Interacting Mindreaders

Stephen A. Butterfill

<s.butterfill@warwick.ac.uk>

December 16, 2011

Abstract

***

1. Mindreading: the normative project

The question is what constitutes evidence.

Mostly people have supposed that the evidence is available to pure observers — being able to interact with the targets of mindreading confers no in principle advantage.

In challenging this thesis we want to start with goal ascription, which is either a precursor to or a component of mindreading. Recent interest in non-propositional precursors to mindreading makes this appropriate.

2. Goal ascription

Purposive action is action directed to the realisation of one or more outcomes. Goal ascription is the process of identifying which outcomes others' purposive actions are directed to. Because goal ascription has received little attention, we first review some of the benefits that being able to identify others' goals brings.

It is a familiar idea that goal ascription enables one to learn from others' successes. For example, if you know or can guess that another agent's actions are directed to opening a nut, you may then be in a position to infer that the unfamiliar pattern of actions she is performing constitute a means to open nuts.[1] Or (in a different case) knowing the goal of another agent's actions may enable you to discover a new use for a tool.[2] A slightly less familiar idea is that goal ascription enables one to learn from others' failures as well as their successes. For example, suppose that while you are searching for some peanuts another agent attempts but fails to reach for a closed container. In some circumstances, if you know that the goal of the agent's action was to obtain the peanuts then you now have evidence as to where they might be.[3] This is one illustration of how goal ascription could in principle enable us to learn from others' failures.

Goal ascription enables one to predict and manipulate others' actions. ***if goal involved manipulating an object, you can predict where agent will go at some point; you can also manipulate the agents' action by revealling the location of the object. Eg. Liszowski pointing

Goal ascription is also instrumental for mindreading: knowing which outcomes an action is directed to may constrain hypotheses about what an agent intends as well as potentially providing information concerning what the agent knows, believes or desires. For example, if we know that the goal of an agent's action is to retrieve some peanuts, and if we also know where all the peanuts are, we may be able to infer that she does not know where the peanuts are, or that she falsely believes that some of the peanuts are over there.[4] (Of course this can also work the other way: information about an agent's beliefs or other men-

---

[1] *** (Horner & Whiten 2005)

[2] There is currently much

[3] Hare & Tomasello (2004) exploit this principle in testing chimpanzees' abilities to ascribe goals.

[4] Wimmer & Mayringer (1998) exploit this possibility in testing children's abilities to ascribe false beliefs.

tal states may support conclusions about her goals. Belief- and goal-ascriptions are mutually constraining.)

Despite the close connection between goal ascription and mindreading, goal ascription does not necessarily involve representing representations. To see why not we first need to be careful about the term 'goal'. *** [do the detour here]

The fact that goal ascription does not involve metarepresentation raises the possibility that goal ascription is possible independently of mindreading. Consider an agent who has no communicative skills and no metarepresentational abilities. Nothing prevents her from *representing* goals, of course. But could she actually ascribe goals? Is there any evidence which could be available to her and which would support goal ascriptions? Some philosophers have argued[5] or implied[6] that there could not be. On their view, the goal ascription and mindreading are interdependent in this sense: there is no evidence for hypotheses narrowly about goals, only evidence for more complex hypotheses concerning both goals and mental states (such as beliefs). So, on this view, those without metarepresentational abilities are not in a position to know the goals of other agents' actions.

3.   The problem of opaque means

While we lack a detailed theory of the evidential basis of goal ascription, it is certain that the evidence for goal ascription sometimes includes considerations about which ends actions are means to. Suppose an observer faces an action but cannot identify ends to which it could be a means. This may prevent her from recognizing the action's goal[7] by depriving her of evidence. To illustrate, contrast two cases of tool use. In one case, someone

---

[5]   *Bennett

[6]   *Davidson

[7]   It is possible that some actions have more than one goal. To reduce parenthetical qualifications we shall write as if actions had only one goal. All of our key claims and arguments are consistent with the possibility of actions with more than one goal.

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. As this illustrates, ignorance about to which ends actions are means can be an obstacle to goal ascription.[8] Call this the problem of opaque means.

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. If goal ascription is based on entirely observation (so that the possibility of interaction is ignored), the problem of opaque means is likely to arise in both cases. We have just seen an illustration of how the problem of opaque means arises where tools are used to unfamiliar ends. Relatedly, it 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

---

[8] This is not to say that no observer could ever identify the goal of any action she fails to recognise as a means to achieving that goal. Surely opaque means are not in every case an insurmountable obstacle to goal ascription. Our point is just that opaque means *sometimes* deprive observers of evidence and so prevent goal ascription.
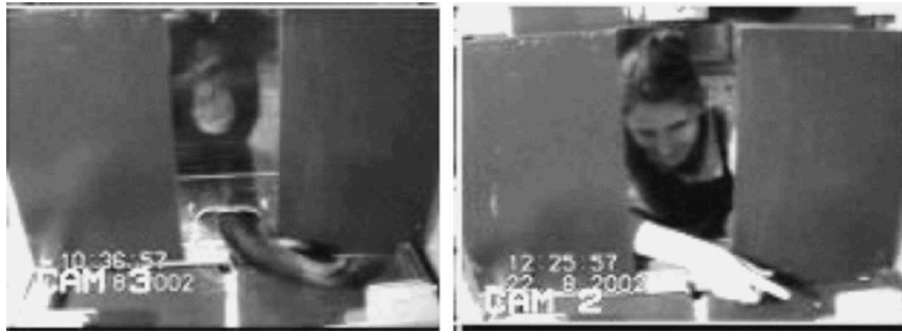
Figure 1: A failed reach (left) and a helpful point (right). Reproduced from Hare & Tomasello (2004, p. 557, figure 4).

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.[9] 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.[10] Whatever the truth about the 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 interpreters given the standard, purely observational models of interpretation. It is not our intention to suggest that the problem of opaque means is a problem *for models of interpretation*. Rather, it is a potential problem *for interpreters*. The existence of this problem shows, we shall argue, that models of interpretation based on pure ob-

---

[9] 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).

[10] 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.

servation are less powerful than models taking into account the possibility of interaction. They are less powerful in this sense: some routes to knowledge of the goals of actions are available only where a model of interpretation takes the possibility of interaction into account. In what follows we shall suggest that abilities to engage in joint action with others provides a route to knowledge of the goals of other agents' actions which avoids the problem of opaque means. This is one way in which moving away from a purely observational model of interpretation yields a richer evidential base for ascriptions.

4.   Joint action

Our overall aim is to defend this claim: there are routes to knowledge of the goals of others' actions which are closed to interpreters who merely observe their targets but open to interpreters capable of interacting with their targets. As a preliminary to defending this claim, we need to specify the types of interaction which are relevant. We shall focus on joint action.

   ***Much disagreement about what it is. Necessary condition is distributive goal. If people want to disagree, this is only terminological.

5.   Your-goal-is-my-goal: a route to knowledge

If an interpreter is able to interact with her targets, if she is not limited to merely observing them, how might this provide her with a route to knowledge of the goals of their actions? The intuitive idea we started with hinges on joint action, a particular form of interaction in which the actions of two or more agents are directed to a single goal (see section 4). Our intuitive idea was this: if an interpreter is engaged in joint action with her target, it's easy for the interpreter to know what the goal of her target's actions is because this goal is the goal of her own

6

actions. So if she knows the goal of her own actions and she knows that she is engaged in joint action with her target, then she already knows what the goal of her target's actions are.

Of course this intuitive idea won't work as it stands. For the inference it captures relies on the premise that the interpreter and her target are engaged in joint action. But for the interpreter to know this premise—for her to know that they are engaged in joint action—it seems she must already know which goal her target's actions are directed to. Worse, on many accounts of joint action the mere truth of the premise would require the interpreter to know the goal of her target's actions.[11] Apparently, then, engaging in joint action presupposes, and therefore cannot be a source of, knowledge of others' goals.

Fortunately there is a way around this. For there are various cues which signal that one agent is prepared to engage in joint action with another. Seeing a stranger with the haggard look of a new parent struggling to get his twin pram on to the bus, you grab the front wheels and make eye contact, raising your eyebrows and smiling. (The noise of the street rules out talking, and you can barely speak the local language anyway.) In this way you signal that you are about to act jointly with the stranger, to lift the pram with him. Since the stranger is fully committed to getting his pram onto the bus, he knows what the goals of his own actions will be. This enables him to infer the goal of your imminent actions: your goal is his goal, to get the pram onto the bus.

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

1. You are willing to engage in some joint action[12] or other with me

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

---

[11] *ref

[12] *What notion of joint action is needed here? Any will do as long as it involves distributive goals.

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 the *your-goal-is-my-goal inference.* To say that this inference 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. We shall consider these points in turn

Is it ever possible to know the premises without first knowing the conclusion? Consider the first premise. Sometimes in the right contexts an individual can recognize in another's facial expressions, engaging gestures or synchronized bodily movements that she are about to engage in joint action with him. Exploiting these indicators does not typically depend on knowing the particular contents of any of their beliefs, desires or goals. Expressions, gestures and movements can naturally indicate imminent jointness in much the way thay can also naturally indicate emotions.[13] Of course these indicators provide no guarantee that others are genuinely willing to engage in joint action. But they are sufficiently reliable to ground knowledge in some cases. The existence of such indicators shows that knowing the first premise of the above inference does not require already knowing which particular goals the other has.

Not everything needs to rest on indicators, however. It is sometimes possible to know that others are about to willing in joint action with you even without relying on such indicators. Thanks to widespread dispositions to act jointly, in some situations it is reasonable to take for granted that others are willing to act jointly. For example, this is often so for children surrounded by family or familiar adults who are struggling with a coat. And in at least some subcultures people using public transport can reasonably take for granted that, within limits,

---

[13] Ideas along these lines are suggested by the discussion of *emergent coordination* in Knoblich, Butterfill & Sebanz (2010).

those around them will act jointly with them when the need is clear. Of course dispositions to engage in joint action may vary between cultural groups and situations. This may be fatal for the peripatetic cosmopolitan, but for others what matters is not whether the dispositions are universal but only that they are sufficiently widespread to be predictable.

Turning to the second point, does knowing the premises of the your-goal-is-my-goal inference sometimes put one in a position to know the conclusion? Of course the inference is not deductive and will only work when certain background conditions are met. These background conditions include the other having largely true beliefs concerning which goal your actions are or will be directed to. After all, the other agent may be willing to act jointly with you while being entirely mistaken about the goals to which your actions will be directed. Where this happens, the premises of the inference might be true but the conclusion false. (We return to this point in section 8 on page 14.) But there are situations in which it is reasonable to ignore this possibility, as for instance when stereotypes, conventions or simplicity should and do make the goal of your actions obvious to the other agent. Given the right conditions, that others are willing to engage in joint action with you is sometimes sufficient reason to hold that you will end up acting jointly with them even where the goal to which your actions will be directed is already fixed.

In short, then, the two requirements for the your-goal-is-my-goal inference to characterise a route to knowledge are met. In some cases it is possible to know the premises without already knowing the conclusion thanks to natural expressions of willingness to engage in joint action. And knowing the premises sometimes puts one in a position to know the conclusion thanks to the fact that, when things are going well, another's willingness to engage in joint action with you can be based on an accurate assessment of the goals of your actions.

The your-goal-is-my-goal route to knowledge is characterised by an inference. However, exploiting this route to knowledge

may not require actually making the inference or knowing the premises. Depending on what knowing requires, it may be sufficient to believe the conclusion because one has reliably detected a situation in which the premises of the inference are true; it may not be necessary to think of this situation as a situation where the premises are true, nor even to be able to think of it in this way.

Our aim in identifying the your-goal-is-my-goal is not to defend a detailed hypothesis about the mechanisms and processes involved in mindreading. Instead our present concerns are limited to a normative question about the evidential basis of mindreading. The your-goal-is-my-goal inference matters not because it describes how interpreters actually assign goals but because its existence reveals that interacting mindreaders can have access to evidence unavailable to those relying entirely on observation.

6.   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 3 on page 3). Showing how your-goal-is-my-goal makes avoiding this problem possible is a way of demonstrating the potential value to mindreaders of interaction.

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 (on page 7). Suppose the stranger starts tipping the pram in a way that the inexperienced parent fails to recognise as a means, indeed the only means, of getting it onto the bus. Outside the context of joint action this might give the parent sufficient evidence to reject the idea that the goal of the stranger's actions is to get the pram onto this bus. Or suppose the stranger starts by pulling the pram away from the bus in order to better position it for entry, and the novice parent does not realise that this is necessary. Outside the context of joint action, his evidence might on balance support 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 a means to getting the pram onto the bus, this really is the goal of her actions. Of course this additional evidence will not necessarily trump other evidence. But it does provide a possible way around the problem of opaque means by providing evidence for goal ascriptions that is independent of a mindreader's understanding of which ends actions are means to.

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, Katarina, 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 Katarina that we are willing to 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 willing 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 Katarina 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 has better understanding of relevant means-ends relations, such as actions involving novel tools, the your-goal-is-my-goal route to knowledge of other's goals is sometimes the only option.

The distant promise of all this is that understanding how interaction widens the evidential basis for mindreading 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 other agents without assuming that rich communicative skills or sophisticated forms of mindreading must already be present.

7. Communicative gestures

When introducing the problem of opaque means (in section 3 on page 3) we suggested that it could sometimes affect communicative actions. To illustrate this suggestion we drew on an experiment by Hare & Tomasello (2004) in which the chimpanzee participants 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 our 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.

Let us put ourselves in the chimpanzee's position for a moment. We witness the pointing action. With our eyes we follow

12

the point to a container (see Moll & Tomasello 2007, p. 6). But in selecting a container we are no more likely to choose this one than the other. So we associate the pointing action with its target, but we fail to ascribe any goal which would clue us in to the relevance of this target for our own actions. Now suppose that before pointing the agent had used engaging facial gestures to signal willingness to engage in joint action with us, that we were able to think of retrieving the food as a possible distributive goal[14] of our actions, and that we had exploited the my-goal-is-your-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, was very helpful).

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.[15] What we are suggesting is not that this must be wrong, but only that participants must also have been unable to exploit the your-goal-is-my-goal inference.

Consider a related experiment by Leekam, Solomon & Teoh (2010). Again the participants, who were two- and three-year-old children, had to retrieve a reward from one of several closed containers. This time they 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).

---

[14] On *distributive goal* see section 4 on page 6.

[15] See also footnote 10 on page 5. 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).

Why did the engaging facial expression enhance performance? The authors consider the idea that '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' (Leekam et al. 2010, p. 118). But 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 was not enough to motivate them to select this container, of course) and have exploiting my-goal-is-your-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 misunderstood the action as like the failed pointing in being a failed attempt to retrieve the reward.

*Point is not to defend conjecture about actual performance but to ...

*productive misunderstandings can result in understanding communicative gesture.s
***

Note that I am not suggesting that young children might fail to understand communicative intent. I am suggesting that they might first understand the goals of pointing actions without understanding communicative intent. But of course once they understand the goals of pointing actions within the context of joint action, it's likely that they will be able to understand the goals of pointing actions outside the context of joint action too.[16]

Joint action may explain how individuals starting with a simple understanding of goals end up understanding communicative intentions.

This is one illustration of how capacities for joint action, even very simple forms of joint action, might be relevant to ex-

---

[16] Contrast Csibra's 'two stances' idea. The referential action understanding involves a "stance" (p. 455); teleological and referential action interpretation "rely on different kinds of action understanding' (?, p. 456); they are initially two distinct 'action interpretation systems' (although of course they come together later in development) (?, p. 456). In relation to Csibra's, my suggestion is not that there is no referential stance. It's rather that the referential stance might emerge from what he calls the 'teleological stance' together with abilities to engage in the sort of joint actions that are characterised by shared goals.

plaining the development or evolution of richer forms of cultural cognition.

8.   The problem of false belief

We have just seen that failure to identify to which ends actions are means can impair goal ascription; this was the problem of opaque means. Another problem affecting goal ascription given purely observational models of interpretation arises from the interdependence of beliefs and goals. To illustrate, imagine sitting at a table on which two closed opaque boxes each contain an object; one contains an owl, the other a cat. If the goal of your action is to retrieve the owl, and you believe that the owl is in the north-most box, then (unless things are going very badly) you will reach into the north-most box. But of course if you had believed instead that the owl was in the other box, then, in acting on the same goal, you would have reached into the other box. Now consider someone observing your actions. If she is has good reason to believe, falsely, that you know the owl is in the south-most box, then she may be justified in supposing, incorrectly, that when you reach into the north-most box the the goal of your action is to retrieve the cat. As this illustrates, differences in belief between observers and protagonists can be an obstacle to goal ascription. Call this the problem of false belief.

There is disagreement over whether it is possible to knowledgably identify the goals of an agent's actions without also ascribing some beliefs to that agent. Bennett (1976, pp.48–50) and Davidson (1984) both appear to hold that this is impossible, that identifying goals cannot be done independently of ascribing beliefs. By contrast, Gergely, Nadasky, Csibra & Biro (1995), *Meltzoff?, Baillargeon, Scott & He (2010) and Woodward (1998) (among many others) appear to assume the opposite, that it is possible to identify goals without even being able to ascribe beliefs. For what it is worth, we tentatively favour this latter po-

sition.[17] However, this debate is not directly relevant to our concerns here. Both sides can agree that differences in belief between observer and protagonist are sometimes an obstacle to goal ascription. This is all that the problem of false belief requires.

In what follows we shall argue that abilities to engage in joint action provide an interpreter with a route to knowledge of the goals of other agents' actions which does not depend on her abilities to ascribe beliefs. This is not because abilities to engage in joint action provide a way to avoid the problem of false beliefs altogether. Rather, as we shall see, abilities to engage in joint action provide a way to shift the burden of resolving the problem of false belief from an interpreter to her target.

9. OLD

These simple facts about goal ascription raise many questions. Some concern mechanism, how in fact one subject is able to discover facts about which outcomes another agent's actions are directed to. Another set of questions focuses on the evolution of goal ascription and the costs and benefits of being able to ascribe goals and of being a potential target of goal ascription. Our concern here is not directly with any of these questions. Instead we shall focus on a more narrowly epistemic question. What evidence could support hypotheses about the outcomes to which actions are directed? And how would the evidence

---

[17] It is striking that, as far as we can tell, neither Bennett nor Davidson offers an argument for this claim. They do note that beliefs and goals make an interdependent contribution to observed action. But this by itself does not show that goal ascription cannot in some cases involve justifiably ignoring the possibility of differences in belief between interpreters and their targets. For instance, suppose that two people are sitting opposite each other at a low table which is sparsely populated with objects. The objects are all out in the open; manifestly, both can clearly see them. If one reaches to grasp one of these objects (the duck, say), must the other ascribe beliefs in order to knowledgeably identify the goal of her action? On the face of it, she need not. Even if she had no ability to ascribe beliefs, she might nevertheless be in a position to acquire knowledge of the goal of the other's action.

support the hypotheses?[18]

Of special interest is evidence available independently of any knowledge of mind or language. We want to know how it is possible to identify goals even without knowing what an agent believes or desires and even without understanding their communicative actions. Accordingly we will adopt the perspective of a goal ascriber who knows nothing about the mental states of her target agent that would distinguish this agent from any other. We will also stipulate that there is initially no common ground, shared culture or conventions. And we will stipulate that the goal ascriber is initially unable to understand any communicative actions.

There are two sorts of motivation for these restriction on the evidential basis. One is simply that developmental and comparative research indicates that goal ascription does appear to take place in such circumstances.[19] This makes it important to understand the evidence on which such ascriptions could be based. (Of course identifying evidence that could support such ascriptions would not all by itself enable us to explain how goal ascriptions are actually made, but identifying evidence is necessary if we are ever to explain the reliable success of mechanisms for goal ascription.) Another source of motivation is the conjecture that goal ascription is a prerequisite for the more sophisticated mindreading activities which reveal mental states and meanings. The coherence of this conjecture depends on the possibility of knowing something about which outcomes an agent's actions are directed to independently of knowing what she believes or desires and independently of understanding her communicative actions.[20]

---

[18] These questions are versions of those Davidson constructs a theory of interpretation to answer. While what follows draws on Davidson's insights, our aims here are more modest than his. For we are concerned only with a fraction of the problem of ascribing mental states and meanings; and, unlike Davidson, we are not concerned with larger claims about the nature of mind. See Davidson (1973, 1990); Lepore & Ludwig (2005).

[19] *refs

[20] *Compare and contrast Davidson? He did think relational attitudes (holding true) are the foundation for interpretation. But he also thought that interpretation had to happen all at once.)

So what evidence could support goal ascription by someone who knows nothing discriminating about her targets' mental states or communicative actions? Ordinary third-person goal ascription, simplified and idealized, works like this.[21] Faced with an action, the would-be goal ascriber first asks which outcomes this action could be a means to realising. She then considers which of these outcomes are potentially beneficial for, or desirable to, the agent. Any such outcomes are identified as goals to which the action is directed. So the fact that an action is a means to realising some outcome which is potentially beneficial or desirable is evidence for the conclusion that this outcome is one to which the action is directed. Schematically, the proposal is that:

(**E$_1$**) Action a is a means of realising outcome G.

and:

(**E$_2$**) The occurrence of outcome G is potentially beneficial for, or desirable to, the agent of a. (And there is no other outcome, G', which action a is a means of realising and which would be more beneficial for, or more desirable to, the agent of a.)

jointly constitute evidence for the conclusion that:

(**C**) G is a goal to which action a is directed.

This proposal might be extended in various ways. For instance, Southgate, Johnson and Csibra offer a 'principle of efficiency' according to which:

Dennett (1987, p. 17) 'Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally your predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many—but not in all—instances yield a decision about what the agent ought to do; that is what you predict the agent will do.'

[21] *ref? Dennett?

'goal attribution requires that agents expend the least possible amount of energy within their motor constraints to achieve a certain end.' (Southgate, Johnson & Csibra 2008, p. *)

If this is a correct principle of goal attribution, we could extend the proposal above to incorporate it:

(E$_3$) No alternative action, a', is a means to realising outcome G and would involve expending less energy than a.

Now the proposal is that (E$_1$) to (E$_3$) are jointly evidence for (C).

In at least some cases goal attribution is likely to be more complicated than this proposal allows. To illustrate, note that some agents may weigh the efficiency of alternative actions against their possible side effects and how reliable they would be as a means to realising an outcome. Where this is true, identifying the evidential basis for goal ascription may require a similar weighing of these factors in inferring backwards from actions to their goals.[22] Specifying exactly what should be weighed and how is beyond the scope of this paper, (and may also be something which varies between species of agent). We can mark the gap with an alternative to (E$_3$) which uses an unspecified notion of 'better' as a placeholder:

(E$_{3'}$) No alternative action, a', is a better means to realising outcome G.

This, then, is the standard approach to answering our question about goal attribution: (E$_1$), (E$_2$) and (E$_{3'}$) jointly constitute evidence for (C) given that these approximate conditions under which it would be rational to perform a in order to realise G and given that agents approximate to performing a in order to realise G rather than any other outcome under these conditions.

---

[22] This is loosely related to what Csibra and Gergely call 'the principle of rational action'. As they formulate the principle, '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' (Csibra & Gergely 1998, p. *; cf. Csibra, Bíró, Koós & Gergely 2003).

References

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110–118.

Bennett, J. (1976). *Linguistic Behaviour.* Cambridge: Cambridge University Press.

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.

Davidson, D. (1974 [1984]). Belief and the basis of meaning. In *Inquiries into Truth and Interpretation* (pp. 155–170). Oxford: Oxford University Press.

Davidson, D. ([1984] 1973). Radical interpretation. In *Inquiries into Truth and Interpretation* (pp. 125–139). Oxford: Oxford University Press.

Davidson, D. (1990). The structure and content of truth. *The Journal of Philosophy, 87*(6), 279–328.

Dennett, D. (1987). *The Intentional Stance.* Cambridge, Mass.: MIT Press.

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

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

Horner, V. & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (pan troglodytes) and children (homo sapiens). *Animal Cognition, 8*, 164–181.

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.

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.

Lepore, E. & Ludwig, K. (2005). *Donald Davidson: Meaning, Truth, Language, and Reality.* Oxford 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.

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

Wimmer, H. & Mayringer, H. (1998). False belief understanding in young children: Explanations do not develop before predictions. *International Journal of Behavioral Development, 22*(2), 403–422.

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