Intentions and Intention Recognition in Intelligent Agents

Dr. Gary Berg-Cross
Knowledge Strategies
gbergcross@gmail.com

Dr. Christopher Crick
Brown University
chriscrick@cs.brown.edu

1. INTRODUCTION

PerMIS 2009 included a special session that explored R&D work using the Theory of Mind (ToM) concept. Simply stated the ToM hypothesis is that intelligent agents attribute mental states to other agents in order to reason in a theory-like fashion about the causal relation between these unobservable mental states and the agents' subsequent behavior [53]. Such theories grow in part out of the consideration of the richness and complexity of primate social interactions, which have long been seen as a driver for the evolution of primate intelligence [34, 33]. Child research also suggests that as human infants develop they use knowledge of their own mental function as a model for how other agents function. When infants see others acting 'like me,' they construct and test a representational correspondence hypothesis that others have the same mental experience generating their behavior [44]. This is enhanced by the regularities of perceptions and actions of social interaction, where others act as if they are governed by a similar type of mind. Having a ToM is readily useful because it affords the possibility of profitably applying judgments, originally made about one's self, to others. The PerMIS session explored whether the ToM hypothesis can be testing and if the concept is useful to the goal of highly competent systems able to achieve goals in a relatively autonomous way [6].

The use of constructs like ToM follows a broad direction of research for intelligent systems (IS) and robotics. Over time we have seen a movement from deliberative robots, rooted in an early AI model, to one of reactive robots, followed by behavior-based robots and more recently intentional and motivational robots [42]. It is the potential of this latest thrust into intentionality and "intentional robots" that are discussed in this paper. Intentional plans and intentionality in the everyday sense of pursuit of plans and goals has long played a key role in "the folk ontology of mind". Starting with a vaguer formulation serving goal satisfaction, the concept of intention has matured through its long application in robotics, AI, cybernetics and IS. In part because of its clarified role in rational behavior, intentionality becomes a key concept in a wide range of disciplines from cognitive science to psychological theory. Indeed within the last ten years the development of intentional action and its understanding provided an integrative element to research as diverse as: imitation, early understanding of mental states and their properties, ToM and the recognition of others' intentions. Other work includes studies of goal-directed behavior in nonhuman animals, executive function, language acquisition, play and narrative understanding [68]. A good example of intention understanding within a ToM in animal cognition is Clayton et al's [15] study of scrub jay hiding and stealing cached food. In their study one group of scrub-jays sees another bird caching food and then is given an opportunity to make off with those caches. A second group is allowed to observe food caching, but has no opportunity to pilfer the caches. Several months later, the same 2 groups of scrub-jays are given the opportunity to cache food themselves under 2 conditions – either when observed or when alone after which they are allowed recover their cached food. Clayton et al. [15] report that only the birds with pilfering experience re-hid their caches in new sites. Such hiding would function to prevent future theft, and a parsimonious explanation for why only those birds with pilfering experience do this is because they exhibit a theory of other jay intentions and project (as in a ToM) this experience onto other scrub jays. This amounts to predicting future behavior based on their "knowledge" of likely intentions.

This paper reviews some of the thinking on, and evidence for, the importance of intentionality in cognitive agents. The remaining paper is organized into 4 following sections. First we provide a baseline review of developmental and primate work on intentions and intention recognition. These provide a firmer basis for understanding the role of intentions in cognition and why it is important for an IS. Following this we discuss some of the psychophysics of intentionality. The next section discusses intentionality involved in cognitive-social robots. A particular applied and practical focus is how to better understand the current and projected application of theories and models of biological intention to build intelligent adaptive systems that afford intention. Within this we discuss how useful intentions and their recognition appear in agent social interactions. As part of a developmental robotics approach, we sketch out a development architecture that incorporates intentions and protoplans. A final section sums up the discussion and provides some thoughts on the direction of future work.

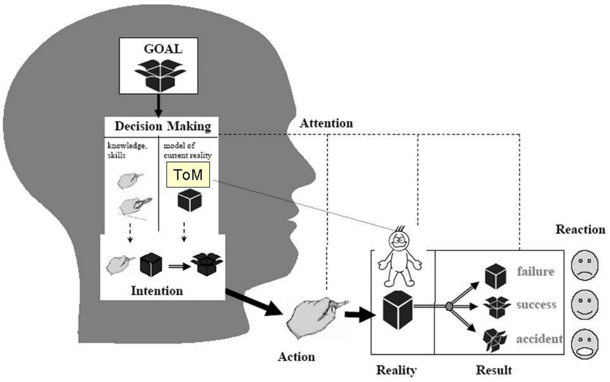

Figure 1 Simple model of Human intentional action (adapted from Tomasello et al 2005).

Figure 1: Simple model of human intentional action (adapted from Tomasello et al 2005)

2. REVIEW OF PRIMATE AND DEVELOPMENTAL WORK ON INTENTIONS AND INTENTION RECOGNITION

We start with a simple model in Figure 1 showing a central role of intentions integrating human perception, actions and problem solving (from [61]. Depicted cybernetically there, intentions support integrated perception, actions and decision making (e.g. choosing proper sub-goals from alternatives) which use models of current reality (e.g. environmental constraints) and goals (part of the mental constraints). At its extreme such models would include a ToM which could help orchestrate social interactions. This is shown as ToM ibeing used inside the Decision making process as a model of the current reality of an external person who is the subject of attentional focus. An agent's goal is shown an open (meaning extendable or modifiable in the future) box; reality is a closed (unchanging) box.

The actor chooses a means (plan), depicted as small hands doing things, which forms an intention. The resulting action (now the big hand) causes a change to reality, which leads to a reaction from the actor (shown as emotion icons). What Figure 1 depicts is an expansion of Bratman's [10] commonsense interpretation of human reasoning that an intention is a form of a plan – one of action that an organism chooses and commits itself to in pursuit of a goal. An intention thus includes both a means (action plan) as well as a goal. This idea of intentions was concretized as part of Bratman's belief-desire-intention (BDI) paradigm. The BDI framework has inspired designing in intentions as part of intelligent agents in an effort to achieve some balance between deliberative processes (deciding what to do) and means-ends reasoning (working out how to do things). In such BDI architectures, however, a relatively simple interpretation of intentions for traditional, directed problem solving domains must be hardwired in. A more dynamic model is represented in Figure 1. In this conceptualization things labeled (internal) goal, action, and attention (aka perceptual monitoring) are all seen as components within a larger, recursive cybernetic/adaptive system that serves to regulate an agent's behavioral interactions with the environment. Such a model allows intentions to play a central organizing role which provides a basis for a range of things form parsing perceptual input, understanding language, making moral judgments, interpretation linking external states and actions with internal ones, and predicting the future behavior of agents [4]. This is useful, for example, in the social-perceptual sphere since without intention recognition it is not evident how the exact same agent movements may be understood as "moving", "giving" or "loaning" an object. This recognition depends on the goals and the intention model which an agent consults along with its stored knowledge/skills and its mental model of current reality. This entire cognitive complex is "relevant" to recognizing the goals of action. A common example of misunderstanding goals occurs with young children where a caretaker may see a child's orientation to something like a box and think that the child desires something in the box. If the caretaker opens the box to give the child the contents, a negative reaction of the child may indicate that opening the box itself was the goal or part of the goal. In this and other cases the agent's chosen action is internally "rational" to the extent that it effectively accommodates the complex of knowledge, skills, and model of current reality.[1]

Taken as a whole then it seems plausible that we have a broad role for intentions in meaningful, social thought and motivated behavior that is beyond its use as simple tool for goal satisfaction. Two ways to investigate and understand intentional ability better are to investigate its evolutionary/phyologenetic development and how it develops in children ontogentically. Each of these research areas includes methods that do not rely on language expression, although those that do obviously show significant difference between verbal children and non-human animals. Early work relied on verbal reports and anecdotes, which were used to capture intuitions on intentional hypotheses [51]. More recently tight non-verbal, behavioral measures, such as gaze, have firmed up a converging, empirical and conceptual basis for intentionality. Likewise intention's relations to other concepts have come into sharper focus through its role in social cognition and the interpretation of everyday events [61]. Thus at least some intentions result in typical concrete behavioral action (represented by the large hand icon in Fig. 1) which are often accompanied by signifiers of persistent, purposive effort and/or evidence of attention [61]. This is usually clearest in social events involving agents that routinely engage objects and other agents. Such experimental results support a commonsense view that social cognition and what we know about others – the ability to perceive, interpret, and explain the actions of others – relies on the use of intention and intentionality to structure agent interactions which are understood as purposeful.

Increasingly studies directed at understanding intentionality compare the behavior of non-human primates with that of human children [64]. As previously noted, principle reasons to investigate intentions and intention recognition in non-human animals is to better understand its evolutionary

---

[1] Tomasello et al (2005) notes that 3 types of engagement emerge sequentially:

1. Dyadic action engagement: Sharing behavior and emotions.
2. Triadic engagement: Sharing goals and perception.
3. Collaborative engagement: All of the above plus joint intentions and attention.

roots and to check specific ideas and hypotheses. Generally speaking, an evolutionary argument for the emergence of animal cognition provides that a richer mental or inner world allows those animals so endowed to simulate a number of different actions ahead of time in order to test out possible consequences and evaluate them [3, 32, 29]. It is, for example, a tool for future-directed cooperation which involves dynamically generated, new goals that lack fixed values. Looked at this way the representation of intentions may serve as a bridge from the present to future possibilities as it embodies a form of symbolic, independent representations the coordinates means and goals [11]. On the whole, however, opinions remain divided on the question of whether non-human primates have a full, working theory of intentionality [24]. The chimp Sarah's success at solving problems after watching videos of a person solving various problem (getting out of a locked cage) lead Premack and Woodruff [54] to interpret this at evidence that she understood the objectives of the human in the films. The alternate explanation is that Sarah was less aware and just stored associations between the pictures and the problematic situations without having an understanding of human intentions.

The previously cited work of Tomasello et al [61], however, is illustrative of subsequent work on these issues, suggesting a degree of continuity in higher processes between humans and non-human primates. The rational "imitation" ability of agents' behavior seen by Gergely et al. [26] in young children is apparently shared to a large extent by chimpanzees [12]. Both children and non-human primates can be tested in naturalistic social settings with subjects being nonverbally requested to help the experimenter achieve goals, such as picking up a dropped sponge or opening a box. Earlier studies by Call et al. demonstrate that chimpanzees, like children, are able to distinguish between a person unwilling to perform a task, from one who is unable to help [13]. Tomasello et al. [61] advanced this work by a chimpanzees study using a food test for intention recognition. In the procedure a human began giving food to an ape through a hole in a transparent wall, but sometimes refusing to give it to the ape and sometimes attempting to give it to the ape but "failing" in the attempt. The chimpanzees gestured more and left the area earlier for humans in the unwilling situations than the unable. In the unable situation chimps tended to wait patiently throughout the unsuccessful attempts. The interpretation of these results supports the hypothesis that chimpanzees have some representation of the intentions of others – they understood the intention behind the human's failure behavior. Chimps also imitate human behavior on acts when they are successful in achieving a goal and not when they fail [49] . However, as Gardenfors [24] points out a limitation of the experiment is that what was tested was whether an action is intentional or not. It does not directly test show that the chimps had a mental representation of the contents of a specific intention. Such representations, Gardenfors believes, are a capability characteristic of human children but not of apes. As he notes it remains "for the development of new experimental paradigms before it can be judged to what extent other species understand the intentions of others." Interest in such better measures is consistent with the theme of PerMIS in pursuing a more rigorous definition of intentions and how they are recognized.

One limitation of apes seems to be some aspect of language involved in shared or social intentions. Warneken & Tomasello [64] have been able to show that chimpanzees, like 18 month-old children, recognize the intent of simple requests (e.g. picking up a dropped object) about equally. But children are able to respond to more complex language-like requests. This limits the full range of collaborative activities to human children. Shared intentions (important for games) may emerge when individuals, who understand one another as intentional agents and interact socially. These may become collaborative interactions where participants have a shared goal (such as building towers with blocks) and coordinated action (such as one child taking the block-gatherer role as a help to pursue the overall shared goal of building a tower).

Methodologically both the child and animal studies have expanded ways in which we can identify both intention-in-action [58] and well as intention recognition through performance. Intention-in-action roughly corresponds to judging the biological plausibility of observed self-executed motor sequences (e.g. an effector tracing a continuous spatiotemporal path towards the object to be manipulated) to decide they are evidence for agent intentions. Early studies by Newtson & Engquist [50] demonstrate that human adults watching videoed action show a high degree of interrater agreement as to the beginnings and endings of intentional actions. Subsequently this finding, supported by child studies using infant looking times as evidence for perceived boundaries in a behavior stream, shows that they have similar boundaries to those rated by adults [1]. And we now know that even young children (18 months) readily distinguish between such intentional and unintentional behavior; identify the intentions underlying others' behavior; explain completed actions with reference to intentions, beliefs, and desires; and evaluate the social worth of actions using the concepts of intentionality and responsibility [41]. Developmentally there is now a large amount of evidence for a theory of infant development reflecting the type of inter-subjective behavior (i.e. inter-affective, inter-attentional, and inter-intentional) afforded by the cognitive architecture shown in Figure 1 [60]. There is evidence for a natural sequence to these. For example, communicative intentions can be recognized in others' behavior before the content of these intentions is accessed or inferred [20]. Intention recognition uses all of the processes shown in Figure 1 starting with an infant's recognitions that others as agents are capable of spontaneous actions – i.e. acting animatedly [67]. This is followed in 9-12 month olds who infants understand the basics of goal-directed action. They recognize that others are pursuing goals and that they will persistent upon failure, accidents or around obstacles. They seem to understand that goal success means that directed actions will stop. Later they understand that others are rationally choosing between which of various plans to implementat, and that this is an intentional act that fits with perceived reality as depicted in Figure 1. Intentionality thus provides the interpretive framework to explain why we perceive agent behavior as humans do. Children, for example, are skillful in social discernment of what others are perceiving, intending, desiring, what they know and believe. There are a wide range of phenomena to consider, but developmentally it can be argued that the foundational skill is that of understanding intentions.

Intention recognition (IR) is the process by which an agent becomes aware of the intent of others. IR is clearly important because as connected devices and intelligent agents advance and multiply we face an increasing coordination challenge. IR will be necessary for truly smart agents that anticipate needs to help agents negotiate over perceived reality and action plans.

Simple psychophysical processes that underlie the detection of others' intentional states appear to be irrepressible instincts that develop quickly in children, around the age of 9 months [55]. Through habituation studies of infants watching animated circles moving about on blank screens, Rochat, Striano and Morgan found that "infants do appear to become increasingly sensitive to subtle changes potentially specifying intentional roles" in abstract, low-context displays. When shown a video in which one circle chased another, the older (8-10 month) children were more likely than the younger ones to notice when the circles' roles (chasing vs. fleeing) suddently switched.

This low-level processing skill develops early in children, and is accompanied by the development of other social skills [21, 27, 22], such as the attribution of agency and intentionality. Csibra found that children (again, at nine months but not at six) were able to interpret goal-directed behavior appropriately. An experimental group was shown a small circle proceeding across the visual field toward another circle, making a detour around an interposing rectangle. When the rectangle disappeared, the infants expected to see the circle move straight at the other, and were surprised when the circle moved in the same detour path that it took around the obstacle.

Csibra and his group conducted several other experiments to control for various confounds, such as the possibility that infants merely preferred the circle's straight-line approach, all else being equal. For example, by habituating the children to a circle that did not take a direct line toward a goal when one was available, they found that the differential dishabituation noted above disappeared. This, they reasoned, was because the infants no longer believed the behavior of the circle to be rational and goal-oriented. The social cognition undertaken by these children depends entirely upon their interpretation of the relative motion of these simple figures.

Attempting to determine and to model when and how typical children develop the social skills that these animations demonstrate is one avenue of research; another emerges when one looks at what happens to people whose ability to do so is impaired. Heberlein and Adolphs [30] investigated the performance of one such person, a victim of the rare Urbach-Weithe disease, which causes calcifications in the anteromedial temporal lobes [59] and completely ravaged the subject's amygdala and adjacent anterior endorhinal cortex. Asked to describe a simple animation similar to those used by Heider and Simmel [31], she used language that almost completely ignored any of the social or intentional implications that control subjects invariably mentioned, treating it entirely in terms of the abstract motion of abstract shapes. However, her inability to anthropomorphize reflexively in this way was not due to a general social deficit – she was able to describe social scenes involving dogs and children perfectly well. The authors conjecture that her amygdala damage impaired a deep automatic social-attribution reflex, while leaving intact her ability to reason and deliberately retrieve declarative social knowledge.

Researchers working with individuals on the autism spectrum have used Heider and Simmel animations to uncover similar social attribution deficits. Ami Klin [36] presented the animation to twenty each of autistic, Asperger's and normally developing adolescents and adults, and gave them each the opportunity to describe and answer questions about their interpretation of the shapes' activity. He found a marked difference between the tendency of the clinical groups and the controls to attribute agency and intentionality to the shapes, and to situate the scenario in a social milieu. For example, a typical narrative from a normally developing adolescent: "The smaller triangle more like stood up for himself and protected the little one. The big triangle got jealous of them, came out and started to pick on the smaller triangle." In contrast, an autistic adolescent with comparable verbal IQ: "The small triangle and the circle went around each other a few times. They were kind of oscillating around each other, maybe because of a magnetic field." In all, Klin found that the clinical groups noticed only about a quarter of the social elements usually identified by the controls.

3. THE PSYCHOPHYSICS OF INTENTIONALITY

The study of intention recognition draws from and contributes to investigations of the fundamental cognitive processing modules underpinning perception and interpretation of motion. Sixty years ago, the Belgian psychologist Albert Michotte performed hundreds of experiments investigating our tendency to attribute causality to low-context motion [46, 47]. Using a primitive animation apparatus consisting of a rotating disc, some number of colored lines painted in various swirls, and a slit which caused the lines to appear as points to the observer, he found and characterized a number of irresistable perceptual effects. For example, when one animated circle approaches another, stationary, circle, and upon reaching it stops its motion while the previously fixed one begins to move, observers invariably perceive the second circle's motion to caused by the impact of the first, an effect Michotte named "launching".

We appear to possess special-purpose perceptual modules responsible for our rapid and irresistable computation of physics-based causality. Gelman [25] and Tremoulet [62] showed that trajectories alone are enough to stimulate the perception of animacy or inanimacy. When shown small particles moving about, subjects reported greater animacy from objects that made large speed or direction changes, especially if such changes violated our physical intuitions about Newtonian mechanics. True, context has an effect – when subjects had access to orientation information as well as trajectories, they made somewhat different determinations. In later work, Tremoulet attempted to investigate the role of context more thoroughly [63]. Subjects watched a particle move along an angled trajectory, while another object was also present. That object's location affected the perception of the particle's animacy: if the particle's trajectory appeared to turn either toward or away from the object, the event looked much more animate and intentional.

Cognitive neuroscientists have looked for evidence of these animacy processing modules in the brain. Blakemore [7] conducted an fMRI experiment where subjects watched an animation where two objects interacted. One object colliding with another, causing a Michottean launching effect, activated the middle temporal gyrus. When the objects changed direction of their own accord, consistent with animacy perception, subjects' right lingual gyrus lit up. And one object orienting itself towards the other's motion, interpreted as attention, involved the superior parietal lobe. These results lend additional credence to the idea that these sorts of perceptions are the result of localized, bottom-up neural processing.

Another demonstration of animacy processing in the brain comes from an fMRI experiment by Wheatley [65]. Her subjects watched an object move on a cartoon background with recognizable features. A simple triangular object moved in a figure-8 pattern, in front of a background that looked either like a frozen lake or a tabletop strewn with toys. The idea was for the same motion to be taken as animate in one context (an ice skater on a pond) and inanimate in another (a spinning top). And indeed, nearly all of the subjects reported the expected bias, given the background. Strikingly, the whole suite of brain regions that have been identified as contributing to social cognition, from the amygdala to the medial frontal gyrus, was much more active when the stimulus was taken to be animate. The authors speculate that animacy perception primes the whole apparatus of social cognition.

If this is true, a deficit in perceiving animacy may disrupt all manner of social development. Rutherford [56] reports on experiments with autistic children that suggest that animacy perception in individuals with social deficits is less reflexive and automatic than for typically developing children and those with non-social developmental disorders. Rutherford presented these three categories of children (typically-developing, socially-unimpaired but developmentally disabled, and autism spectrum) with animations of moving circles. In the control condition, the children were asked to determine which of two balls was heavier, using Michottean stimuli such as one ball rolling towards another at high speed and hitting another, which rolls away slowly. All three groups of children learned to distinguish the heavier from the lighter ball at about the same rate. In the experimental condition, where, for example, one circle moves out of the way of another, the children were rewarded for picking the circle that moved in animate fashion. The autistic children took, on average, twice as many trials as the typically developing children to learn to distinguish animate from inanimate. Interestingly, though autistic children had a harder time with the animacy task than the weight task, typical ones picked up on animacy much faster. However, once the children had learned to distinguish animate from inanimate motion appropriately, there was no difference in the performance of the various groups when faced with previously-unseen test data. Thus, these animacy cues are available to perception, and can be learned, even by autistic children. Modeling the interpretation of these perceptual cues, as our system does, may help illuminate both the initial difficulties and the learned compensatory mechanisms observed in autistic individuals.

These modules appear responsible for our rapid and irresistable computation of physics-based causality [14], as well as facile, subconscious individuation of objects in motion independently of any association with specific contextual features [39] [57] [48]. Furthermore, different processing modules appear to attend to different levels of detail in a scene, including global, low-context motion [40].

4. INTENTIONS IN COGNITIVE-SOCIAL ROBOTS

These psychophysical results are immediately useful in designing computational models of intentionality. Pantelis and Feldman [52] designed artificial agents governed by simple interaction rules, and present behaviors to test subjects for evaluation. They thereby determine how humans assign intentions and personality to behaviors under explicit programmatic control. For example, perceptions of hostility vs. friendliness depended very much upon the agents' behavior with respect to others at a very specific and narrow radius of interaction.

Peter Todd and his colleagues have used simple animations of the Heider and Simmel variety for testing simple social classification heuristics [28, 2]. They asked subjects to generate animations via a game: two people sat at computers and controlled two insect-like icons using a mouse, producing simple socially significant scenarios such as "pursuit and evasion", "play" or "courting and being courted". Using these animations, they tested the ability of adults and children to categorize the interactions properly. They discovered that three-year-old children are able to distinguish, say, chasing from fighting, at rates above chance, and four- and five-year-olds perform better still. A sample of German adults was able to distinguish the six different tested intention regimes around 75% of the time, significantly better than children. Furthermore, the authors performed a cross-cultural study, looking at the categorization performance among Amazonian hunter-gatherer tribespeople. Their scores were nearly identical to those of the German adults, suggesting that the ability to intuit intention from this low-context data is not merely a cultural artifact.

Crick and Scassellati [16, 17, 18, 19] investigated intention recognition in real-world scenarios of actual playgraound games. Their system generated genuine narratives to explain the roles, goals, intentions and rules to in-progress playground games. In addition, the simplicity of the stimuli and their computational tractability enabled them to implement an intention-recogntion system as a robot controller, enabling them to explore the relationship between *watching* a game and *participating*. When taking an active part, the system was able to probe uncertainties in its learning, collapse ambiguity by performing experiments, and explore how motor control relates to social interaction [66].

Computational intentional models such as these are an active area of development. Whether with real robots such as Crick's, or in simulated worlds such as Kerr and Cohen's CAVE model [35], a fundamental question is how intentions develop within a cognitive architecture. Recently the development and use of intentions has been investigated by ontogenetic developmental robotic (DR) approaches. A DR approach takes its inspiration from developmental psychology and developmental neuroscience as opposed to the earlier behavior-based robots whose design includes fixed "motivations" that are hardwired in their structure. DR work includes studying prolonged epigenetic development to elucidate general mechanisms of intelligence development, starting with proposed cognitive development mechanisms such as motivation, shared attention, intention recognition etc. Taken as a whole DR work provides a potentially interesting and useful perspective on how to build an artificial adaptive agent, as well as to better test particular developmental the-

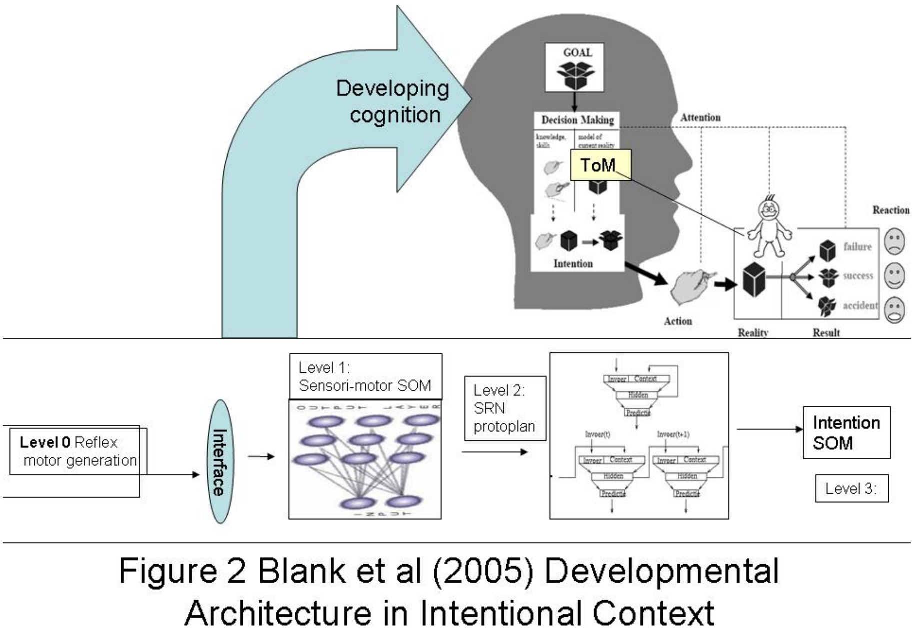

Figure 2: After Blank et al (2005)

ories including intentions. Typically work involves embedding a suitably adaptive[2] and developmentally capable IS within an ecological context so it can learn in a dynamically changing environment filled with multiple streams of raw, uninterpreted sensory information. Developmental agents are endowed with general abilities such as hierarchies of abstraction capabilities that over time enable it to control its actions and focus its attention on the most relevant features of its environment. Based on these abstractions, the robot will be able to predict how its environment will change over time, and go beyond simple reflex behavior to more purposeful behavior. DR sees this as a natural part of a developmental process driven by internally-generated motivations and instrumental tools such as ToM. These push and pull a developing agent toward ever higher-level abstractions and more complex predictions. The idea is that, as an ongoing ontogenetic development of an agent's control structure proceeds, it constrains and simplifies learning enough be make them tractable [45]. Just this idea of an agent's ongoing emergence and continued refinement of its skills has been featured at previous PerMIS workshops [5, 8]. While the role of intentions was not a specific focus it is easily incorporated into the existing architecture vision. What this work fleshes out is how a hierarchical, "developmental architecture" (Figure 2) embodies related algorithms and bottom up processes, generating abstract intentional plans from interactions with the environment. As shown in the figure, the agent starts out at Level 0 with just reflex endowments that are analogs of Piaget's first sensory-motor circular reactions, as well as some specific higher levels that look down on the lower levels and recursively learn from their activity as the agent engages with the environment. The robot is driven by two competing motivations – to avoid boredom and achieve control over its actions. Over time these are afforded by the higher levels. Level 1 is made of a self-organizing map (SOM) which maps a "high"-dimensional input vector to a cell in a lower-dimensional matrix [37]. In this architecture the SOM "observes" the sensor and motor values that are produced as Level 0 controls the robot. Through this observation, Level 1 begins to form associations between sensors and motors and abstractions about sensors and motors within its self-organizing map. Eventually, when Level 1 has successfully captured the control information from agent interactions, these more effectively control the developing robot. A Level 2 structure observes the sensor/motor associations developed within the self-organizing map of Level 1. This structure allows the robot to anticipate events, enabling the robot agent to predict its own future. Level 2 uses a simple recurrent network (SRN) architecture [23] to recognize sequences of sensor-motor associations through time, enabling prediction what the next Level 1 state will be given its current state.

Previous work [43] has shown that this type of simple recurrent network develops representations of multi-step behaviors that might be termed protoplans. One nice thing about this architecture is that engagement with other agents allows the developmental architecture to create protoplans that reflect the actions and others. This in turn is built on by Level 3 which observes the protoplans developed by Level 2 and uses algorithms to categorize them, forming abstract version of plans. Driven by the need to master interactions with other intelligent agents these could include other agent intentions to predict agent behavior, and could leverage models of their own behavior to do this. Taken as a whole the developmental architecture of Blank et al [8] provides a parsimonious, if preliminary, model by which

---

[2] E.g. an intelligent agent should be motivationally autonomous and include adaptive control structures that manage dynamic interactions with its environment.

intentions recognition may be formed developmentally in a robot. It also provides insight into a cognitive architecture that can support and generate a ToM.

5. SUMMARY AND DISCUSSION

This final section sums up the discussion and provides some thoughts on the direction of future work. In this paper we argued for a central role of intentions and intention recognition in IS work. In the first section, we discussed some of baseline work on primate and child cognition studies that provide a basis for this role. In the second section we outlined studies on the neurological, psychological, and physiological foundations of intentions and IR. Follwing this we discussed some efforts to apply these ideas to cognitive robots and sketched out one developmental architecture that indicates how such a capability may emerge. We are still early in the work of development of computational models of intention-recognition. However Bello et al [4] has already shown how consideration of intentionality requirements focuses principled design choices for the construction of IS. E.g.:

- The relation of intention to other mental concepts self reflection, joint attention, communication, imitation
- Modular vs. explanatory theory formulations of intentional robots.

It should be noted that Kozma et al. [38] also provide some initial, dynamic models as ideas for how a ToM and intentions may emerge in human cognition and be implementable in intentional robots.

Questions that remain are how important is intention to levels of autonomy and other basic issues in robotics. These await formal testing and are obviously subject to clarification of autonomy levels which is not represented in either of the architectures of Figure 1 or 2. Autonomy and even several levels of autonomy is implied, however, since such an agent's behavior is directed based on initial learning experiences in specific aspects of their environment. It can be expected that work in DR will provide some useful data on this issue. Of particular use would be an applied and practical focus showing how to better understand the current and projected application of theories and models of biological intention to build intelligent adaptive systems that afford intention.

Consistent with the theme of PerMIS, research and development on intentions needs to pursue more rigorous definition of intentions and how they are recognized. A start on this has been made, but much more development is needed. We believe that development of more rigorous metrics to quantify, and indeed to detect and recognize, intent in man-made devices would be an important step towards the benchmarking progress. An obvious challenge is that measures of such rationality may not be obvious to an outside observer because they are not observable, at least before actions. This makes intention-centered systems a challenge to naive performance models and thus a very interesting topic for PerMIS consideration. Boesch [9], for example, addressed the methodological problem in many study comparisons between humans and other apes. He notes that in tests of social cognition, humans are used as stimuli and caregivers are present when during testing human infants. Boesch concludes that we still need more studies on the cognitive capacities of humans applying the same conditions as those used for animals before we conclude on "evidence" for the absence of cognitive abilities.

6. REFERENCES

[1] D. Baldwin and A. Baird. Action analysis: a gateway to intentional inference. In P. Rochat, editor, *Early Social Cognition*, pages 215–240. Lawrence Erlbaum Associates, 1999.

[2] H. Clark Barrett, Peter M. Todd, Geoffrey F. Miller, and Philip W. Blythe. Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, 26:313–331, 2005.

[3] L. W. Barsalou. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22:577–609, 1999.

[4] Paul Bello, Nicholas Cassimatis, and Kyle McDonald. Some computational desiderata for recognizing and reasoning about the intentions of others. In *Proceedings of the AAAI 2007 Spring Symposium on Intentions in Intelligent Systems*, 2007.

[5] Gary Berg-Cross. Developing knowledge for intelligent agents: exploring parallels in ontological analysis and epigenetic robotics. In *Performance Metrics for Intelligent Systems (PerMIS)*, 2006.

[6] Gary Berg-Cross. Is an agent theory of mind (tom) valuable for adaptive, intelligent systems? In *Performance Metrics for Intelligent Systems (PerMIS)*, 2009.

[7] S.-J. Blakemore, P. Boyer, M. Pachot-Clouard, A. Meltzoff, C. Segebarth, and J. Decety. The detection of contingency and animacy from simple animations in the human brain. *Cerebral Cortex*, 13:837–844, 2003.

[8] D. Blank, D. Kumar, L. Meeden, and J. Marshall. Bringing up robot: fundamental mechanisms for creating a self-motivated, self-organizing architecture. *Cybernetics and Systems*, 36:125–150, 2005.

[9] C. Boesch. What makes us human (homo sapiens)? the challenge of cognitive crossspecies comparison. *Journal of Comparative Psychology*, 121:227–240, 2007.

[10] M. E. Bratman. *Intentions, plans and practical reason.* Harvard University Press, 1987.

[11] I. Brink and P. Gardenfors. Co-operation and communication in apes and humans. *Mind and Language*, 18:484–501, 2003.

[12] D. Buttelmann, M. Carpenter, J. Call, and M. Tomasello. Enculturated chimpanzees imitate rationally. *Developmental Science*, 10:31–38, 2007.

[13] J. Call. Inferences about the location of food in the great apes (pan paniscus, pan troglodytes, gorilla gorilla, and pongo pygmaeus). *Journal of Comparative Psychology*, 118:232–241, 2004.

[14] Hoon Choi and Brian J. Scholl. Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. *Perception*, 35:385–399, 2006.

[15] N. S. Clayton, J.M. Dally, and N. J. Emery. Social cognition by food-caching corvides: the western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society (Biological Sciences)*, 362:507–522, 2007.

[16] Christopher Crick, Marek Doniec, and Brian Scassellati. Who is it? inferring role and intent from agent motion. In *Proceedings of the 11th IEEE Conference on Development and Learning*, London UK, 2007. IEEE Computational Intelligence Society.

[17] Christopher Crick and Brian Scassellati. Inferring narrative and intention from playground games. In *Proceedings of the 12th IEEE Conference on Development and Learning*, Monterey CA, 2008. IEEE Computational Intelligence Society.

[18] Christopher Crick and Brian Scassellati. Intention-based robot control in social games. In *Proceedings of the Cognitive Science Society Annual Meeting*, 2009.

[19] Christopher Crick and Brian Scassellati. Controlling a robot with intention derived from motion. *Topics in Cognitive Science*, 2:114–126, 2010.

[20] G. Csibra. Recognizing communicative intentions in infancy. *Mind and Language*, 25:141–168, 2010.

[21] Gergely Csibra, Gyorgy Gergely, Szilvia Biro, Orsolya Koos, and Margaret Brockbank. Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, 72:237–267, 1999.

[22] Verena Dasser, Ib Ulbaek, and David Premack. The perception of intention. *Science*, 243:365–367, 1989.

[23] J. L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[24] P. Gardenfors. *Consciousness transitions: phylogenetic, ontogenetic and physiological aspects*, chapter Evolutionary and developmental aspects of intersubjectivity, pages 281–305. Elsevier, 2007.

[25] R. Gelman, F. Durgin, and L. Kaufman. *Causal cognition: A multidisciplinary debate*, chapter Distinguishing between animates and inanimates: Not by motion alone, pages 150–184. Clarendon Press, Oxford, 1995.

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

[27] Gyorgy Gergely, Zoltan Nadasdy, Gergely Csibra, and Szilvia Biro. Taking the intentional stance at 12 months of age. *Cognition*, 56:165–193, 1995.

[28] Gerd Gigerenzer and Peter M. Todd. *Simple heuristics that make us smart*. Oxford University Press, New York NY, 1999.

[29] R. Grush. The emulator theory of representation: motor control, imagery and perception. *Behavioral and Brain Sciences*, 27:377–442, 2004.

[30] Andrea S. Heberlein and Ralph Adolphs. Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Sciences*, 101:7487–7491, 2004.

[31] F. Heider and M. Simmel. An experimental study of apparent behavior. *American Journal of Psychology*, 57:243–259, 1944.

[32] G. Hesslow. Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, 6:242–247, 2002.

[33] N. K. Humphrey. *Growing points in ethology*, chapter The social function of intellect. Cambridge University Press, 1976.

[34] A. Jolly. *The evolution of primate behaviour*. Macmillan, 1985.

[35] Wesley Kerr and Paul Cohen. Recognizing behaviors and the internal state of the participants. In *Proceedings of the IEEE 9th International Conference on Development and Learning*, 2010.

[36] Ami Klin. Attributing social meaning to ambiguous visual stimuli in higher functioning autism and asperger syndrome: The social attribution task. *Journal of Child Psychology and Psychiatry*, 41:831–846, 2000.

[37] T. Kohonen. *Self-organizing maps*. Springer, 2001.

[38] R. Kozma, T. Hunstberger, and H. Aghazarian. Implementing intentional robotics principles using ssr2k platform. In *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007.

[39] Alan M. Leslie, Fei Xu, Patrice D. Tremoulet, and Brian J. Scholl. Indexing and the object concept: developing 'what' and 'where' systems. *Trends in Cognitive Sciences*, 2(1):10–18, 1998.

[40] Jeff Loucks and Dare Baldwin. Sources of information in human action. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, pages 121–126, Austin TX, 2008. Cognitive Science Society.

[41] Bertram F. Malle, Louis J. Moses, and Dare A. Baldwin. *Intentions and intentionality: foundations of social cognition*. MIT Press, 2001.

[42] R. Manzotti and V. Tagliasco. From "behavior-based" robots to "motivations-based" robots. *Robotics and Autonomous Systems*, 51:175–200, 2005.

[43] L. Meeden. *Towards planning: incremental investigations into adaptive robot control*. PhD thesis, Indiana University, Bloomington, 1994.

[44] A. Meltzoff. *Perspectives on imitation: from neuroscience to social science*, chapter Imitation and other minds: the like me hypothesis, pages 55–77. MIT Press, 2005.

[45] G. Metta, F. Panerai, R. Manzotti, and G. Sandini. Babybot: an artificially developing robotic agent. In *From Animals to Animats: the Sixth Conference on Simulation of Adaptive Behavior*, 2000.

[46] A. Michotte. *La perception de la causalite*. Institut Superior de Philosophie, Louvain, 1946.

[47] A. Michotte. *Feelings and emotions: The Mooseheart symposium*, chapter The emotions regarded as functional connections, pages 114–125. McGraw-Hill, New York, 1950.

[48] Stephen R. Mitroff and Brian J. Scholl. Forming and updating object representations without awareness: evidence from motion-induced blindness. *Vision Research*, 45:961–967, 2004.

[49] M. Myowa-Yamakoshi and T. Matsuzawa. Imitation of intentional manipulatory actions in chimpanzees (pan troglodytes). *Journal of Comparative Psychology*, 114:381–391, 2000.

[50] D. Newtson and G. Engquist. The perceptual organization of ongoing behavior. *Journal of Personality and Social Psychology*, 12:436–450, 1976.

[51] Richard E. Nisbett and Timothy D. Wilson. The halo effect: evidence for unconscious alteration of judgments. *Journal of Personality and Social*

Psychology, 35:250–256, 1977.

[52] Peter C. Pantelis and Jacob Feldman. Exploring the mental space of autonomous intentional agents. In *Proceedings of the Cognitive Science Society Annual Meeting*, 2010.

[53] D. C. Penn and D. J Povinelli. Causal cognition in human and nonhuman shape and material. *Animal Learning and Behavior*, 15:423–432, 2007.

[54] D. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4:515–526, 1978.

[55] Philippe Rochat, Tricia Striano, and Rachel Morgan. Who is doing what to whom? young infants' developing sense of social causality in animated displays. *Perception*, 33:355–369, 2004.

[56] M. D. Rutherford, Bruce F. Pennington, and Sally J. Rogers. The perception of animacy in young children with autism. *Journal of Autism and Developmental Disorders*, 36:983–992, 2006.

[57] Brian J. Scholl. Can infants' object concepts be trained? *Trends in Cognitive Sciences*, 8(2):49–51, 2004.

[58] J. Searle. *Intentionality: an essay in the philosophy of mind*. Cambridge University Press, 1983.

[59] Claudio C V Staut and Thomas P Naidich. Urbach-wiethe disease (lipoid proteinosis). *Pediatric Neurosurgery*, 28:212–214, 1998.

[60] D. Stern. *The interpersonal world of the infant*. Basic Books, 1985.

[61] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28:675–735, 2005.

[62] Patrice D. Tremoulet and Jacob Feldman. Perception of animacy from the motion of a single object. *Perception*, 29:943–951, 2000.

[63] Patrice D. Tremoulet and Jacob Feldman. The influence of spatial context and the role of intentionality in the interpretation of animacy from motion. *Perception & Psychophysics*, 68(6):1047–1058, 2006.

[64] F. Warneken, F. Chen, and M. Tomasello. Cooperative activities in young children and chimpanzees. *Child Development*, 3:640–663, 2006.

[65] Thalia Wheatley, Shawn C. Milleville, and Alex Martin. Understanding animate agents: Distinct roles for the social network and mirror system. *Psychological Science*, 18:469–474, 2007.

[66] Daniel M Wolpert, Kenji Doya, and Mitsuo Kawato. A unifying computational framework for motor control and social interaction. *Philisophical Transactions of the Royal Society B*, 358(1431):593–602, March 2003.

[67] A. L. Woodward. Infants selectively encode the goal object of an actor's reach. *Cognition*, 69:1–34, 1998.

[68] P. D. Zelazo, J. W. Astington, and D. R. Olson, editors. *Developing theories of intention: social understanding and self-control*. Erlbaum, 1999.