Interacting Mindreaders: CEU Talk

Stephen A. Butterfill
<s.butterfill@warwick.ac.uk>

Abstract

Could interacting mindreaders be in a position to know things which they would be unable to know if they were manifestly passive observers? This paper argues that they could. Mindreading is sometimes reciprocal: the mindreader's target reciprocates by taking the mindreader as a target for mindreading. The paper explains how such reciprocity can significantly narrow the range of possible interpretations of behaviour where mindreaders are, or appear to be, in a position to interact. A consequence is that revisions and extensions are needed to standard theories of the evidential basis of mindreading. The view also has consequences for understanding how abilities to interact combined with comparatively simple forms of mindreading may explain the emergence, in evolution or development, of sophisticated forms of social cognition.

1. On the evidential basis of mindreading

Thank you for inviting me. I'm visiting until Christmas to immerse myself in your work here, with the aim continuing one collaboration and perhaps starting new collaborative work.

My work is mostly on mindreading and joint action. Here I'm aiming to bring the two together by discussing mindreading by the agents of a joint action. The question I aim to answer is whether interacting mindreaders could be in a position to know things which they would be unable to know if they were manifestly passive observers.

But before I get to the question, let me step back to explain the wider issues that motivate it.

2. slide: Vygotskian Intelligence Hypothesis

Moll and Tomasello say …

Coming at things from a very different starting point, Sebanz and Knoblich make a different but related claim. They say …

Faced with these bold conjectures, it is natural to ask, How could that work? How could social interaction drive or ground sophisticated forms of cognition such as those involved in communication by language or sophisticated forms of mindreading?

I think we might be able to make a small step towards understanding these conjectures by considering the role of social interaction in simple forms of mindreading.

3. slide: Question

So this talk is based on a narrow question. Contrast a mindreader who is, or appears to be, capable of interacting with her targets and a mindreader who can manifestly only observe. Is it possible that the interacting mindreader is in a position to know things which she would be unable to know if she were unable to interact with her targets?

Most philosophers who have written about the evidential basis of mindreading have assumed that the answer is 'no'. Of course they would allow that interaction can facilitate mindreading in practice. But their theories on the evidential basis of mindreading assume that interacting mindreaders have no advantage *in principle* over mere observers. So on their theories, a purely passive mindreader observing from behind a one-way mirror is on a par with a mindreader who, actually or apparently, could interact with those she seeks to interpret. The two are on a par in this sense: in principle the same evidence could be available to each, and each can exploit the same routes to knowledge in moving from evidence to ascriptions of thought and action.

In this talk I aim to show that mindreaders actually or apparently capable of interacting with their targets are at an advantage not only in practice but also in theory. Their ascriptions could exploit routes to knowledge which would be unavailable if they were entirely passive observers.

4. slide: Reciprocity

Why suppose, in advance of considering the details, that interacting mindreaders might know more? A mindreader's target is often also a mindreader and may sometimes reciprocate by taking the mindreader as a target for min-

dreading. It ought to be possible, in mindreading, to make use of this reciprocity. But how could such reciprocity facilitate mindreading? If we assume the mindreader is merely observing her target, that there is manifestly no potential for interaction, then it seems that any way of exploiting reciprocity would involve higher-order ascriptions. The mindreader would ascribe to her target beliefs (say) about the mindreader's own beliefs and other mental states. And if her target reciprocates, she might escalate by ascribing to the target beliefs about her own beliefs about the target's beliefs about her beliefs. While this might be useful in some situations, the nesting this approach requires quickly becomes dauntingly complex. And the basic intuition about reciprocal mindreading goes unsatisfied. Reciprocal mindreading should sometimes result in something like a meeting of minds rather than an escalation of higher-order ascriptions. Perhaps fully exploiting reciprocity in mindreading requires the mindreaders to be in a position to interact with each other; perhaps in some cases being or appearing poised to interact can somehow enable mindreaders to exploit reciprocity without first having to ascribe higher-order mental states. This is the hunch we develop in what follows.

5. slide: Goal Ascription

In most discussions of mindreading, the focus is on ascription of beliefs and other mental states to individuals. But in what follows I want to focus on goal ascription.

Goal ascription is the process of identifying to which outcomes others' purposive actions are directed. To illustrate, suppose that Hannah kicks a ball thereby both preventing her sisters from scoring and also breaking a window. Asked about the episode, Hannah might protest, truthfully, that the goal of her action was not to break the window but only to reverse the others' advance. As this illustrates, among the actual and possible outcomes of an action, only some are outcomes to which the action is directed. Goal ascription is the process of identifying those outcomes.

Some might claim that goal ascription is not mindreading, perhaps because identifying relations between actions and the outcomes to which they are directed does not necessarily involve ascribing mental states. I'm happy if you want to define mindreading so as to exclude mere goal ascription.

Still, everyone should agree that mindreading very often depends on goal ascriptions. After all, evidence for ascriptions of belief and of other mental states often includes the occurrence (and non-occurrence) of goal-directed actions. Identifying these actions typically requires goal ascription. In some cases, then, evidence for the ascription of a goal will indirectly support the ascription of a belief or other mental state—the evidence will support a goal

ascription which will in turn support the ascription of a belief (say) to the agent of the action. Accordingly, even those who deny that goal ascription is mindreading should agree that sometimes evidence for goal ascription is indirectly evidence for ascription of belief and other mental states.

6. slide: Evidence for goal ascription

What evidence could support ascriptions of goals to actions?

Some philosophers seem to assume that goal ascription depends on knowledge of the agents' intentions which in turn depends on knowledge of their beliefs, desires and other mental states. If this claim were true, it would make no sense to discuss the evidential basis of goal ascription except in discussing the evidential basis of mindreading more generally.

However, the falsity of this claim is presupposed in both developmental and comparative research on goal ascription.[1]

To show that goal ascription is possible without ascribing mental states, we need to say what evidence someone could have that would enable her to know to which goals an action was directed.

According to what Csibra and Gergely call 'the principle of rational action',

> 'an action can be explained by a goal state if, and only if, it is seen as the most justifiable action towards that goal state that is available within the constraints of reality.'[2]

This principle plays two distinct roles. One role is mechanistic: this principle forms part of an account of how infants (and others) actually ascribe goals. Another role is normative: this principle also identifies grounds on which it would be rational to ascribe a goal. I'm concerned only with its normative role.

As Csibra and Gergely formulate it, the principle might seem simple. But actually their eloquence is hiding some complexity. It is perhaps worth spelling out what might be involved in applying this principle. Let me try to spell it out as an inference with premises and a conclusion ...

1. action a is directed to some goal;

---

[1] Compare Gergely et al. (1995), Woodward (1998) and Penn & Povinelli (2007) among many others.

[2] Csibra & Gergely (1998, p. 255); cf. Csibra et al. (2003). A related but different 'principle of efficiency' has been formulated by Southgate et al. (2008, p. 1061): 'goal attribution requires that agents expend the least possible amount of energy within their motor constraints to achieve a certain end.'

2. actions of a's type are normally capable of being means of realising outcomes of G's type in situations with the salient (to any concerned) features of this situation;

3. no alternative type of action is both typically available to agents of this type and also such that actions of this type would be normally be significantly better[3] means of realising outcome G in situations with the salient features of this situation;

4. the occurrence of outcome G is typically desirable for agents of this type;

and

5. there is no other outcome, G', the occurrence of which would be at least comparably desirable for agents of this type and where (2) and (3) both hold of G' and a

may jointly constitute defeasible evidence for the conclusion that:

6. G is a goal to which action a is directed.

We suggest that the above inference, from (1)-(5) to (6), is a route to knowledge of the goals of actions in this sense: in some cases it would be possible to know the premises without already knowing the conclusion; and, in some of those cases, knowing the premises could put one in a position to know the conclusion.[4]

[*skip paragraph*] Why accept this? Suppose that (1)-(5) are facts and that G is not a goal to which action a is directed. Then from (1) we know that a has a goal, call it G'. And from (5) we can infer that either (i) the action, a, was not of a type that is normally capable of being a means to the outcome to which it was directed, or (ii) a significantly better (for instance, a more reliable or less effortful) means of achieving G' is typically available to agents of this type, or (iii) G''s occurrence would be significantly less desirable than G's occurrence for agents of this type. So the agent's action fell short of being appropriate or optimal in some way, or the agent has atypical preferences or capabilities,

[3] An action of type a' is a better means of realising outcome G in a given situation than an action of type a if, for instance, actions of type a' normally involve less effort than actions of type a in situations with the salient features of this situation and everything else is equal; or if, for example, actions of type a' are normally more likely to realise outcome G than actions of type a in situations with the salient features of this situation and everything else is equal.

[4] Knowledge of the conclusion may not require explicit knowledge of what is salient to others, what is desirable to them, or what makes actions better for them. It is arguably sufficient that the individual ascribing a goal is entitled to rely on being sufficiently similar to the target of her ascription with respect to these things.

or the situation is not normal. While any of these is possible—and perhaps in some cases even to be expected—there are also circumstances in which it is reasonable to suppose that none obtains, that the situation will be normal in relevant respects, that agents will have typical preferences and capabilities and that they will act in ways that are appropriate and (in the limited sense in play here) optimal. In such circumstances, the above inference can serve as a route to knowledge of the goals of actions. The existence of this route to knowledge shows that goal ascription doesn't invariably depend on mental state ascription.

Of course the inference is not deductive. It is possible for the premises to be true and the conclusion false. But this is typically of any inference describing a route to knowledge.

I'm not suggesting, of course, that this inference describes how anyone actually ascribes a goal. My concern is with a related, but more narrowly theoretical question: what evidence could support goal ascription.

I also want to concede that this can't be the whole story about goal ascription because this inference has only limited applications. For example, it cannot be used when an action is the best available means of achieving two or more comparably desirable outcomes.

Anyway, here is one route to knowledge of the goals of actions. My aim in what follows is to show that there is another route to knowledge of the goals of actions that is available only to interacting mindreaders and not to those who merely observe. But first let me explain why we might want another route to knowledge of the goals of actions.

To apply Csibra and Gergely's principle in a particular case where you first observe an action and then conclude that its goal is G, you need to know that the action is a 'justifiable action towards' G. In other words, you need to know that actions of this type are a means to achieving G.

Whatever you think of the details of this principle of rational action, it seems plausible that the evidence for goal ascription sometimes includes considerations about which ends actions are means to.

7. The problem of opaque means

This the problem of opaque means: When an observer faces an action but cannot identify ends to which it could be a means, she might be unable to recognise the action's goal.

Ignorance about to which ends actions are means can be an obstacle to goal ascription.

Of course the opaque means are not always an insurmountable obstacle to goal ascription, and they may only rarely be a problem for human adults who can rely on culture, convention, or communication. Where goal as-

cribers lack such sophistication, the problem of opaque means is likely to be more pressing.

Some of the most plausibly unique aspects of human cognition depend on our abilities to recognise the goals of novel behaviours involving tools, and of communicative gestures. The problem of opaque means is likely to arise in both cases if goal ascription is based entirely on observation (so that the possibility of interaction is ignored) and if the goal ascriber lacks sophisticated social skills.

To illustrate, contrast two cases of tool use. In one case, someone uses a reamer to juice a lime; in the other, someone else scores shag with a lame to prevent a loaf from cracking. Without communication, repetition or convention, an observer familiar with reamers but not lames may be able to identify the goal of the first action only.

The problem of opaque means is also likely to arise where actions involve multiple steps that do not form a familiar sequence, can occur in various orders and can be interspersed among other activities; as in preparing spirit from grain, for example.

The problem of opaque means also affects communicative actions because these characteristically have goals which the actions are means to realising only because others recognise them as means to realising those goals (a Gricean circle). To illustrate, consider an experiment from Hare & Tomasello (2004, experiment 3) whose two main conditions are depicted in figure 1 on the following page. 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 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] This may be because of the problem of opaque means. One theoretically possible explanation of these findings is that the participants could identify to which end a failed reach might be a means, but not to which end a communicative gesture might be a means.[6] Whatever the truth about the

[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).

[6] Hare and Tomasello consider several explanations for their findings including 'the hypothesis that chimpanzees do not understand the communicative intent of a cooperative-communicative experimenter' (2004, p. 580). Moll & Tomasello (2007, pp. 5–7) argue for a hypothesis along these lines by appeal to a range of related findings.

Figure 1: A failed reach (left) and a helpful point (right). Reproduced from Hare & Tomasello (2004, p. 557, figure 4).

chimpanzees' performance, this possibility illustrates how the problem of opaque means can be an obstacle to exploiting communicative gestures.

This, then, is the problem of opaque means: failures to identify to which ends actions are means can impair goal ascription. The problem is potentially a problem for mindreaders given the standard, purely observational theories of mindreading.

In what follows we shall explain how being able to interact with another sometimes makes available a route to knowledge of the goals of her actions which avoids the problem of opaque means. This will support our claim that an interacting mindreader might be in a position to know things which she would be unable to know if she were only observing.

8. Interactions involving distributive goals

As a preliminary to defending this claim, we need to specify which types of interaction are relevant.

Let us stipulate that an outcome is a *distributive goal* of two or more agents' actions just if two conditions are met. First, this outcome is one to which each agent's actions are individually directed. Second, each agent's actions are related to the outcome in such a way that it is possible for all the agents (not just any agent, all of them together) to succeed in bringing about this outcome.

To illustrate, suppose that, while doing some gardening, we find a turnip too big for any of us to easily pull out of the ground alone. We each individually intend that we all pull up the enormous turnip and we act on this intention.[7] There is a single outcome, freeing the turnip, to which each of

[7] For an argument that it is possible for each of us individually to intend that we do something, see Bratman (1997) and Bratman (2012).

our actions is individually directed. And it is possible for all of us to succeed in bringing about this outcome. So this outcome is a distributive goal of our actions.

In defining 'distributive goal' we stipulated that it must be possible for all of the agents of actions with a distributive goal to succeed relative to that outcome. Consider a variation on our first turnip-pulling story. Instead of each intending that we pull up the turnip, we each simply intend to pull up the turnip. In this case, some might claim that this amounts to us each intending that he or she pulls up the turnip. And this claim might be taken to support the further claim that there is no outcome to which each of our actions are directed in virtue of our having these intentions such that we could all succeed relative to this outcome. We take no stand on whether either claim is correct. For our purposes all that matters is that there are distributive goals. And the existence of distributive goals follows from the possibility of each of us intending that we pull up the turnip (rather than simply intending to pull it up), and from the possibility of each of us intending to contribute to pulling up the turnip or pull it up.

For two agents' actions to have a distributive goal it is sufficient that their actions constitute a joint action (at least this is true on almost any account of joint action).[8] However, the converse does not hold. As the first variation of the turnip-pulling story indicates, two or more agents' actions may have a distributive goal even though the agents do not know about each others' intentions or actions. In fact our actions might have a distributive goal even though none of us is aware of this, or even of the others' existence. (We resist the temptation to contrive a truly gigantic turnip and a pitch dark, very stormy night; readers can probably guess how this would go.) One consequence of this is that two or more agents' actions may have a distributive goal even though they are not engaged in joint action (at least not on any standard account of joint action).[9]

---

[8] We know of two definitions of joint action on which it is not straightforward that all joint actions involve distributive goals. One is provided is provided by Ludwig (2007, p. 366) and quoted in footnote 9; the other is due to Sebanz et al. (2006, p. 70). According to Sebanz and colleagues' 'working definition', 'joint action can be regarded as any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment' (2006, p. 70). On one way of reading this, the latter part is equivalent to 'a change in the environment is such that: to bring it about two or more individuals coordinate their actions'; this does suggest that the actions have a distributive goal. On other standard accounts it is straightforward that joint action involves distributive goals. Compare Bratman (1992, pp. 329-31), Searle (1990, pp. 96-7) and Gilbert (2009, pp. 168-9).

[9] On many accounts of joint action, each agent involved in a joint action must believe, expect or know something about the jointness of her action. See Bratman (1993, p. 103), Butterfill (2012, p. 40), Kutz (2000, p. 10), Miller (2001, p. 56) and Roth (2004, p. 361). Exceptions include Pacherie (2011) and Ludwig (2007). According to Ludwig, 'The concept

The notion of a distributive goal is a narrowly technical one. In introducing this notion by stipulation we are not aiming to match anyone's intuitions about anything. What matters for now is just that the definition is theoretically coherent and that it is possible for actions to have distributive goals. Our aim in introducing distributive goals is to capture a minimal requirement on the sort of interactions relevant to characterising a route to knowledge of the goals of actions available only to interacting mindreaders.

9. Your-goal-is-my-goal

If a mindreader is able to interact with her targets, if she is not limited to merely observing them, how might this enable know things about the goals of their actions?

Here is an intuitive idea that doesn't quite work: if a mindreader is engaged in an interaction with her target that involves a distributive goal, it may be easy for the mindreader to know what the goal of her target's actions is because this goal is the goal of her own actions. So if she knows the goal of her own actions and she knows that she is engaged with her target in an interaction involving a distributive goal, then she already knows what the goal of her target's actions are.

Of course this intuitive idea is no use it stands. For the inference it captures relies on the premise that the mindreader and her target are engaged in actions with a distributive goal. But for the mindreader to know this premise it seems she must already know which goal her target's actions are directed to.

Fortunately there is a way around this. For there are various cues which signal that one agent is prepared to engage in some joint action or other with another, and joint actions involve distributive goals. 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.

---

of a joint action as such is just that of an event of which there are multiple agents' (2007, p. 366). Depending on what events are and what it is to be the agent of an event, it may turn out that, on Ludwig's definition, our actions' having a distributive goal is sufficient for us to be engaged in joint action.

Our 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[10] 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. Second, in some of those cases knowing the premises would put one in a position to know the conclusion.

10. Avoiding the problem of opaque means

The problem of opaque means was this: failures to identify to which ends actions are means can impair goal ascription (see section 7 on page 6). Showing how your-goal-is-my-goal makes it possible to avoid this problem is a way of demonstrating the potential value of interaction for mindreading.

In our earlier example a novice parent is struggling to lift his heavy twin pram onto a bus when a stranger joins in and they lift the pram together (see page 10). Suppose the stranger starts tipping the pram in a way that the novice parent fails to recognise as a means, indeed the only means, of getting it onto the bus. Outside the context of joint action, failure to realise that the action is a means to getting the pram onto the bus might have the consequence that the parent's evidence on balance supports the conclusion that the goal of the stranger's actions is to take the pram off the bus. (Perhaps the stranger is impatient to get onto the bus herself.) But in the context of joint action, your-goal-is-my goal gives the parent additional evidence for supposing that, even though the stranger's actions do not seem to him to be

[10] 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 (see further footnote 8 on page 9). Attempts to characterise joint action in ways relevant to explaining development include Tollefsen (2005), Carpenter (2009), Pacherie (2011) and Butterfill (2012). The last of these not only shows how a primitive form of joint action can be characterised without ascriptions of higher-order mental states but also shares our focus on relations between actions and goals rather than between agents and intentions.

a means to getting the pram onto the bus, this really is the goal of her actions. Of course this additional evidence will not always outweigh other evidence. But it will do sometimes, and this is all we need. We have identified evidence for goal ascription that is available independently of a goal ascriber's knowing which ends actions are means to, and we have illustrated how this will sometimes enable interacting mindreaders to avoid the problem of opaque means

We saw earlier that the problem of opaque means may impair goal ascription where actions involve novel uses for tools. How could your-goal-is-my-goal mitigate the problem in such cases? Imagine we are interacting with a young child, Ayesha, and want her to understand how a new tool is used. It is difficult to convey this to her directly. So we first get her interested in achieving an outcome that would require the new tool, knowing that she will perform actions directed to achieving this outcome. We then signal to Ayesha that we will act jointly with her. Now she is in a position to know what the goal of our action will be when we deploy the tool. She is able to identify this goal despite being unable to recognize it as an end to which our tool-using action is a means. She is able to identify this goal because she knows that this is her goal and that we were attempting to engage in joint action with her. This is one illustration of how interacting mindreaders have at their disposal ways of identifying the goals of actions involving novel uses of tools which are unavailable to mindreaders who can only observe.

As this example indicates, exploiting your-goal-is-my-goal can shift the burden of identifying goals from a mindreader to her target. In the example Ayesha is the focal mindreader and we are her target; but her success in identifying the goal of our actions depends on this, that our willingness to act jointly with her is based on *our* knowledge of the goals of *her* actions. In purely observational mindreading, the target's beliefs about the goals of the mindreader's actions are not normally relevant (except, of course, when the mindreader is ascribing such beliefs). But interacting mindreaders who rely on your-goal-is-my-goal thereby rely on their targets having correctly identified the goals of their actions. Of course this is sometimes a reason not to rely on your-goal-is-my-goal. But where the target understands relevant means-ends relations, such as actions involving novel tools, the your-goal-is-my-goal route to knowledge of others' goals may sometimes be the only option.

Here we have a first illustration of how interacting mindreaders can exploit reciprocity without relying on ascriptions of higher-order mental states. The mindreader's knowledge of the goal of her target's actions depends on the target's knowledge of the goal of the mindreader's actions. What makes this dependence possible is not each knowing that the other knows something about the goal of her actions; it is the actual or apparent possibility of interaction.

The distant promise of all this is that understanding how interaction widens the evidential basis for goal ascription and mindreading more generally may eventually enable us to explain the origins, in evolution or development (and ideally both), of abilities to learn novel and opaque uses for tools from others without assuming that rich communicative skills or sophisticated forms of mindreading must already be present.

11. Communicative gestures

When introducing the problem of opaque means (in section 7 on page 6) we saw that it could affect communicative actions. To illustrate this suggestion we drew on an experiment by Hare & Tomasello (2004) in which chimpanzees had to find a reward and were helped by being shown either a failed reach or a helpful point to the target location. Strikingly, for chimpanzees the helpful point is no help at all—even though it superficially resembles the failed reach, which did help. 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 enables mindreaders to make the transition from a simple understanding of goals to an early understanding communicative actions.[11]

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. (In principle we might perfectly understand the pointing action while failing to react to it in any systematic way because we are uncertain about the agent's integrity; but let us discount this possibility for the sake of illustration.) Now 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 retrieve the food. 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,

[11] Note that although our discussion borrows an experimental paradigm, our aim is not to argue for empirical hypotheses about chimpanzee social cognition. Our aim is only to argue for the theoretical significance of interaction for mindreading by showing that your-goal-is-my-goal could in principle enable individuals to make the transition from a simple understanding of goals to an early understanding of communicative actions. Of course it would powerfully demonstrate the relevance of our theoretical argument if we could provide evidence to show that this actually happens. But for now we are concerned with more narrowly conceptual issues.

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.

It is natural to suppose that the difficulty chimpanzees have in Hare and Tomasello's experiments with responding appropriately to helpful pointing but not to failed reaching is due to a failure to understand communicative intention.[12] What we are suggesting is that participants must also have been unable or unwilling to exploit the your-goal-is-my-goal inference.

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.[13] 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.

So far we have illustrated how your-goal-is-my-goal enables responding appropriately to communicative gestures with two examples, pointing and holding up a replica. The pattern of reasoning generalises to a wider range of communicative gestures including single-word utterances. The basic re-

[12] See footnote 6 on page 7. Relatedly, in their discussion of these findings Moll and Tomasello suggest that '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). Assuming this is right, our suggestion is that individuals could reliably respond appropriately to pointing actions in the context of joint action without understanding pointing.

[13] 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.'

quirement 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.[14] 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 a mindreader supposes that her target 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 thing associated with 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.[15] Of course we have only argued that this transition is theoretically possible; we have not attempted to defend any hypothesis about anything's actual evolution or development. But widening the evidential basis for goal ascription to include facts about interactions between mindreaders and their targets does reveal the coherence of novel hypotheses about a role for interaction in the emergence of communicative gestures.

---

[14] Such associations appear to be in place by six months of age or earlier (Tincoff & Jusczyk 1999, 2011). There is debate on when infants are first able to rapidly associate words with novel objects (Werker et al. 1998; Friedrich & Friederici 2011). Interestingly, a possible developmental decalage separating two measures of whether infants have formed a word-object association, preferential looking and comprehension as measured by reaching in response to a request (Gurteen et al. 2011), may suggest that infants can form a word-object association before they can comprehend communicative actions involving the word.

[15] Contrast Csibra's claim that, early in human development, goal ascription ('teleological understanding' in his terms) and identifying the referents of communicative gestures ('referential understanding') 'rely on different kinds of action understanding' and are initially two distinct 'action interpretation systems' (Csibra 2003, p. 456). We have not shown that this claim wrong. But we have shown that there is another possibility: referential understanding may emerge from the teleological understanding together with abilities to engage in simple forms of joint action.

12. The problem of false belief

So far we have been arguing that interaction can matter for goal ascription because it makes available routes to knowledge of the goals of others' actions which avoid the problem of opaque means. There is another reason for supposing that interaction matters for goal ascription and (at least indirectly) for mindreading more generally. To introduce this reason we must first describe another problem affecting goal ascription, the problem of false belief.

Recall that the problem of opaque means occurs when mindreaders must rely entirely on observation and cannot identify to which ends actions are means (see section 7 on page 6). A yet more familiar problem affecting goal ascription arises from the interdependence of beliefs and goals. To illustrate, imagine sitting at a table. On the table are two closed opaque boxes. One box contains an owl, the other a cat. If the goal of your action is to retrieve the cat, and you believe that the cat is in the north box, then (unless things are going very badly) you will reach for the north box. But of course if you had believed instead that the cat was in the south box, then, in acting with the same goal, you would have reached for the south box. Now consider Ayesha who is observing your actions. Suppose Ayesha has sufficient reason to believe, falsely, that you know the cat is in the south box. Then she may be justified in supposing, incorrectly, that the goal of your action, in reaching for the north box, is to retrieve the owl. As this illustrates, differences in belief between observers and protagonists can impair goal ascription when the goal ascriber is unaware of those differences. Call this the problem of false belief.

The your-goal-is-my-goal route to knowledge sometimes enables mindreaders to avoid the problem of false belief. To illustrate consider a counterfactual alternative to the above example. Ayesha dislikes the owl and is concerned with retrieving the cat. As you reach for the north box, your facial gestures signal willingness to engage in joint action with Ayesha. She then concludes, correctly, that your goal is her goal, to retrieve the cat. So despite the difference in belief, the possibility of interaction may mean that Ayesha can knowledgeably identify the goal of your action.

Exploiting your-goal-is-my-goal does not make it possible to avoid the problem of false belief entirely, it only shifts the problem. To illustrate, in the above example Ayesha's ability to correctly identify the goal of your action depends on your correctly anticipating the goals of her actions—on your knowing that she is concerned with retrieving the cat. If you incorrectly anticipated that her actions would be directed to retrieving the owl, Ayesha would have been mistaken in taking your actions to be directed to retrieving the cat. As this indicates, exploiting your-goal-is-my-goal can shift both *which* differences in belief have the potential to impair goal ascription and

also *who* needs to be aware of those differences. This is a second illustration of how the possibility of interaction may enable even individuals with only limited insights into others' minds and actions to exploit reciprocity in mindreading—to exploit the fact that their targets are sometimes themselves mindreading the mindreaders.

In short, then, abilities to engage in joint action make available a route to knowledge of the goals of other agents' actions which does not depend on the goal ascriber knowing what her target believes even when their beliefs relevantly differ.[16] This is not because abilities to engage in joint action provide a way to avoid the problem of false beliefs altogether. Rather they shift the burden of resolving the problem of false belief from a goal ascriber to her target. This is potentially valuable for those who have limited insight into their targets' beliefs, or who lack abilities to track differences in belief altogether.

13. Conclusion

Our aim was to show that interaction can facilitate mindreading in this sense: some routes to knowledge are closed to mindreaders who rely exclusively on observation but open to interacting mindreaders. In pursuing this aim we focused on interactions which are joint actions. Our suggestion was this. Where a mindreader recognizes that her target is about to attempt to engage in joint action with her and where the mindreader is unwilling to change which goals her own actions will be directed to, she may be in a position to know that the goals of her target's actions will be the goals of her own actions. This is the your-goal-is-my-goal route to knowledge (see section 9 on page 10).

To show that this route to knowledge is potentially valuable, we argued that it enables mindreaders to overcome two problems, the problem of opaque means and the problem of false belief. These problems may rarely arise for human adults thanks to sophisticated linguistic communication and extensive knowledge of how things can be achieved. But interest in the evolution or development of mindreading motivates focusing on those with absent or fledgling communicative skills, limited insight into others' minds and narrow knowledge of how things work. For such individuals the problems of opaque means and false beliefs may easily arise, particularly in cases involving novel tools and communicative gestures. Both problems can be overcome by exploiting the your-goal-is-my-goal route to knowledge. It follows

[16] The her-goal-is-his-goal inference (see section ?? on page ??) shows that mindreaders who only observe may also be able to exploit a route to knowledge of the goal of another agent's actions which does not depend on knowing what that agent believes but rather on knowing what a further agent interacting with the first believes.

that expertise with tools and communicative gestures does not presuppose knowledge of others' intentions and beliefs. This shows that it is coherent to suppose that expertise with tools or communicative gestures (or both), far from presupposing sophisticated forms of mindreading, might instead play a role in explaining their emergence in evolution or development.

How might this work? Here is a wild conjecture about a path from simple forms of interaction to sophisticated forms of mindreading. We have seen that abilities to ascribe goals to actions together with abilities to engage in simple forms of joint action involving distributive goals may be sufficient to explain how individuals come to productively misunderstand communicative gestures in the context of joint action (see section 11 on page 13). This might be a step towards responding appropriately to, and using, communicative gestures in other contexts, which might then play a role in explaining how early forms of communication emerge. But we know that early, non-linguistic forms of communication are essential for the later appearance of linguistic communication, and abilities to communicate by language in turn appear to play an important role in the development (at least) of sophisticated forms of mindreading.[17] Now clearly this wild conjecture is not one that we have provided any reason for accepting. Our point in mentioning it is merely to illustrate how **better understanding the roles of interaction in goal ascription and mindreading more generally might eventually lead to discoveries on how sophisticated forms of cognition emerge from conceptually and cognitively undemanding forms of interaction. Take cultures of tool use, the flexible use of communicative gestures, or insight into interpersonal differences in belief and other mental states: if any or all of these originate in abilities to interact, then we need to know how interaction could lead to their emergence. And one small step towards acquiring this knowledge might just be understanding how interacting mindreaders could come to know things which they would not have been able to know if they were only observing.**

References

Astington, J. & Baird, J. A. (Eds.). (2005). *Why Language Matters for Theory of Mind*. Oxford: Oxford University Press.

Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review*, *101*(2), 327–341.

Bratman, M. (1993). Shared intention. *Ethics*, *104*, 97–113.

[17] A range of evidence is reviewed in contributions to Astington & Baird (2005).

Bratman, M. (1997). I intend that we J. In R. Tuomela & G. Holmstrom-Hintikka (Eds.), *Contemporary Action Theory, Volume 2: Social Action*. Dordrecht: Kluwer. Reprinted in Bratman, M. (1999) *Faces of Intention*. Cambridge: Cambridge University Press (pp. 142-161).

Bratman, M. E. (forthcoming 2012). *A Theory of Shared Agency*.

Butterfill, S. (2012). Joint action and development. *Philosophical Quarterly*, *62*(246), 23–47.

Carpenter, M. (2009). Just how joint is joint action in infancy? *Topics in Cognitive Science*, *1*(2), 380–392.

Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions: Biological Sciences*, *358*(1431), 447–458.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*(1), 111–133.

Csibra, G. & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, *1*(2), 255–259.

Friedrich, M. & Friederici, A. D. (2011). Word learning in 6-Month-Olds: fast Encoding–Weak retention. *Journal of Cognitive Neuroscience*, *23*(11), 3228–3240.

Gergely, G., Nadasky, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies*, *144*(1), 167–187.

Gurteen, P. M., Horne, P. J., & Erjavec, M. (2011). Rapid word learning in 13- and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology*, *109*(2), 201–217.

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

Knoblich, G., Butterfill, S., & Sebanz, N. (2010). Psychological research on joint action: Theory and data. In B. Ross (Ed.), *Psychology of Learning and Motivation*, volume 51. Academic Press.

Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research*, *61*(1), 1–31.

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.

Ludwig, K. (2007). Collective intentional behavior from the standpoint of semantics. *Nous*, *41*(3), 355–393.

Miller, S. (2001). *Social action: a teleological account*. Cambridge University Press.

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

Pacherie, E. (2011). Framing joint action. *Review of Philosophy and Psychology*, *2*(2), 173–192.

Penn, D. C. & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society B*, *362*(1480), 731–744.

Roth, A. S. (2004). Shared agency and contralateral commitments. *The Philosophical Review*, *113*(3), 359–410.

Searle, J. R. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in Communication* (pp. 90–105). Cambridge: Cambridge University Press. Reprinted in Searle, J. R. (2002) *Consciousness and Language*. Cambridge: Cambridge University Press (pp. 90–105).

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and mind moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.

Southgate, V., Johnson, M. H., & Csibra, G. (2008). Infants attribute goals even to biomechanically impossible actions. *Cognition*, *107*(3), 1059–1069.

Tincoff, R. & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-Month-Olds. *Psychological Science*, *10*(2), 172–175.

Tincoff, R. & Jusczyk, P. W. (2011). Six-Month-Olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432–444.

Tollefsen, D. (2005). Let's pretend: Children and joint action. *Philosophy of the Social Sciences*, *35*(75), 74–97.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Stager, C., & Casasola, M. (1998). Acquisition of Word-Object associations by 14-Month-Old infants. *Developmental Psychology*, *34*(6), 1289–1309.

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