Original citation:

Permanent WRAP URL:
http://wrap.warwick.ac.uk/74606

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher’s statement:
© APA 2016

Link to the authoritative document on the APA website:
http://psycnet.apa.org/index.cfm?fa=buy.optionToBuy&id=2016-08871-001

"This article may not exactly replicate the authoritative document published in the APA journal. It is not the copy of record."

A note on versions:
The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher’s version. Please see the ‘permanent WRAP URL’ above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk
Is Goal Ascription Possible in Minimal Mindreading?

Stephen A. Butterfill
Department of Philosophy, University of Warwick

& Ian A. Apperly
School of Psychology, University of Birmingham

Author note

Ian A. Apperly, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

Stephen A. Butterfill, Department of Philosophy, University of Warwick, Coventry, CV4 7AL, UK.

Correspondence should be addressed to Ian Apperly, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK. Email, i.a.apperly@bham.ac.uk, tel. +44 0121 414 3339, fax. +44 0121 414 4897
Goal ascription in minimal mindreading

Abstract

In this response to the commentary by Michael and Christensen, we first explain how minimal mindreading is compatible with the development of increasingly sophisticated mindreading behaviours that involve both executive functions and general knowledge, and then sketch one approach to a minimal account of goal ascription.

Keywords: theory of mind, goal attribution, two systems theory, cognitive development, minimalism
Introduction

We very much welcome the commentary from Michael and Christensen (M&C henceforth). We agree, of course, that in offering a minimal account of belief ascription we said little about goals, that an account of mindreading is incomplete without an account of goal ascription, and that an account of minimal mindreading is incomplete, and indeed, would be threatened if goal ascription could never be achieved in a cognitively efficient manner. However, at the end of their commentary M&C offer two potential ways in which we might respond to their critique: either we must abandon the idea that minimal mindreading is strongly encapsulated and accept that it might consist of multiple systems which require integration via the use of executive functions, or else we must appeal to ‘ad hoc representations’ in offering an account of cognitively efficient goal ascription. In the first part of our response we explain why do not accept the terms of this dilemma; in particular we will clarify the degree of encapsulation that we see as necessary for minimal mindreading and explain where we think that increasing knowledge and executive processes may contribute to minimal mindreading. We then briefly sketch one approach to minimal goal ascription with the aim of showing how existing research already gives grounds for thinking that efficient goal ascription is possible, that this does not require appeal to ad hoc representations, and that such a suggestion makes distinctive, testable predictions.

It is informative to make an analogy with the well-studied case of number cognition in order to see how a relatively encapsulated cognitive process might interact with other
Goal ascription in minimal mindreading

processes. One part of infants’ early numerical abilities is the ability to enumerate sets of up to 3 objects (e.g., Feigenson, Dehaene, and Spelke 2004). This ability is initially only apparent in the relatively simple behaviours observable in young infants, such as habituation and preferential looking. But as infants develop more sophisticated motor and cognitive abilities, the same capacity influences the number of times that they will search for objects hidden in a box (Feigenson and Carey 2003) and their strategic decisions to crawl towards the more potentially rewarding of two boxes containing hidden food items (Feigenson, Carey, and Hauser 2002). Interestingly, in the latter case infants’ decisions are based on their estimate of the total quantity of available food made up by items of varying size, and not on the number of items per se. However, infants are only able to make this calculation if the total number of items in each box does not exceed 3. This result implies that infants’ ability to enumerate small sets acts in concert with an ability to quantify the amount of material in the objects comprising the set (food in this case), and also with an ability to make and act on a strategic decision to select the larger total amount of food. For our current purposes this case is helpful as an illustration of how a relatively encapsulated process comes to enable a growing range of responses and an increasingly rich set of judgements over development. At least some of these developments are likely to depend on infants’ increasing knowledge and increasing executive function, of course. But note that this does not affect the ability to enumerate sets of up to 3 objects, whose signature limit of 3-4 items remains unaffected by these other developments. This example shows that it is possible for relatively encapsulated processes to be among the causes of behaviours requiring knowledge and placing demands on executive function.
A similar situation may exist for minimal mindreading abilities. These abilities are cognitively efficient in the sense that, compared with full-blown mindreading, they trade some flexibility for reduced demands on scarce cognitive resources such as working memory and inhibitory control. We have proposed that minimal mindreading achieves cognitive efficiency by use of a stripped-down model of the mental involving registrations rather than beliefs as standardly conceived, and by being informationally encapsulated to some degree. Minimal mindreading may underpin relatively automatic belief tracking in adults (Kovács, Téglás, and Endress 2010; van der Vel, Sebanz, and Knoblich 2014; Schneider, Nott, and Dux 2014) as well as infants’ various expectations about the behaviours of agents with false beliefs (Onishi and Baillargeon 2005; Southgate, Senju, and Csibra 2007). Ascriptions of registrations may also explain older infants’ ability to help someone with a false belief (as in Buttelmann, Carpenter, and Tomasello 2009) and to selectively intervene by pointing (Knudsen and Liszkowski 2012). The latter presumably depend upon abilities to ascribe registrations operating in concert with infants’ increasingly sophisticated abilities to plan, decide and act in social situations, and success presumably depends, among other things, on infants’ increasing social knowledge and increasing memory and executive function. There is no contradiction between such developing sophistication and the idea that registrations are computed by a relatively encapsulated and cognitively efficient process, provided this process itself does not become increasingly dependent on knowledge, memory or executive function. But how could one tell whether this is indeed the case? As for the case of number cognition, positive evidence would come from finding that the signature limits remained constant even while the outputs come to have increasingly sophisticated
consequences. The evidence base on this question remains small, but it is noteworthy that both the anticipatory eye movements of 3-year-olds on an action prediction task (Low et al. 2014) and of adults on a reference resolution task (Mozuraitis, Chambers, and Daneman 2015) appear sensitive to whether a protagonist has a false belief about the existence of an object but not to whether a protagonist has a false belief about the identity of an object. This is a key signature limit of minimal theory of mind.

Just as an encapsulated process may act in concert with a variety of other abilities, so it is also coherent to suppose that the influences on an encapsulated processes may be relatively sophisticated in adults, and become increasingly sophisticated through development. There is already evidence that automatic visual perspective taking (Samson et al. 2010) and some other automatic social processes can be influenced by knowledge about whether the protagonist who is the target of these processes can see or not (e.g., Furlanetto et al. 2015; Teufel, Fletcher, and Davis 2010). These effects do not entail that representations from flexible mindreading form direct inputs to more efficient mindreading processes, but we see no decisive reason in theory why even this should be ruled out under appropriately limited circumstances. Although our account may have emphasized the need for minimal mindreading to exhibit a significant degree of encapsulation (which perhaps led to some misunderstanding of our position: De Bruin and Newen 2012; Christensen and Michael forthcoming), our project is not (and was not) to stipulate that minimal mindreading processes operate entirely independently from more flexible mindreading processes, either in their online operation or during development (for example, Apperly, 2010, pp. 137, allows that minimal mindreading processes may provide important inputs into the development of full-blown
mindreading). Rather our project was to suggest that minimal mindreading processes are independent from other cognitive processes to a degree sufficient to explain their efficient operation. In our view, such independence does require the possibility of there being a single scenario in which different mindreading processes in a single individual can result in incompatible expectations and predictions (as in Clements and Perner 1994; and Wang, Hadi, and Low 2015, for example). And it does entail that minimal mindreading processes will be subject to distinctive signature limits. But this does not preclude the possibility that the overall effects of minimal mindreading sometimes depend on knowledge.

There is also a more specific side to M&C’s critique that perhaps has most force against our account of minimal mindreading. M&C note that our account presupposes the ability to ascribe goals to actions, and that goal ascription itself can depend on tracking beliefs (perhaps by representing registrations). M&C express doubts that such goal ascription could be achieved within a minimal framework whereby the processes are relatively encapsulated. Our primary response to this challenge will be to sketch one account of minimal goal ascription, but first we must make a clarification.

The claim that an individual can only track beliefs by representing registrations does not entail that she must be similarly restricted to minimal mindreading in tracking the goals of actions. On a two-systems account of mindreading, adult humans have both relatively efficient and relatively flexible systems for mindreading, and they can track beliefs using the relatively flexible system. Why not suppose that infants also have one or more relatively flexible systems for mindreading? Perhaps some of the most convincing evidence is that typically developing children who pass tasks requiring relatively flexible
and accurate reasoning about desire, perception or pretence nevertheless fail very similar tasks involving belief (Custer 1996; Gopnik and Slaughter 1991; Gopnik, Slaughter, and Meltzoff 1994; Rakoczy, Warneken, and Tomasello 2007). Together with findings that responses based on relatively flexible belief ascription can dissociate from responses based on the efficient system within the same individual (Clements and Perner 1994; Low and Watts 2013), this indicates that belief does not typically feature in flexible mindreading until some time after abilities to reason flexibly about other mental states have been acquired. By contrast, there is presently no comparable body of evidence concerning goal ascription. We are therefore provisionally open to the possibility that one- or two-year-olds’ goal ascription may sometimes rely on a flexible system. Evidence for this would not be more surprising than the evidence just mentioned that flexible desire ascription appears earlier in development than flexible belief ascription. The development of mindreading probably involves multiple transitions if it involves any. So contra part of what we take M&C to be suggesting, the existence of some flexible goal ascription in infancy would not, by itself, be inconsistent with our account. That said, we do agree with M&C’s central claim that an account of minimal mindreading is incomplete without an explanation of how goal ascription could sometimes be achieved in a cognitively efficient manner. How might minimal mindreading involve goal ascription?

**A proposal: stage one**

A *goal* is an outcome to which an action is, or will be, directed (not to be confused with a *goal-state*, which is a state, such as an intention, in virtue of which an action has a
Goal ascription in minimal mindreading

goal). Goal ascription is the process of identifying an outcome to which an observed or anticipated sequence of bodily configurations and joint displacements are directed. How could goal ascription sufficient to explain success on a range of false belief tasks be cognitively efficient?

As success on different false belief tasks appears to involve goal ascription in several ways, let us start with the simplest. Consider a case in which identifying, at least approximately and partially, the goals of a means-end action must be done before tracking beliefs (e.g. Träuble, Marinović, and Pauen 2010). One strand of existing research on this topic hinges on the idea that motor processes and representations are involved in some cases of goal ascription. Control of action involves motor representations of outcomes that are relatively distal from bodily configurations and joint displacements, outcomes such as the grasping of a handle or the movement of an object from one place to another (Hamilton and Grafton 2008; Cattaneo et al. 2009). Importantly, motor representations of an outcome can occur not only when agents are performing an action directed to that outcome but also when they are passively observing such an action (Rizzolatti and Sinigaglia 2010). How does it come about that, often enough, an outcome represented motorically in the observer of an action is actually a goal of the observed action? One possibility is that this is due to the occurrence of motor processes in action observation. It is well-established that motor processes are planning-like in two respects. First, they involve computing means from representations of ends (Bekkering, Wohlschlager, and Gattis 2000; Grafton and Hamilton 2007). Second, they are planning-like in that they involve computing a best way to do something now in the light of things that will be done later, such as grasping a cup with a slightly awkward
posture in order to avoid a more uncomfortable posture after turning it around (Jeannerod 2006; Zhang and Rosenbaum 2007; Rosenbaum et al. 2012). Further, such planning-like processes also occur in action observation, as has been demonstrated in studies of interference effects (e.g. Brass et al. 2000; Costantini, Ambrosini, and Sinigaglia 2012), and also by measuring predictive gaze (e.g. Flanagan and Johansson 2003; Ambrosini, Costantini, and Sinigaglia 2011; Costantini et al. 2014). Planning-like processes generate expectations about how observed actions will unfold, and about their sensory consequences. Errors in these expectations would be evidence that the outcome represented motorically in the observer is not a goal of the observed action. Suppose such errors weakened the representation of an outcome in the observer: then it could be that planning-like motor processes in action observation are what ensure that, often enough, outcomes represented motorically in the observer of an action are actually goals of the observed action. In this way, a limited but useful kind of goal ascription could be achieved in which the only representations are motor representations (see further Sinigaglia and Butterfill, forthcoming).

The conjecture that motor representations and processes underpin a kind of goal ascription is relevant to us because such processes enable rapid, online action predictions, indicating that they may be cognitively efficient in the sense required for minimal mindreading. Because there are clear limits on the kinds of outcomes that can be represented motorically (which are linked to abilities to act), the conjecture also generates many testable predictions (Ambrosini, Sinigaglia, and Costantini 2012; Beets, Rösler, and Fiehler 2010; Urgesi et al. 2007; see also Michael et al. 2014). Whether or not the conjecture about motor representations and processes enabling minimal goal
Goal ascription in minimal mindreading

ascription is correct, its existence does show that theorising about cognitively efficient goal ascription need not require ad hoc postulations: instead there is a testable conjecture that coheres with a large and rapidly growing body of evidence.

Stage two: belief tracking informs goal ascription

So far we have considered only how goal ascription might inform belief tracking in situations where one can first identify the goal of an action and then ascribe a belief or belief-like state. A further challenge arises from the fact that in some false belief tasks, success apparently requires belief tracking to inform goal ascription. To illustrate, suppose Ayesha performs an action the goal of which is to retrieve some chocolate. She falsely believes that the chocolate is behind the red occluder, whereas actually it is behind the green occluder. What happens if someone attempting to identify the goals of Ayesha’s action in the way described above ignores her false belief? In this case, fixing on retrieving the chocolate as a goal of her action would generate incorrect expectations about how her action will unfold. Therefore ignoring Ayesha’s false belief would prevent correct goal ascription. But how could belief tracking inform cognitively efficient goal ascription?

Suppose the above conjecture about motor representations and processes underpinning a cognitively efficient kind of goal ascription is right. Where this kind of goal ascription occurs in false belief tasks, it must be possible for a representation of a registration to modulate expectations concerning how an observed action will unfold in much the way that a perceptual input would. But could such modulation occur? We know that when planning-like motor processes generate expectations concerning how actions
will unfold, they take into account various facts about the agent’s environment. The fact that such processes occur in motor imagery (see Jeannerod 2006) suggests that planning-like motor processes are not tied to the actual environment but can also generate expectations based on non-actual environments. Accordingly, it is at least coherent to conjecture that, when observing an agent, planning-like motor processes in the observer generate expectations by taking into account not only facts about the actual environment but also facts about the environment as specified by the agent’s registrations. This is one way in which planning-like motor processes might be modulated by belief tracking processes.³ We mention this conjecture not because our view commits us to accepting it, but only as an illustration of how belief tracking may inform goal ascription where both belief tracking and goal ascription are underpinned by cognitively efficient systems.

**Stage three: preferences and expecting actions**

In offering an account of minimal goal ascription we have yet to address one aspect of M&C’s challenge. As they note, belief tracking can involve not merely ascribing goals to actions but also forming expectations about the goals to which an agent’s actions are likely to be directed. Having observed someone repeatedly reaching for the ball when she could have been reaching for the teddy (as in Woodward 1998), you may form an expectation about the goals any future actions will have. Forming such expectations about future actions goes beyond ascribing goals, although of course such ascriptions are typically formed in part as a consequence of goal ascription and cannot be met or violated without further goal ascription. Part of M&C’s discussion asks how cognitively efficient
mindreading might support forming expectations about the goals to which an agent’s actions are likely to be directed.

This question could be answered by invoking a notion of preference. Consider this fragment of a minimal theory of preferences in which \textit{prefers} is a relation between an agent, a pair of outcomes \((o_1, o_2)\), and a probability. Intuitively, the idea will be that \textit{prefers} tracks the probability that the agent will perform an action directed to \(o_1\) rather than to \(o_2\). More carefully:

1. When an agent could perform an action with a goal that is one outcome, \(o_1\), or an action with another goal, \(o_2\), and in fact performs an action of the former kind, the probability associated by \textit{prefers} with \(a\) and the pair \(<o_1, o_2>\) increases.

2. When the agent, \(a\), could perform an action directed to either of two outcomes, \(o_1\) or \(o_2\), and the probability associated by \textit{prefers} with \(a\) and \(<o_1, o_2>\) meets a certain criterion, she will not perform an action directed to \(o_2\).

Consider a minimal mindreader whose ability to track preferences depends on a system implementing a model this fragment partially characterises. She has been observing an agent repeatedly reaching for the ball when the agent could have been reaching for the teddy. The above two principles entail that she will now expect the agent not to reach for the teddy; thus observing her reach for the teddy would violate an expectation in the minimal mindreader, consistent with the observed behaviour in (Woodward 1998).

Applying the two principles above involves being able to track which actions an agent could perform. How might a minimal mindreader do this? One possibility is that
past experience plays a role (the actions an agent could perform are the actions others have performed in this sort of situation in the past); another, complementary possibility is that motor representations and processes play a role here too (e.g. Costantini, Committeri, and Sinigaglia, 2011; Cardellicchio, Sinigaglia, Costantini, 2012).

While this fragment is far from fully characterizing the potential of minimal mindreading in relation to preferences, it does serve to show how it is possible to extend the construction of minimal theory of mind to incorporate a simple but useful preference relation.4

**Stage four: belief tracking and preferences**

Up to this point we have kept things simple by treating belief-tracking and preference-tracking as independent. But this is likely to be too simple where the aim is to characterise efficient belief-tracking in humans. The final challenge M&C offer concerns cases in which preference tracking and belief tracking jointly inform expectations about an agent’s future actions. Is such a case beyond the scope of minimal mindreading?

No. Consider observing a caterpillar whose past behaviour enables a minimal mindreader to track its preference for approaching cheese over approaching an apple (as in Surian, Caldi, and Sperber 2007). Now suppose that a situation is contrived in which the caterpillar has false beliefs about the locations of the cheese and the apple, and that the minimal mindreader is able to track these false beliefs. The above ideas about preference-tracking enable the minimal mindreader to expect that when the caterpillar can act either with the goal of approaching the cheese or with the goal of approaching the apple, it will act with the former goal. Further, the part of minimal theory of mind that
Goal ascription in minimal mindreading

deals with registration enables the mindreader to expect that, when the caterpillar acts on a goal involving the cheese, its actions will accord with what it registers concerning the cheese rather than only with the facts about the cheese. So while we expect that adding further principles linking preferences, registrations and the goals of actions will be necessary in constructing a minimal theory of mind, the theory as we have elaborated it so far already enables representations of preferences and registrations to have a common effect on a mindreader’s expectations about another’s future actions.

Conclusion

We find ourselves in strong agreement with M&C that an account of minimal mindreading must concern not only belief-like states but also goals and preferences; emotion is surely also critical. But our approach differs from theirs in some ways. We are not yet convinced that the forms of goal ascription M&C discuss are beyond the theoretical limits of minimal mindreading. More fundamentally, whereas M&C are primarily concerned with evidence from mindreading in infants, our objective in developing an account of minimal mindreading is to make sense of evidence indicating cognitively efficient mindreading in adults as well as precocious mindreading abilities in infants. An important upshot of this is that the account generates predictions about signature limits which allow minimal mindreading to be identified across different paradigms and different participant groups, including adult and infant humans as well as non-humans. In the present response we hope to have illustrated the usefulness of this approach by drawing from research on motor cognition in adults to sketch the first steps in the construction of minimal goal ascription.
Goal ascription in minimal mindreading
Goal ascription in minimal mindreading

References


Goal ascription in minimal mindreading


Goal ascription in minimal mindreading


Goal ascription in minimal mindreading


Goal ascription in minimal mindreading


Goal ascription in minimal mindreading
Footnotes

1 M&C suggest that goal ascription depends on ascribing agents beliefs about means-ends relations in this case; we suggest this is unnecessary as facts about means-ends relations may be taken for granted.

2 As M&C anticipate, we think minimal goal ascription is to some degree ‘limited in the extent to which it can incorporate functional type information’; however, the possible link between minimal mindreading and motor processes shows that minimal goal ascription can sometimes reflect both the context in which actions occur and also functional properties of objects (Fogassi et al. 2005; Iacoboni et al. 2005; Cattaneo et al. 2007; Hamilton and Grafton 2008; Bonini et al. 2011; Costantini et al. 2011). Minimal goal ascription is not, contra M&C, ‘unable to incorporate situational goal attribution.’

3 Van der Wel, Sebanz, and Knoblich (2014) provide evidence that information about another’s belief can systematically perturb processes underpinning action performance. This may lend some initial plausibility to our conjecture that information about another’s belief can modulate motor processes underpinning goal ascription.

4 The fragment of a minimal theory of preferences also illustrates one way in which minimal theory of mind might unproblematically involve what M&C call ‘agentic linking’ or ‘the association of a goal with an agent’.