Joint Action and Knowing Others' Minds

s.butterfill@warwick.ac.uk

Abstract

How could abilities to engage in joint action foster developments in understanding minds? Several researchers have argued that joint action does in fact foster mentalistic understanding, but few if any have explained in detail how it could do so. In this talk I elaborate a modest conjecture: abilities to engage in joint action provide a route to knowledge of others' goals distinct from ordinary third-person interpretation. The existence of this route may partly explain how joint action fosters developments in understanding minds.

Introduction

the Vygotskian intelligence hypothesis:

> "the unique aspects of human cognition … were driven by, or even constituted by, social co-operation" (Moll and Tomasello 2006: 2–3).

Problem for VIh

This hypothesis faces an immediate problem … by social cooperation these authors do not mean to refer to the behaviours which are merely co-operative and co-ordinated, such as: the fungus farming activities of some ants, (Schultz 1999); the ambush of much larger insects by ants with

carefully constructed traps;[1] and perhaps also the group hunting adventures of Taï forest chimps (Moll and Tomasello 2006).

Rather, social co-operation is to be understood as involving the sharing of goals or intentions.

The leading account of social co-operation, which these authors also endorse is Bratman's account of shared intentional activity which requires shared intentions. And this *presupposes* some "unique aspects of human cognition", in particular it requires intentions about other's intentions:

> for us to share an intention that we J it is sufficient that:
>
> "1. (a) I intend that we J and (b) you intend that we J
>
> "2. I intend that we J in accordance with and because of la, lb, and meshing subplans of la and lb; you intend that we J in accordance with and because of la, lb, and meshing subplans of la and lb
>
> "3. 1 and 2 are common knowledge between us" {Bratman, 1993 #1356@View 4}

So the **first puzzle** we face is how to characterise social co-operation in a way that doesn't (i) reduce to the sorts of co-operative and co-ordinated behaviours found in ants and other non-humans and doesn't (ii) presuppose unique aspects of human cognition.

---

[1] See (Dejean, Solano, *et al.* 2005): "the ant Allomerus decemarticulatus uses hair from the host plant's stem, which it cuts and binds together with a purpose-grown fungal mycelium, to build a spongy 'galleried' platform for trapping much larger insects. Ants beneath the platform reach through the holes and immobilize the prey, which is then stretched, transported and carved up by a swarm of nestmates. To our knowledge, the collective creation of a trap as a predatory strategy has not been described before in ants."

Solution to puzzle

In earlier work I suggested a solution to this puzzle …

The first step is to distinguish between goal-directed activities and intentions. Representing goal-directed activity requires less conceptual sophistication than representing intentions: roughly speaking, for a behaviour to be goal directed is for it to have a teleological or conventional function; the goal it is directed to is its function. Understanding goal-directed behaviour therefore does not require understand intentions or understanding any mental states at all. (The goal is a property of the behaviour rather than of the agent.)

The second step towards solving the puzzle is to focus on a most basic kind of social co-operation. To do this we start with the notion of a **plural activity**.

Plural activities are those in which two or more agents' activities are individually organised around a single outcome which occurs as a common effect of all the agents' activities. Plural activity in this sense can be found in ants and other nonhuman animals.

Successful plural activity generally requires coordination. How is this coordination achieved? In the case of ants such coordination may be achieved **hormonally**. In humans, who can voluntarily engage in plural activities with novel outcomes, coordination can usually only be achieved **psychologically**. This is what shared goals are for. Their function is to coordinate plural activities with novel outcomes.

Which states realise shared goals?

(a) *many goals, one outcome—*
two or more agents' individual goals all require for their fulfilment a single outcome, where the outcome and goal

are related in such a way that it is normally obvious to any agents of these kinds that the goal's fulfilment requires the outcome;

(b) *identification*—
each agent can identify each of the other agents in a way that doesn't depend on knowledge of goals involving the outcome;

(c) *expectations about goal-directed activities*—
on balance[2] each agent expects each of the other agents she can identify to perform a goal-directed activity whose success obviously requires the outcome; and

(d) *expectations about a common effect*—
on balance each agent expects this outcome to occur as a common effect of all of their goal-directed actions.

To illustrate, suppose that we each individually want to move a heavy two-handled basket across the room. Each of us individually has a goal whose fulfilment requires the basket to move across the room; each of us expects the other to lift the basket and for this lifting to serve its function only if the basket moves across the room; and each of us expects that, if the basket does move across the room, it will be a common effect of both of our efforts.

In favourable circumstances the goals and expectations in these four conditions could serve to coordinate the goal-directed plural activities of two or more agents.

… this simpler conception of joint action **resolves the puzzle** because it (i) allows for joint actions directed at entirely novel ends, and (ii) doesn't presuppose more than expectations about others' goals and of common effects,

---

[2] The 'on balance' qualification in conditions (c) and (d) rules out cases where agents do have the specified expectations about goals and outcomes but also have further, conflicting expectations which outweigh them.

which are plausibly widespread in humans and non-humans.

Challenge for VIh

Having removed the obstacle we now need to say in detail *how* social co-operation might drive "unique aspects of human cognition"

Obstacle to responding to challenge (no romantic sharing)

[Short version:] can't literally share a goal or attention in the sense that we can share a bottle of wine; ultimately to talk of sharing in psychological contexts needs to be cashed out in terms of individuals' attitudes; and so—even though there sharing goals can involve distinctive phenomenology—there is no room for romantic ideas about the magic of sharing.

[long version, see v.1] Here I think *instrumentalism* provides an obstacle … [***idea is that we can't take *sharing* as explanatory without saying what we mean by sharing, and when we do it is hard to see how *that* can be explanatory]

So the problem is to say in detail *how* social co-operation might drive "unique aspects of human cognition". Needless to say I have only a very modest suggestion to offer.

My suggestion concerns ascriptions of goals [strictly speaking: identifying the functions of goal-directed behaviours]. Some of the most plausibly unique aspects of human cognition depend on our abilities to recognise the goals of novel behaviours involving tools and gestures. In particular, communicative actions depend on our abilities to recognise as the goals of behaviours goals which the behaviours can serve only because we recognise the

behaviours as serving those goals [the Gricean circle]. It is here that I think joint action can play a role in explaining aspects of human cognition. My suggestion will be that abilities to engage in joint action provide a route to knowledge of others' goals which is distinct from ordinary third-person interpretation.

To explain this suggestion in detail I first need to identify a problem …

Problem of opaque means

What is the problem of opaque means? Suppose you cannot gain knowledge of another's current goals through linguistic communication.

Suppose the goal is relatively novel (there are no stereotypical indications -- you're not a chimpanzee presented with a grape).

In this situation we cannot generally do better to work backwards from her behaviours to her goals: we determine which outcomes her behaviour is likely to bring about and then suppose that her goal is to bring about one or more of these outcomes (Dennett 1991).

But this method doesn't work when:

> we don't know which outcomes the observed behaviour is likely to bring about;
>
> we, or the agent under observation, have false beliefs relevant to which outcomes this behaviour is likely to bring about; or
>
> there are many possible outcomes.

Cases where the method is particularly likely to fail include:

> use of novel tools which we don't recognise as tools or don't know the functions of;

multi-step activities where the steps can occur in various orders and do not form a natural sequence (it takes young children months to work out the need to move a switch in order to enable buttons to cause sounds by themselves [useful example because they pick this up right away if you show them how it works]);

communicative actions where there is no apparent connection between an act of meaning and its intended outcome (e.g. Tomasello, Call and Gluckman 1997; Leekam forthcoming)

The problem of opaque means is especially likely to affect infants and younger children.

Here's how I propose that joint action bears on the problem of opaque means …

The your-goal-is-my-goal route to knowledge

How could abilities to engage in joint action provide us with knowledge of others' goals?[3] The intuitive idea I started with was this: if you're engaged in joint action with me, it's easy for me to know what your goal is … because your goal is my goal.

This intuitive idea isn't quite right as it stands. For to be engaged in joint action requires that I have expectations about your goals [strictly speaking, the goals of your behaviours]. So engaging in joint action presupposes rather than explains knowledge of others' goals. Or so it seems.

But there is a way around this. For there are various cues that you can give me which signal that you are about to

---

[3] I do not mean to suggest that *engaging* in joint action can provide this knowledge; engaging in joint action requires expectations about others' goals and so cannot be a route to knowledge of them.

engage in joint action with me. Seeing me struggling to get my twin pram on to the bus, you grab the front wheels and make eye contact, raising your eyebrows and smiling. In this way you signal that you both disposed to help and are about to engage in joint action with me. This makes it trivial for me to know what the goal of your behaviour is: your goal is my goal, to get the pram onto the bus.

My suggestion, then, is that the following inference characterises a route to knowledge of others' goals:

1. We are about to engage in some joint action or other (for example, because you have made eye contact with me while I was in the middle of attempting to do something).

2. The other agent or agents are disposed to help me with what I am currently attempting to achieve.

3. What I am attempting to achieve obviously requires a certain outcome.

Therefore:

4. The others will each have goals whose fulfilment requires that this outcome occur.

Call this the 'your-goal-is-my-goal' route to knowledge. To say that this inference characterises a route to knowledge implies two things. First, in some cases it is possible to know the three premises, 1–3, without already knowing the conclusion, 4. Second, in some cases knowing the three premises puts one in a position to know the conclusion. I take both points to be true.

The your-goal-is-my-goal route to knowledge is characterised by an inference. However, exploiting this route to knowledge may not require actually making the inference or knowing the premises. Depending on what knowing requires, it may be sufficient to believe the conclusion because one has reliably detected a situation in

which the premises of the inference are true without necessarily being able to think of this *as* a situation where the premises are true.

First illustration of your-goal-is-my-goal: multi-step actions

[Do it for the frog visual aid with the concealed on-off switch]

The relevance of your-goal-is-my-goal to fostering understanding of other minds is most readily seen by putting ourselves in the shoes of adults interacting with young children. You are demonstrating how a toy works and want little Molly to know what your goal is. It is difficult to convey this to her directly. So you first get her interested in the outcome your goal was directed to achieving, knowing that she will adopt what was your goal as her own goal. Then you show Molly in gesture that you are going to act with her. Now she is in a position to know what your goal is. She knows what your goal is not because she was able to infer your goal by observing you acting on it but because you adopted her goal and then expressed willingness to engage in joint action with her. Now you can take a step towards achieving the goal which Molly would normally fail to realise is connected with your goal. Because engaging in joint action gives her knowledge of the goal of your action which doesn't depend on her recognising this action as a means to achieving the goal, Molly is poised to discover that the step you took is a novel means of achieving the goal.

Note what has happened here: your-goal-is-my-goal shifts the problem of inferring goals from the observer to the observed.

Note the structure I'm suggesting: ostensive cues -> shared goal -> learning. The proposal is a friendly extension of views about natural pedagogy.

Second illustration of your-goal-is-my-goal: communicative actions

Prefix by saying that word—object association is not yet to understand word as a referential device, that is as something uttering which can be used to direct people to the object.

The difficulty of making the step is an instance of the problem of opaque means—it is just not obvious that actions such as uttering the word "bed" can be means to direct someone to look for something in a bed, or to perform any number of other bed-related actions depending on the context.

There are two reasons why this fails to be obvious. First, one can associate "bed" with bed without knowing that the two are associated. Second, aside for special cases, uttering "bed" to direct someone to look for something in a bed only works if they recognise that uttering "bed" can serve this function [the Gricean circle].

I want to suggest that abilities to engage in joint action may explain how we make the step from knowing word—object associations to recognising words as referential devices.

It's the step from word—object association to knowing that uttering the word is a way of directing someone to the object. This is the step that joint action helps us make.

I am looking for a teddy and you say "in the bed". I assume that you are helpfully engaging in joint action with me (not trying to distract me from what I am doing) and I associate "bed" with bed. Your utterance thereby takes me

to the bed even without me understanding anything about referential intentions or communicative actions. Within the context of joint action, I don't need to understand your utterance as referential. The mere fact that you have grabbed my attention and triggered the association is enough.

So I do not have to understand your goal prior to our joint action, or to think of your utterance as a referential device, in order for your action to be effective.

But if I am in the business of trying to identify the goal of your utterance of "bed" then the context of joint action can be helpful. For the your-goal-is-my-goal reasoning allows me to infer that you acted on a goal whose fulfilment requirement me to the find the toy. So I have greatly narrowed down the possible range of goals that you could have had in uttering the word "bed".

Key point is that *inside joint action* don't need to understand it [the action] as referential/communicative. But once is has been given a function, it can be used to serve that function outside joint action contexts. And so it becomes genuinely communicative.

NB communicative does not generally occur in the context of joint action.

Here, then, are two ways in which abilities to engage in joint action enable infants to learn about goals despite the problem of opaque means. The general idea is that it is often easier to know that one is engaged in joint action than to figure out what another's goal is—but once one is engaged in joint action, one thereby knows what the other's goal is.

conclusion

The question was to understand how social co-operation might in principle drive "unique aspects of human cognition"

My modest suggestion is only supposed to be a very small part of the answer.

It [my suggestion] is that abilities to engage in joint action provide us with a route to knowledge of others' goals which is distinct from ordinary third-person interpretation. I called this the your-goal-is-my-goal route to knowledge. This route to knowledge is distinct from ordinary third-person interpretation in the following sense: gaining knowledge via this route doesn't depend on being able to identify what others' behaviours are a means to achieving (so the problem of opaque means doesn't arise).

This may be particularly significant for children who have no or limited abilities to communicate by language and lack extensive knowledge of which behaviours are effective means to which outcomes.

Part of the significance of joint action may be that it enables children to identify others' goals, and so learn how to achieve things with words and other tools despite their fledgling linguistic skills and limited insight into how things are done.

[Csibra on referential understanding of gaze]

Why do infants follow gaze etc only following eye contact or other signals?[4] *Because they are signals of cooperation*.

---

[4] "In human infants, however, gaze following also serves communication. This is evidenced by the fact that young infants tend to follow gaze shifts only when these are preceded by an ostensive signal such as eye contact or infant-directed greeting 29 A. Senju and G.

"referential expectation may also originate from conceiving gaze shifts as communicative acts (Csibra, 2003)" {Csibra, 2008 #1791@8}

"infants … may have concluded that the woman on the screen, who apparently attempted to communicate with them, referred to, and implicitly indicated the presence of an object *for* them." {Csibra, 2008 #1791@8}

speculation: "even young infants, who have only limited understanding of the conditions of visual access, show evidence of such an expectation which suggests that infants follow the gaze of others not in order to find out what the other can see but in order to find out what the other is communicating about." {Csibra, 2008 #1791@10}

"early understanding of referential actions evident in 'gaze following' phenomena originates not from a rich comprehension of the link between a mental state of an agent and its referent but from a blind tracking of motion cues in communicative contexts" {Csibra, 2003 #1441@454}

The referential action understanding involves a "stance" (p. 455); teleological and referential action interpretation "rely on different kinds of action understanding" {Csibra, 2003 #1441@546}; they are initially two distinct "action interpretation systems" (although of course they come together later in development) (p. 456).

Csibra, Gaze following in human infants depends on communicative signals, Curr. Biol. 18 (2008), pp. 668–671. Article | PDF (283 K) | View Record in Scopus | Cited By in Scopus (17)[29] (Figure 2). In fact, infants prefer to watch a person making object-directed to non-object-directed gaze shifts, but only if these shifts start from a direct gaze position [30]." {Csibra, 2009 #1790}