize. *Child Development*, *76*, 614–631.

Huang, C., Heyes, C., & Charman, T. (2002). Infants behavioral reenactment of failed attempts: Exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology*, *38*, 840–855.

Kaminka, G. A. (2013). Curing robot autism: A challenge. In *The International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-13).*

Kelley, R., King, C., Tavakkoli, A., Nicolescu, M., Nicolescu, M., & Bebis, G. (2008). An architecture for understanding intent using a novel hidden markov formulation. *International Journal of Humanoid Robotics, Special Issue on Cognitive Humanoid Robots*, *5*, 203–224.

Kelley, R., Tavakkoli, A., King, C., Ambardekar, A., Nicolescu, M., & Nicolescu, M. (2012). Context-based bayesian intent recognition. *IEEE Transactions on Autonomous Mental Development*, *4*, 215–225.

Kiraly, I., Jovanovic, B., Prinz, W., Aschersleben, G., & Gergely, G. (2003). The early origins of goal attribution in infancy. *Conciousness and cognition*, *12*, 752–769.

Lesh, N., & Etzioni, O. (1995). A sound and fast goal recognizer. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1704–1710).

Lopes, M., Melo, F., & Montesano, L. (2007). Affordance-based imitation learning in robots. In *Proceedings of 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Madole, K., & Cohen, L. (1995). The role of object parts in infants attention to form function correlations. *Developmental Psychology*, *31*, 637–648.

Madole, K., Oakes, L., & Cohen, L. (1993). Developmental changes in infants attention to function and formfunction correlations. *Cognitive Development*, *8*, 189–209.

Marhasev, E., Hadad, M., Kaminka, G. A., & Feintuch, U. (2009). The use of hidden semi-markov models in clinical diagnosis maze tasks. *Intelligent Data Analysis*, *13*, 943–967.

Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*.

Meltzoff, A. (2002). Imitation as a mechanism of social cognition: origins of empathy, theory of mind, and representation of action. In U. Goswami (Ed.), *Blackwells Handbook of Childhood Cognitive Development*. Blackwell.

Meltzoff, A. (2007). The "like-me" framework for recognizing and becoming an intentional agent. *Acta Psychologica*, *124*, 26–43.

Meltzoff, A., & Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London*, (pp. 491–500).

Meltzoff, A., & Gopnik, A. (1993). The role of imitation in understanding persons and developing a theory of mind. In H. Baron-Cohen, S. Tager-Flusberg, & D. Cohen (Eds.), *Understanding Other Minds* (pp. 335–366). Oxford University Press.

Meltzoff, A., Gopnik, A., & Repacholi, B. (1999). Toddlers' understanding of intentions, desires and emotions: exploration of the dark ages. In P. Zelazo, J. Astington, & D. Olson (Eds.), *Developing theories on intention: social understanding and self control*. Lawrence Erlbaum Associates.

Meltzoff, A., & Moore, M. (1992). Early imitation within a functional framework: The importance of person identity, movement and development. *Infant Behavior and Development*, *15*, 479–505.

Meltzoff, A., & Moore, M. (1994). Imitation, memory and the representation of persons. *Infant Behavior and Development*, *17*, 83–99.

Meltzoff, A., & Moore, M. (1995). Infants' understanding of people and things: From body imitation to folk psychology. In J. Bermdez, A. Maarcel, & N. Eilan (Eds.), *The Body and the Self* (pp. 43–69). The MIT Press.

Nehaniv, C., & Dautenhahn, K. (Eds.) (2007). *Imitation and Social Learning in Robots, Humans, and Animals: Behavioural, Social and Communicative Dimensions*. New York: Cambridge University Press.

Nielsen, M. (2009). 12-month-olds produce others intended but unfulfilled acts. *Infancy*, *14*, 377–389.

Oztop, D., E.and Wolpert, & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, *22*, 129–151.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, *1*.

Quinn, P. (1994). The categorization of above and below spatial relations by young infants. *Child Development*, *65*, 58–69.

Ramirez, M., & Geffner, H. (2010). Probabilistic plan recognition using off-the-shelf classical planners. In *Proceedings of the 10th AAAI Conference on Artificial Intelligence*.

Sahin, E., Cakmak, M., Dogar, M., E., U., & Ucoluk, G. (2007). To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior December*, *15*, 447–472.

Sitskoorn, M., & Smitsman, A. (1995). Infants' perception of dynamic relations between objects: Passing through or support? *Developmental Psychology*, *31*, 437–447.

St. Amant, R. (1999). User interface affordances in a planning representation. *Human Computer Interaction*, *14*, 317–354.

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

Wang, Z., Deisenroth, M., Amor, H. B., Vogt, D., Scholkopf, B., & Peters, J. (2012). Probabilistic modeling of human movements for intention inference. In *Proceedings of Robotics: Science and Systems, VIII*.

Watson, J. (2005). The elementary nature of purposive behavior: Evolving minimal neural structures that display intrinsic intentionality. *Evolutionary Psychology*, *3*, 24–48.

Woodward, A. (1998). Infants selectively encode the goal object of an actors reach. *Cognition*, *69*, 1–34.

Woodward, A., Sommerville, J., & Guajardo, J. (2001). How infants make sense of intentional action. In B. Malle, & L. Moses (Eds.), *Intentions and Intentionality: Foundations of Social Cognition* (pp. 149–169). MIT Press.

Ye, L., Cardwell, W., & Mark, L. (2009). Perceiving multiple affordances for objects. *Ecological Psychology*, *21*, 185–217.