

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 commitments 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.**

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, reaching 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 is 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 action¹ 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 alternative).

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?**

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.' (?, p. 170)