Heidelberg: Mindreading = Minimal Theory of Mind section (I of II)

§ introduction

Theory of mind abilities are widespread.

From early in their second year children point to inform others in ways that reflect the others' knowledge or ignorance; and they can predict actions based on false beliefs about the locations of objects;

scrub-jays selectively re-cache their food in ways that deprive competitors of knowledge of its location;

and chimpanzees select routes to approach food which conceal them from a competitor's view and retrieve food using strategies that optimise their return given what a dominant competitor has seen.

Note here that I am talking about theory of mind abilities only. It is essential for my purposes to distinguish theory of mind *abilities* from theory of mind *cognition*. **Theory of mind *abilities* are abilities that exist in part because exercising them brings benefits obtaining which depend on exploiting or influencing facts about others' mental states.** Theory of mind *cognition* is cognition of mental states; **it paradigmatically involves ascribing propositional attitudes such as beliefs, desires and intentions in order to give rationalising causal explanations of thought and action.**

This distinction is important because the facts about other minds which theory of mind abilities exploit are not necessarily the facts which are represented in theory of mind cognition. To illustrate, it may be possible to exploit facts about what others perceive by tracking their

lines of sight and so without representing perceptions as such.

I want to suggest that these findings about theory of mind abilities give rise to a puzzle. The puzzle is about how to explain the theory of mind abilities manifested by young children, chimpanzees and scrub jays. **What do these agents represent that enables them, within limits, to track others' beliefs and other propositional attitudes?**

The most straightforward answer would be to suppose that they represent perceptions, knowledge states, beliefs and other propositional attitudes.

But I don't think this explanation can be right because.[1] this kind of *theory of mind cognition is hard (in two senses).* A body of evidence with humans suggests that reasoning about beliefs and other propositional attitudes requires *conceptual sophistication*,

> for it has a protracted developmental course stretching over several years, and
>
> its acquisition is tied to the development of **executive function and language**—things which two-year-olds, scrub jays and chimpanzees are deficient in.
>
> Development of reasoning about beliefs in humans may also be **facilitated by explicit training and environmental influences**, such as siblings.

---

1 *Lack of systematicity*. Eg two-year-olds generally fail to take into account what their addressee's can see when choosing referring expressions (but three- and four-year-olds choice of referring expressions do take into account what the addressee can see) {Matthews, 2006 #1781}. I wouldn't want to put too much weight on this consideration because there are well-known cases where adults fail to take into account what their addressees can see {e.g \Keysar, 2003 #1782}.

Ascribing propositional attitudes also appears to be ***cognitively demanding***, requiring **attention** and **working memory** in fully competent adults.

It makes sense that propositonal attitudes should be conceptually and cognitively demanding. After all, these are states which form **complex causal structures, have arbitrarily nest-able contents, and are individuated by their causal and normative roles in explaining thoughts and actions.** If anything should take years to acquire and consume scarce cognitive resources it is surely states with that combination of properties.

So the puzzle is this:

> Abilities to solve tasks which hinge on facts about what others see, know and believe are widespread
>
> The most straightforward way to explain these abilities would be to suppose that the agents in question are representing perceptions, knowledge states, beliefs and other propositional attitudes as such.
>
> But abilities to reason about propositional attitudes appear to be require cognitive resources and conceptual sophistication. This may justify caution in supposing that even human adults' theory of mind abilities always depend on representing perceptions, knowledge states and beliefs as such.

So what do infants, chimpanzees and scrub-jays represent that enables them, within limits, to track others' perceptions, beliefs and other propositional attitudes?

To understand how minds appear to infants and other animals, we need to construct notions which resemble propositional attitudes like knowledge states, beliefs, desires and intentions in some respects but representing which is less conceptually and cognitively demanding.

The approach I am proposing is not entirely new …

§ predecessors

On this question there have already been several intuitve suggestions.[2] But on the whole these suggestions imply signature limits which infant theory of mind abilities have since been shown to exceed. In particular, none of the existing suggestions cannot readily explain abilities to track false beliefs.

Minimal theory of mind is an attempt to build on these suggestions in saying what it is that infants, chimpanzees and scrub jays represent that enables them, within limits, to track beliefs and other mental states.

[*acknowledge Ian]

§ Minimal theory of mind

I will describe it with a series of concepts and principles.

The first concept is the **field**. An agent's field is a set of objects related to the agent by proximity. The extent of

---

[2] Juan-Carlos Gomez has suggested awareness of primitive intentional relations to objects established by gaze may be important (Gomez 2007: 730), Daniela O'Neil and Martin Doherty have discussed a notion of engagement with objects (O'Neill 1996; Doherty 2006), Josep Call and Michael Tomasello have suggested that chimpanzees track the 'likely target' of others' visual access and understand something about its effects on behaviour (Call and Tomasello 2005: 58), and Andrew Whiten uses the notion of an "intervening variable" to explain primitive theory of mind notions (Whiten 1994, 1996). **Add Bartsch & Wellman (our proposal is also similar to theirs.)

the field depends on things like lighting and the agent's orientation. To say that an agent is **encountering** an object just means that it is in her field. The notion of encountering is a proxy for perception: within limits, an agent perceives an object just when she encounters one. But to think about encountering doesn't require understanding perception. **In particular it doesn't require understanding perceptual modalities, perspectives or appearances.**

The next concept is **goal-directed** action. The term 'goal-directed action' can be used to mean several things. One is intentional action. This notion is no use for constructing a minimal theory of mind. To represent intentional actions as such you also have to represent intentions and related propositional attitudes such as belief (Davidson 1999). Constructing a minimal theory of mind requires a basic notion of goal-directed action.

[*presented in two phases. First: have goal-directed actions when have sequence of units of object-directed behaviours where the sequence has a function which is not a function of any individual unit.

*The question is then what we mean by function here. *Smuggling* danger. Either teleological or whatever notion of function infants have from tool use.]

We stipulate that for *g* to be the goal of a unit of object-directed behaviours is for two conditions to be met:

> (i) *g* would be the outcome *of the whole unit* if *g* occurred;
>
> and
>
> (ii) units of this type occur in order to bring about *g* (that is, *g* is the function of this unit).

This is not an account of full-blown goal-directed action. It is what someone who has only a minimal grasp of goal-directed action might understand.[3]

The **first principle** links encountering with goal-direction actions. It says that you can only act on a goal involving an object if you have encountered it. So if you know that someone hasn't encountered an object, you can predict that they won't look for it.[4] Equally, one way to prevent someone from being able to act on an object is to prevent her from encountering it. This principle isn't true, but it approximates a truth and is a useful heuristic for predicting action.

[Explains Level-1 perspective taking]

For the next principles we need the concept of **registration**. This is a relation between agents, objects and locations. An agent registers an object at a location just if she last encountered it there. Registrations can be correct or incorrect. A registration is incorrect when an agent registers an object at a location but it is not at that location.

Principle Two says that correct registration is a condition of *successful* action. If you have a goal involving an object, you will only succeed if you register the object at its actual location. To illustrate, suppose you see someone encountering an object which is then moved while they are not watching it. In this situation, Principle

---

3 This teleological approach to characterising goal-directed action has been developed in different ways by several philosophers including Taylor (1964), Wright (1976) and Millikan (1993).

4 (In *this* experiment by Brian Hare and colleagues, a subordinate chimpanzee makes predictions about a dominant chimpanzee's ability to retrieve food. They found that the subordinate's predictions take into account whether the dominant's view was blocked while the food was placed. This could be explained by the First Principle. For the subordinate to predict that the dominant will not be able to recover the food, it is sufficient to think: because the dominant did not encounter the food, she will not be able to retrieve it.)

Two allows you to predict that they will not be successful in retrieving that object.

One application of this Principle is to some of Liszkowski and colleagues' pointing studies. [*compressed] In one paper these authors showed that children point in order to provide information to others about the locations of objects needed to perform an action. The adults had previously used the objects so they were not unfamiliar to them; rather, they had recently misplaced them. This could be explained on the hypothesis that the 12- and 18-month-old subjects think of pointing as a way to get others to register objects and understand this Principle that correctly registering an object as necessary for acting with it.

Principle Two doesn't enable you to pass false belief tasks. This is because when people have incorrect registrations, this Principle doesn't allow you to predict what actions they will perform. To do this you need **Principle Three**. With this Principle, we move from thinking of registration as a *condition* on action to thinking of it as a *cause* of action. Principle Three says that agents will act as if objects were actually at the locations they register them as being. What determines where an agent will look for an object is not its actual location but the location the agent registered it as being at. This Principle is sufficient for passing some false belief tasks.

This, then, is minimal theory of mind. My conjecture is that two-year-olds and perhaps adults too are sometimes able to track beliefs by means of representing encounters and registrations. Just as you might represent weight in order to track mass (within limits), so you can represent registrations in order to track beliefs (within limits).

Conclusion [of part I, mindreading]

The motivation for introducing minimal theory of mind was a puzzle about theory of mind abilities. Minimal theory of mind demonstrates that it is possible to succeed on a range of theory of mind tasks without representing perceptions, knowledge states or beliefs as such. It also enables us to understand in a systematic way how there could be degrees or kinds of theory of mind cognition. Finally, minimal theory of mind may be what makes those with limited cognitive resources and little conceptual sophistication, such as infants, chimpanzees or scrub-jays, sensitive to facts about perceptions, knowledge states and beliefs.

A natural question is how minimal theory of mind cognition might develop into full blown theory of mind …