Mindreading and Joint Action: Philosophical Tools
Notes for Lecture 9
Interacting Mindreaders

Stephen A. Butterfill
<ButterfillS@ceu.hu>

December 4, 2012

1.   Introduction

Last week I introduced this conjecture:

> The prior existence of capacities for shared agency partially explains how sophisticated forms of mindreading emerge in evolution or development (or both).

2.   slide

Proponents of this conjecture face an immediate objection:

1. All shared agency involves shared intention.

2. Shared intention requires sophisticated mindreading.

Therefore:

3. The prior existence of capacities for shared agency could play no significant role in explaining how sophisticated forms of mindreading emerge.

3.   slide

In the last lecture I explained how we can defuse this objection by rejecting the first premise.

4.   slide

There is a kind of shared agency which doesn't involve shared intention, but rather involves what I called *shared motor representation*. This was confusing because shared motor representations are not linked to muscle control and so, as some people understand the term, not 'motor' representations; and, nor, of course are they shared.

But, terminology aside, the important point is that there are ways in which joint actions can be coordinated which do not involve mindreading but only goal ascription.

5.   slide

The existence of cognitively and conceptually undemanding forms of shared agency matters because it shows that, at least in principle, we can appeal to joint action in explaining the emergence of sophisticated forms of cognition.

6.   slide

I focus on the analysis of relatively minimal kinds of understanding—on pure goal ascription, on minimal theory of mind and on shared motor representation. This is because starting with less potentially allows you to explain more.

7.   slide

If you follow Tomasello in starting with shared intention and shared intentionality, your philosophical tools commit you to starting with thinkers who know about knowledge and intention and nested structures of these.

But other philosophical are available. The notion of shared motor representation allows us to hang on to Tomasello's ideas about the importance of shared agency without having to start from thinkers who already have sophisticated theory of mind cognition.

8.   slide

So far, then, we have shown that the conjecture is coherent. There is a kind of shared agency that doesn't presuppose sophisticated mindreading. Now the question arises: **How could that work?**

How could shared agency partially explain the emergence of sophisticated forms of mindreading.

I can't give anything like a full answer to this question. I will only gesture in the direction of an answer. If you are hoping for an explanation, the gesture will be unsatisfying. But I hope to say enough to at least show that it is worth considering philosophical tools with more minimal committments than those usually employed.

9. slide

I'm going to break this into two steps. The first step is about how shared agency might get you from pure goal ascription to minimal theory of mind.

This really just a warm up, it's just to introduce the main ideas.

Could there be a role for shared agency in explaining the step from pure goal ascription to minimal theory of mind?

10. slide

Recall that pure goal ascription is the process of identifying outcomes to which purposive actions are directed as outcomes to which those actions are directed independently of any knowledge of mental states.

In an earlier lecture I defended George and Gergo's insight that pure goal ascription involves using one's own planning mechanisms to plan other's actions.

To be more explicit, recall:

> Let $R_M(a, G, s)$ be the relation that holds just if: (i) were M tasked with producing G in situation s, then it would plan action a; and (ii) G is desirable.

I defended the view that:

1. Pure goal ascription involves detecting whether relation R_M holds for some M; and

2. detecting whether R_M holds depends on using one's own planning mechanisms.

11. slide

If this is even roughly right, it tells us something about the limits of pure goal ascription. **Limits on our ability to plan actions are limits on pure goal ascription.**

3

12.   slide

One limit concerns false belief.

To illustrate, imagine sitting at a table. On the table are two closed opaque boxes. One box contains an owl, the other a cat. The goal of Ayesha's action is to retrieve the cat, but she believes, incorrectly and contrary to what you know, that the cat is in the North Box. So when she reaches into the North Box, you cannot rely on pure goal ascription to identify the goal of her action. This is because, from your point of view, reading into the North Box is not a means to retrieving the cat. If you had to plan an action to retrieve the cat, you would not plan to reach into the North Box.

As this illustrates, differences in belief between observers and agents can impair goal ascription. This the problem of false belief.

13.   slide

I want to show that capacities for shared agency, together with an understanding of distributive goals, allows pure goal ascribers to avoid the problem of false belief even without ascribing mental states.

How could this work?

14.   slide

Here's an idea that almost but doesn't quite work. Suppose that you were engaged in joint action with Ayesha, and that the distributive goal of your actions was to retrieve the cat. Then you would know that, when Ayesha reaches into the North Box, her goal is to retrieve the cat. You would know this because your goal is her goal, which is to retrieve the cat.

Now this idea doesn't work. Because knowing that the distributive goal of your actions is to retrieve the cat requires that you know the goal of Ayesha's actions. But there is a way around this.

Fortunately there is a way around this. For there are various ways in which I can know that we are about to engage in joint action independently of knowing the goal of your action. This sometimes involves cues which signal that one agent is prepared to engage in some joint action or other with another, and joint actions involve distributive goals. But it doesn't have to involve cues. Sometimes the situation, or our relationship or history allows me to assume, rationally, that you will engage in joint action with me.

Seeing you struggling to get your twin pram onto a bus and noticing you have the haggard look of a new parent, a passing stranger grabs the front wheels and makes eye contact with you, raising her eyebrows and smiling.

(The noise of the street rules out talking.) In this way she signals that she is about to act jointly with you. Since you are fully committed to getting your pram onto the bus, you know what the sole goal of your own actions will be. But you also know that the stranger will engage in joint action with you, which means that, taken together, her actions and your actions will have a distributive goal. This may enable you to infer the goal of the stranger's imminent actions: her goal is your goal, to get the pram onto the bus.

15.   slide

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

1. You are about to attempt to engage in some joint action[1] or other with me.

2. I am not about to change the single goal to which my actions will be directed.

Therefore:

3. A goal of your actions will be my goal, the goal I now envisage that my actions will be directed to.

Call this inference *your-goal-is-my-goal*. To say that it characterises a route to knowledge implies two things. First, in some cases it is possible to know the premises, 1–2, without already knowing the conclusion, 3.[2] Second, in some of those cases knowing the premises would put one in a position to know the conclusion.[3]

16.   slide

Before going any further, note that we are using the notion of joint action in specifying the content of this inference. What conception of joint action must someone have to be able to use this inference?

---

[1] We leave open the issue of how joint action is to be characterised subject only to the requirement that all joint actions must involve distributive goals.

[2] The key point here is that I can know that we are about to engage in joint action because of either (i) cues (like a play face, or expression of anger) or (ii) our situation or history.

[3] Key points: (i) other must have true beliefs concerning your goal; (ii) other must be willing to pursue your goal (but they wouldn't express willingness to engage in joint action without a salient alterantive).

17.   slide

Actually, they don't need a conception of joint action at all, strictly speaking.

18.   slide

Any conception of an activity which involves the agents' actions having a distributive goal will do.

19.   slide

I claim that this inference, your-goal-is-my-goal, could enable pure goal ascribers to avoid the problem of false belief.

Here's how it works.

20.   slide

Return to the earlier situation with the two boxes, but now suppose things are slightly different. You dislike the owl, and you are trying but failing to retrieve the cat. Ayesha comes along and, from the situation or cues she offers, you can infer that she will engage in joint action with you. Since you are committed to getting the cat, you infer that the goal of her action will be to retrieve the cat. So now when she reaches into the North Box, you have a reason to suppose that the goal of her action is to retrieve the cat. In fact you have a conflict between two routes to knowledge of the goal Ayesha's action:

1. your-goal-is-my-goal tells you that the goal of her action is to retrieve the cat; whereas

2. pure goal ascription tells you that the goal of her action is to retrieve the owl.

The view I'm offering says nothing about how you balance the conflicting evidence. In particular, I don't claim that your-goal-is-my-goal will always win. Just the fact that it will *sometimes* win means that, even without ascribing mental states, you can sometimes be right about the goals of actions despite false beliefs.

So capacities for shared agency can boost your ability to ascribe goals.

21.   slide

But this isn't the key point. The key point is that your-goal-is-my-goal enables goal ascribers to detect **incorrect means**.

Pure goal ascription does not allow for the possibility that an action might be directed to a goal where the action is not a rational or efficient means of achieving that goal. As long as we cannot identify cases in which actions are not rational means to achieving goals, we have no need for a concept of belief.

But your-goal-is-my-goal provides a way of separating goals from means, and so creates space for a conception of belief to do some work. As Davidson almost says somewhere,[4] **the conception of belief is built to explain differences between an agents' actual actions and the actions that would achieve their goals**.

22.   slide

So here one way in which I think shared agency might in principle be involved in the step from pure goal ascription to (minimal) theory of mind. **Capacities for shared agency, plus an understanding of distributive goals, allows pure goal ascribers to identify incorrect means, and so gives the access to the phenomena that beliefs are needed to explain.**

23.   slide

My theme is the idea that capacities for shared agency might explain the emergence of sophisticated forms of culture and cognition. So far we looked at the emergence of minimal theory of mind as a warm up. Now to the main business. **Can shared agency play a role in explaining the emergence of referential communication?**

24.   slide

Consider this experiment by Hare and Tomasello Hare & Tomasello (2004, experiment 3) whose two main conditions are depicted here. The pictures in the figure stand for what participants, who were chimpanzees, saw. The question was whether participants would be able to work out which of two

---

[4] 'The concept of belief thus stands ready to take up the slack between objective truth and the held true, and we come to understand it just in this connection.' (Davidson 1984, p. 170)

containers concealed a reward. In the condition depicted in the left panel, participants saw a chimpanzee trying but failing to reach for the correct container. Participants had no problem getting the reward in this case, suggesting that they understood the goal of the failed reach. In the condition depicted in the right panel, a human pointed at the correct container. Participants did not reliably get the reward in this case, suggesting that they failed to understand the goal of the pointing action.[5]

Taking this paradigm as a case study, we want to suggest that your-goal-is-my-goal might enable us to understand how abilities to engage in joint action could be part of what explains the step from pure goal ascription to understanding referential communication.

Let us imagine ourselves as the chimpanzee for a moment. We witness the pointing action. With our eyes we follow the point to a container (see Moll & Tomasello 2007, p. 6). So we do associate the pointing action with its target. But we are no more likely to choose this container than the other in seeking the reward. So we probably do not think of the pointing action as having any goal which would clue us in to the relevance of the container it indicates.

25. slide

How could we fail to understand the pointing action? Pointing, like all communicative action, involves a Gricean circle. That is, communicative actions characteristically have goals which the actions are means to realising only because others recognise them as means to realising those goals.

Moll and Tomasello stress this:

> 'to understand pointing, the subject needs to understand more than the individual goal-directed behaviour. She needs to understand that by pointing towards a location, the other attempts to communicate to her where a desired object is located; that the other tries to inform her about something that is relevant for her' (Moll & Tomasello 2007, p. 6).

Let me put the point another way. If you observe someone using a novel tool, you might easily fail to know which ends their action can serve. In this sense, their action involves opaque means and so pure goal ascription will be impossible. Similarly, if you are not a communicator and you observe someone performing a communicative action, you might easily fail to know

---

[5] The contrast between the two conditions is not due merely to the fact that one involves a human and the other a chimpanzee. Participants were also successful when the failed reach was executed by a human rather than another chimpanzee (Hare & Tomasello 2004, experiment 1).

which ends this action can serve. And so the action will be opaque to you as a means to an end, and so pure goal ascription will be impossible.

I want to suggest that capacities for shared agency can explain how pure goal ascribers are able to break into the Gricean circle and respond appropriately to communicative actions.

26.   slide

Recall the your-goal-is-my-goal inference. How could exploiting this inference could enable someone to respond appropriately to the pointing gesture?

Suppose that, before pointing, the agent had used facial gestures to signal willingness to engage in joint action with us and that we had exploited the your-goal-is-my-goal inference. Then we would believe, perhaps mistakenly, that a goal of the pointing action was to discover the reward. In which case the pointing action would have been no less helpful in enabling us to succeed than the failed reach—which, as you may recall, was very helpful. So the your-goal-is-my-goal inference can enable a goal ascriber to misunderstand pointing actions as something like failed reaches. This means that, even without any deeper understanding of communication, goal ascribers can respond appropriately to helpful pointing actions in the context of joint action.

27.   slide

Our suggestion is that individuals could reliably respond appropriately to some pointing actions in the context of joint action without understanding pointing.

They do this by productively misunderstanding the goal of the action.

The actual goal of the pointing action is, let's say, that I attend to this object because I recognise that you intend, by means of this gesture, to get me to attend to it. But someone relying on your-goal-is-my-goal would identify the goal as discovering the object.

They would, however, be right about the target of the action, which is the right box. And they would identify the action as a failure, prompting further exploration much as in the case of the failed reach.

28.   slide

So far I've been talking about just one special situation. But the idea generalises not just to other situations involving pointing but also to single-word utterances.

The key to seeing this is to note a distinction between the goal of an action and its target object or target objects. You can be right about a target object while being wrong about

For something to be among the *target objects* of an action is for it to be an object that would normally be intervened on if the action were successful. To illustrate, in grasping an object, I would normally intervene on that object if my action were successful. For a trickier example, suppose that I catch a ball, thereby preventing you from catching it. In this case, you and the ball are both among the target objects of my action. This is true whether or not preventing you from catching the ball was among the goals of my action.

Note that identifying the target objects of actions does not always rely only on goal ascription. It is also possible to identify target objects by association and causal reasoning.

Note also that it is possible to be partially right about the target objects of an action while being quite wildly mistaken about the goals to which it is directed.

29.   slide

Now consider how this works for single-word utterances.

The basic requirement is this: in a particular context, the goal ascriber must associate a communicative gesture with its referent. For instance, she must associate the pointing gesture with the object indicated; or, if (say) she is looking to see who has an object she must associate an utterance of 'daddy' with the daddy, or an utterance of 'food' with the food. These associations make it is possible to identify the target objects of one-word utterances independently of knowing the goals of these actions. We know that such associations can be formed independently of understanding utterances in the sense of knowing how to respond to them. And, as we saw, outside the context of joint action, merely associating a gesture with its referent falls short of being able to respond appropriately.

But if an observer can know that the utterer is attempting to engage in joint action with her, then she may infer that the goal of her target's action is her goal and so be motivated to treat the target object of a communicative gesture as relevant to the goal of her own actions. This will reliably (but not always) enable her to respond appropriately to the communicative gesture even without understanding it as a communicative gesture. And once she has experienced how that communicative gesture works as a tool for guiding others' actions in the context of joint action, she may be in a position to realise, further, that the same tool can be used in other contexts.

This, in barest outline, is how possessing abilities to engage in joint action

means that an individual with an ability to ascribe simple goals only and no understanding of communicative intent might nevertheless reliably respond appropriately to some communicative gestures, and so come be in a position to understand how such gestures can be used to guide others' actions.

29.1.   Producing communicative gestures

It is also possible that someone could be motivated to produce communicative actions without understand communicative intention. Suppose that you are picking some strawberries but have missed a particularly nice looking cluster. Suppose also that I cannot actually reach these strawberries, but could unsuccessfully reach for them. If I knowing that you can exploit your-goal-is-my-goal, then I might be motivated to reach unsuccessfully for the strawberries not because I think I can get them but because I anticipate that you will identify obtaining these strawberries as the goal of my action.

30.   slide

I've been arguing that your-goal-is-my goal would enable someone to **fake an understanding of referential communication**. That is, it enables them to respond to communicative actions in much the way that someone who really understood them would.

Before I finish, I want to explain an application to the natural pedagogy hypothesis. I want to suggest that thinking about referential communication as emerging from shared agency complements the natural pedagogy hypothesis.

31.   slide

On the face of it, this seems wrong because George and Gergo explicitly appeal to communicative intention in characterising natural pedagogy. Natural pedagogy does *not* require mindreading:

> 'infants, by decoding ostensive signals, recognize the communicative intentions of communicators ... Attributing a communicative intention is attributing a second-order intention' (Csibra 2010)

George and Gergo also say that natural pedagogy depends on abilities to attribute knowledge states:

> 'the assumption of relevance requires the learner to decode the teacher's manifestation with respect to his own knowledge. ...

the pedagogical question driving the learner's inferential interpretation of the teacher's demonstration is this: "What is the new information in this manifestation that I don't yet know and would not be able to figure out myself?"' (Csibra & Gergely 2005)

But note, however, that they also seem to say that natural pedagogy can occur without mindreading:

> 'the ability to teach and to learn from teaching is a primary, independent, and possibly phylogenetically even earlier adaptation than ... the ability to attribute mental states.' (Gergely & Csibra 2012)

Whatever exactly their view is, I think that they *can* hold on to the idea that natural pedagogy is phylogenetically prior to mental state attribution by leaning on the idea that capacities for shared agency enable agents fake an understanding communicative of intention and relevance (to act almost as if they did understand these things).

Perhaps natural pedagogy got started by virtue of individuals realising that in the context of joint action, actions which fail can usefully draw attention to their target objects.

I'm not suggesting that there aren't gaps. (*The step from reference to particular to reference to kinds is missing.)

But there is a reason why proponents of natural pedagogy should be interested in the idea that referential communication can be built on top of capacities for shared agency. According to Natural Pedagogy, referential communication's primary function is to get around the problem of opaque means in tool use. And, of course, the hypothesis proposes that humans come to understand and learn about one kind of opaque action, tool use, by introducing another type of opaque action, communication. In effect, the hypothesis says that if we can have one special opaque tool, then we can have them all. So for proponents of natural pedagogy, the question of how humans or their ancestors got to grips with that one special tool, referential communication, is quite pressing.

32.   slide

So the challenge was ...

The conjecture I've been pursuing is that capacities for shared agency ...

As I said at the start, the aim today was only to gesture at ways in which shared agency might matter for the emergence of mindreading and referential communication. I offered two separate gestures, but both make use of your-goal-is-my-goal. The central idea is that pure goal ascription gets a massive boost when even minimal capacities for shared agency are added.

The **first step** was from pure goal ascription to minimal theory of mind. Here the idea was that capacities for shared agency allow individuals capable only of pure goal ascription to identify actions as incorrect means to their goals, and so makes them aware of the gap which the notion of belief is made to fill.

The **second step** was from pure goal ascription to referential communication. Here the idea was that capacities for shared agency allow us to productively misunderstand communicative gestures, and perhaps even to produce actions that, like communicative gestures, function to transmit information.

Probably lots of the details are wrong. But the truth is that capacities for shared agency must play some role in explaining the emergence of insight into interpersonal differences in belief and of referential communication. And we need to know how this works.

33. [*bin] slide

Consider a related experiment by Leekam et al. (2010). Again participants had to retrieve a reward from one of several closed containers, but this time they were two- and three-year-old children. In one condition participants were shown an adult holding up a replica of the target container. Leekam and colleagues found that when this action was accompanied by an engaging facial expression, three-year-old children were significantly better at identifying the correct container compared to when the the action was accompanied by a neutral facial expression (p. 116). Why did the engaging facial expression enhance performance? The authors consider the idea that engaging facial gestures somehow help children to understand communicative intentions.[6] An alternative possibility is that children succeeded without understand the replica as a sign at all. Instead they may have associated the replica with the container it resembled (which by itself is not enough to motivate selecting this container, of course), regarded the engaging facial gestures as expressing willingness to engage in joint action, and exploited your-goal-is-my-goal to infer that a goal of the action of holding up the replica was to find the reward. In this way they might have understood (or misunderstood) the action of holding up the replica as like a failed reach in being an attempt to retrieve the reward.

---

[6] Leekam et al. (2010, p. 118): 'the adult's social cues conveyed her communicative intent, which in turn encouraged the child to 'see through the sign' ... helping them to take a dual stance to it.'

References

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168.

Csibra, G. & Gergely, G. (2005). Social learning and social cognition: The case for pedagogy. In M. H. Johnson & Y. Munakata (Eds.), *Processes of Change in Brain and Cognitive Development. Attention and Performance*, volume XXI (pp. xxx–xxx). Oxford: Oxford University Press.

Davidson, D. (1975 [1984]). Thought and talk. In *Inquiries into Truth and Interpretation* (pp. 155–170). Oxford: Oxford University Press.

Gergely, G. & Csibra, G. (2012). Natural pedagogy. In M. Banaji & S. Gelman (Eds.), *Navigating the Social World: What Infants, Chidren, and Other Species Can Teach Us*. Oxford: Oxford University Press.

Hare, B. & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, *68*(3), 571–581.

Leekam, S. R., Solomon, T. L., & Teoh, Y. (2010). Adults' social cues facilitate young children's use of signs and symbols. *Developmental Science*, *13*(1), 108–119.

Moll, H. & Tomasello, M. (2007). Cooperation and human cognition: the vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B*, *362*(1480), 639–648.