Ascribing Proto-Intentions: Action Understanding as Minimal Mindreading

Chiara Brozzo

ABSTRACT: How do we understand other individuals' actions? Answers to this question cluster around two extremes: either by ascribing to the observed individual mental states such as intentions, or without ascribing any mental states. Thus, action understanding is either full-blown mindreading, or not mindreading. An intermediate option is lacking, but would be desirable for interpreting some experimental findings. I provide this intermediate option: actions may be understood by ascribing to the observed individual *proto-intentions*. Unlike intentions, proto-intentions are subject to context-bound normative constraints, therefore being more widely available across development. Action understanding, when it consists in proto-intention ascription, can be a minimal form of mindreading.

1. Introduction

I watch you move your hand towards a teacup, and I understand that your movements are directed towards picking up that teacup. This is an instance of *action understanding*, namely the process by means of which someone identifies the outcome to which a series of movements are directed.[1] An *outcome* is here to be understood as a possible or actual state of affairs—for example,

---

[1] The notion of *directedness* is here used to distinguish outcomes that are purposely brought about from those that are accidentally brought about. In the example just given, your movements are *directed* towards the outcome of the teacup being picked up. By contrast, if you moved in such a way as to *accidentally* spill

a teacup being picked up—that is the result of a series of movements. *Action* is here used interchangeably with *event*: there is no presupposition that, when the action is understood, it is understood as such—namely, as Anscombe (1957) and Davidson (1963) would have put it, as intentional under a description (see also Smortchkova 2018).

How do we understand other individuals' actions? Answers to this question tend to cluster around two extremes. On the one hand, it may be thought that actions are understood by ascribing to the observed individual a mental state representing the outcome being brought about (Goldman 2006 considers this possibility—see section 3). In the previous example, I would understand your movements as directed to the outcome of the teacup being picked up by ascribing to you, e.g., an *intention* to pick up the teacup, or to drink tea.[2] In other words, action understanding would be a form of *mindreading*, which is standardly conceived as the ascription of mental states—propositional attitudes, but also emotions—to others or to oneself (see, e.g., Stich and Nichols 1992; Goldman 2009).[3]

Connecting action understanding to standardly conceived mindreading requires that an observer engaged in action understanding is equipped with relevant mental state notions, such as that of intention. These notions may in principle be rather cognitively demanding, entailing, e.g., relations to many other mental states. For example, I may ascribe to you the intention to drink

---

[2] The role of intention in the explanation of purposive behaviour has been amply discussed by Kathy Wilkes (see, for example, Wilkes 1989).

the tea contained in the cup, it would not be the case that your movements were directed towards the outcome of the tea being spilled (see Sinigaglia and Butterfill 2015).

[3] Notice that this possibility about how action understanding works does not trivially follow from the definition of action understanding. This is because both the notion of outcome and that of directedness to an outcome are devoid of reference to mental states: a series of movements may be directed to a given outcome without there being any mental states representing that outcome. For example, the movements of a mechanical arm may be directed to the outcome of a teacup being picked up, without there being any mental state representing that outcome.

tea in conjunction with the intention to be a bit more awake, or to be a good host and keep me company in drinking tea, but not in conjunction with ascribing to you an abhorrence of tea, absent a further ascription to you of pressing reasons to nevertheless drink it.

On the other hand, at the other extreme, it may be thought that actions are understood in a way that does not draw on mental state ascription at all. Rather, action understanding is exhausted by relating observed movements to anticipated—and eventually observed—outcomes (Gergely and Csibra 1997, 2003; Roessler and Perner 2010; Spaulding 2013). This would make action understanding more similar to processes whereby most human adults understand physical interactions such as causal ones: to understand that a billiard ball has set another into motion through collision, no mental states are ascribed to either ball.

Lots of experimental research on action understanding, one example of which I shall illustrate later (Behne et al. 2005), suffers from the lack of an intermediate theoretical option. According to the two aforementioned options, actions are either understood through the ascription of cognitively demanding mental states, or without the ascription of any mental states at all. As sections 5.1 and 5.2 will show, neither option seems adequate in some cases of action understanding.

In answer to this impasse, this paper will put forward a proposal about what action understanding could involve that lies midway between the aforementioned two extremes (see also Andrews 2020). According to it, differently from the second extreme option, action understanding would involve the ascription of some mental states. Differently from the first extreme option, however, the mental states ascribed in understanding others' actions would not be as cognitively demanding as in full-blown mindreading.[4]

---

[4] The proposal I am going to put forward has analogous motivations as the proposal made by Butterfill and Apperly (2013; see also Apperly and Butterfill 2009). According to it, some creatures could ascribe to others mental states called *registrations*. These are like beliefs in some respects, but also simpler than beliefs, in a way that will be clarified in section 5. Butterfill and Apperly's proposal is motivated by the need to

In the following sections, I shall, first, describe in detail the possibility that action understanding could consist in full-blown mindreading (sections 2-3). After that, I shall illustrate how action understanding might involve no mental state ascription (section 4). Making clearer the commitments of these extreme options will lay the ground for putting forward my own middle ground proposal (sections 5-ff.).

2. Under what conditions would action understanding be mindreading?

In order to present the first extreme option, according to which action understanding is full-blown mindreading, I want to clarify under what conditions action understanding would be mindreading.

Action understanding would *not* be mindreading in cases such as the following. Suppose that I observe Alice moving towards a teacup. Suppose that I understand Alice's movements as directed to the outcome of the teacup being picked up. Action understanding is complete at this point: the teacup being picked up has been identified as the outcome to which Alice's

interpret certain findings in developmental psychology that the notion of belief is inadequate to explain (e.g., Onishi & Baillargeon 2005). The focus of my proposal, unlike Butterfill and Apperly's, is mental states that are like intentions in some respects, but also simpler than intentions. My proposal is independent of Butterfill and Apperly's: for reasons that will become clear later, the tenability of one does not hinge on the tenability of the other, and vice versa.

Also, my proposal assumes that mental states, whether minimal or full-blown, are representations of sorts, and will therefore not engage with anti-representationalist views of the mind (e.g., Gallagher and Hutto 2008). Lastly, my proposal is not to be seen as an alternative to either Simulation Theory (Gordon 1986; Heal 1986; Goldman 2006) or Theory Theory (Gopnik and Wellman 1992; Gopnik and Meltzoff 1997), as it is, in principle, compatible with both.

movements are directed. Then, subsequently, I additionally ascribe to Alice the intention to have a leisurely cup of tea. But this would not make the previous instance of action understanding an act of mindreading: the act of mindreading (ascribing to Alice the intention to have a leisurely cup of tea) would be *distinct* from that of action understanding (identifying Alice's movements as directed to the outcome of a teacup being picked up). In this example, mindreading begins when action understanding is already over. The moral of this example is that ascribing a mental state once an outcome has already been identified as that to which an observed series of movements are directed does *not* make action understanding an instance of mindreading. By contrast, action understanding would be mindreading if it involved ascribing mental states—either because mental state ascription is part of the process of action understanding, or because it is identical to it. For example, action understanding would be mindreading if ascribing to Alice the intention to have a leisurely cup of tea had a causal role in concluding that her movements are directed towards the outcome of the teacup being picked up.

3. Intention ascription: full-blown mindreading

Let me now present the option according to which action understanding would be full-blown mindreading, in line with the provisos offered in the previous section. I shall present one way for action understanding to be mindreading, which consists in ascribing an intention to an individual performing the action.

How does intention ascription relate to understanding actions? By virtue of the widely shared view that intentions represent or otherwise specify outcomes (see, e.g., Searle 1983; Bratman 1987). For instance, the intention to build a house represents the outcome of a house being built. Therefore, by ascribing an intention to build a house to an individual, one thereby identifies an outcome—the outcome represented by that intention—to which this individual's

action is directed. This leaves it open whether the observed individual actually *has* an intention to bring about the outcome (see Borg 2007; Sinigaglia 2008).

Of course, mental states other than intentions also represent outcomes—for example, beliefs and desires. Here I am assuming that the mental states ascribed in some cases of action understanding are intentions. Why? Because not only do intentions represent outcomes, and to this extent they are akin to beliefs, but, unlike beliefs, intentions represent outcomes with a world-to-mind direction of fit and a mind-to-world direction of causation (see, e.g., Searle 1983): in order for intentions to be fulfilled, the world has to conform to them, and intentions contribute to the required changes in the world. So, they are fit to fulfil the role of causes of the observed behaviour, unlike beliefs.

One could object that the same considerations about direction of fit and direction of causation make desires just as plausible candidate mental states to be ascribed in action understanding. Intentions and desires, however, differ in the following way. According to a standard conception, intentions are tools for planning. This is reflected in their being subject to characteristic normative constraints concerning consistency and rationality—in particular, what Bratman (1987) termed the *strong consistency requirement*. An intention satisfies the strong consistency requirement if and only if it is consistent with the rest of the subject's intentions, as well as with the rest of the subject's beliefs. It is a normative constraint to the extent that intentions should satisfy it in order to fulfil their role as tools for planning, but it is conceivable that intentions may break it (for example, I may intend to get ready in fifteen minutes all the while believing that I will take at least half an hour). If they do, then the subject is guilty of irrationality. No such normative requirement applies to desires (Bratman 1987; Holton 2009). In particular, having conflicting desires does not make a subject irrational. I think that, due to the applicability of these normative constraints, intentions are better suited than desires to account for consistency relationships between ends and means that are recognised by certain subjects, as will be illustrated in the next section with the experiment by Gergely and colleagues (1995). For

this reason, in what follows I shall focus on intentions as candidate mental states to be ascribed in the context of action understanding. I do concede that, if action understanding consisted in desire ascription, it would also be full-blown mindreading, but I leave a discussion of the case of action understanding consisting in desire ascription for another occasion.

Now I shall provide an example of how action understanding could be intention ascription, and therefore full-blown mindreading. This will consist in a specific version of the so-called *generate-and-test* model, introduced by Goldman as follows:

> The attributor begins with a known effect of a sought-after state, often an observable piece of behavior. He generates one or more hypotheses about the prior mental state or combination of states that might be responsible for this effect. He then "tests" […] these hypotheses by pretending to be in these states, feeding them into an appropriate psychological mechanism, and seeing whether the output matches the observed evidence. When a match is found […], he attributes the hypothesized state or combination of states to the target. (Goldman 2006, p. 45)

Action understanding involving intention ascription would take place if the generate-and-test model were instantiated with the following auxiliary assumptions. First, an observer hypothesises that the prior mental state responsible for the observed behaviour of another individual is an *intention* to bring about a certain outcome, and the observable behaviour consists in the bodily movements bringing about that outcome. Furthermore, the observer eventually does find a match between, on the one hand, the bodily movements that the hypothesised intention would produce and, on the other hand, the observed bodily movements. Therefore, the observer ascribes the hypothesised intention to the observed individual. Since the intention represents the outcome to which the observed movements are directed, this would be a case of

action understanding. Therefore, one may understand an action by ascribing an intention.[5] This would make action understanding full-blown mindreading.

4. Mere outcome identification: not mindreading

A second extreme option concerning how action understanding could take place is without any ascription of mental states. I shall call this option *mere outcome identification*. Here is an example of it.

According to Gergely and Csibra (1997; 2003; Csibra and Gergely 1998), outcomes (which they call *goal-states*) are identified by one-year-olds thanks to the *teleological stance*. The teleological stance is an interpretational schema featuring three elements: an outcome, an action (a term that Gergely and Csibra use as synonymous with *a series of movements*—in line with its meaning in *action understanding*) and a set of situational constraints. An individual understanding an action by means of the teleological stance identifies all three elements, and moreover identifies actions as directed to (which may be read as *supposed to bring about*) certain outcomes. Between

---

[5] The generate-and-test model is put forward within the framework of the Simulation Theory of mindreading, according to which an observer ascribes mental states to an observed individual by means of an attempt to replicate the workings of the latter's mind (Gordon 1986; Heal 1986; Goldman 2006; Gallese and Goldman 1998). This is reflected in the fact that, according to the generate-and-test model, the observer tests the hypothesised intention by *pretending to have that intention themselves*. Notice, however, that intention ascription can take place outside the simulationist framework. The generate-and-test model itself could be modified so as not include a commitment to the Simulation Theory, for example as follows. A subject hypothesises that the observed individual has a certain intention, and then draws on a theory about how intentions connect with ensuing bodily movements in order to make the relevant predictions about the movements that she should observe, were the observed individual to have the hypothesised intention.

different actions directed to the same outcome, this individual is further capable of identifying the most rational action for bringing about the outcome given the current situational constraints (e.g., in the absence of an obstacle, approaching a target in a straight line is more rational than approaching it via a curved path). All this, according to Gergely and Csibra, is done without ascribing any mental states.

Here is an example of the teleological stance at work. In a violation-of-expectation study (Gergely et al. 1995), infants were habituated to a computer animation showing a small circle approaching a large one. In this animation, the small circle moved along a trajectory that looked like a jump, through which it approached the large circle while avoiding a rectangular obstacle. In the context of this computer animation, the outcome was the large circle being reached, the action consisted in the movements of the small circle, and the situational constraints consisted in the presence of the rectangular obstacle. After the infants had been habituated to this animation, they were shown two test displays. In both of them, the obstacle was removed. In one of the two test displays, the small circle approached the large circle in a straight line. In the other test display, the small circle approached the large circle following the same trajectory as in the habituation display, i.e. a trajectory that looked like a jump. Infants looked longer (which is taken to indicate surprise) at the latter test display than at the former.

Gergely and Csibra's (1997, 2003) interpretation is as follows. First, infants showed sensitivity to the directedness of movements to an outcome. In particular, they recognised that the small circle's movements were directed to the outcome of the large circle being reached. Furthermore, these infants recognised situational constraints—the presence of the obstacle in the habituation display or of an unblocked path in the test displays. Lastly, infants were capable of recognising the straight-line approach as more rational than the jump-like approach for bringing about the outcome (the large circle being reached) under the given situational constraints (an unblocked path). But this, according to Gergely and Csibra, happened without any representation (and, *a fortiori*, ascription) of mental states on the infants' part, and indeed there was no

presupposition that the moving circle observed by the infant had a mind at all. On the contrary, the infant identifying a certain outcome being brought about is described as a *'mindblind' creature* (Gergely and Csibra 2003, p. 290).

It should at this point be clear that, if action understanding consisted in mere outcome identification (for example, in the form of the teleological stance; see also Perner and Roessler 2010), then action understanding would not be mindreading.

Let me take stock so far: I have singled out two extreme options for how action understanding could occur:

1. *Intention ascription.* An observer identifies an outcome to which a series of movements are directed by ascribing to an observed individual an intention representing that outcome. If action understanding consisted in intention ascription, it would be full-blown mindreading.

2. *Mere outcome identification.* An observer merely identifies an outcome to which a series of movements are directed, without ascribing any mental states. If action understanding consisted in mere outcome identification, it would not be mindreading.

Now I am ready to present my own proposal about how action understanding could occur, situating it midway between full-blown mindreading and the absence of mindreading.

5. A third option: minimal mindreading

Up to now, my discussion has been confined to two rather extreme options: in action understanding, either one ascribes to an observed individual an intention (so that action understanding is full-blown mindreading), or one identifies an outcome to which an action is

directed (so that action understanding is not mindreading). I would now like to point out that the following middle ground should be explored: action understanding could be a minimal form of mindreading. A minimal form of mindreading occurs when minimal forms of mental states—i.e., mental states less cognitively demanding than propositional attitudes—are ascribed (see, e.g., Tomasello et al. 2003; Nanay 2013; Whiten 2013; Butterfill and Apperly 2013).

An example of a mental state less cognitively demanding than a propositional attitude is provided by *registrations*, postulated by Butterfill and Apperly (2013) as part of their proposed Minimal Theory of Mind. A registration is a relation between a subject, an object and a location. Like beliefs, registrations have correctness conditions insofar as an individual correctly registers an object at a location if and only if that object is actually at that location. Due to being relations rather than representations, however, unlike beliefs registrations are not sensitive to different modes of presentation of one and the same object (see Butterfill and Apperly 2013; Low and Watts 2013). Because of this, they are less cognitively demanding (and therefore more widely available within and across species) than beliefs. I will now provide some motivation for exploring minimal forms of intention.

5.1 Some motivation for the ascription of minimal forms of mental states

I shall now illustrate an experimental result that I shall term *unwilling vs. unable* (Behne et al. 2005). I will then explain how a possible interpretation of this result motivates considering the idea that action understanding could consist in the ascription of minimal forms of mental states— specifically, minimal forms of standardly conceived intentions.

Behne and colleagues (2005) tested infants from 6 to 18 months of age as follows. An infant was faced with an adult experimenter in the position to pass them an object. The infant was presented with both the following kinds of scenario at different times: in some of these, the experimenter did not pass the object to the infant because the experimenter was *unwilling* to do so; in other scenarios, the experimenter did not pass the object to the infant because the

experimenter was *unable* to do so (for example, the object slipped out of their hands). While 6-month-olds were not sensitive to the difference between unwilling vs. unable, infants from 9 months of age onwards were more *impatient* in the scenarios in which the experimenter was unwilling to pass them the object than in those in which the experimenter was well-meaning but clumsy, and therefore unable to pass them the object (compare Call et al. 2004 for a similar paradigm with chimpanzees).

There are at least two possible ways of accounting for the different reactions observed in the subjects of the above reported experiment: in terms of mere outcome identification or in terms of mental state ascription. According to an interpretation in terms of mere outcome identification, the infants from 9 months of age onwards identified the outcome to which the experimenter's action was directed. In particular, they understood the experimenter's movements in the *unwilling* condition as directed to the outcome of the object being withheld, and in the *unable* condition as directed to the outcome of the object being passed to them (though the experimenter failed to bring it about). According to an interpretation in terms of mental state ascription, the infants from 9 months of age onwards ascribed a *mental state* to the experimenter, one that represents the outcome to which the experimenter's movements are directed (object being withheld vs. object being passed to the infant). Absent any independent considerations, prima facie there is no reason to exclude an interpretation in terms of mental state ascription (see Michael and Christensen 2016 for doubts that interpretations of similar results in terms of mere outcome identification are adequate). Once this option is on the table, the question now is: if infants do ascribe a mental state to the experimenter, what is the mental state in question?

5.2 Why the ascription of standardly conceived intentions will not do

In section 3, I explored the possibility that the mental state ascribed in action understanding could be a standardly conceived intention. But this does not look like a viable option for the experiment reported in the previous section. Why not? In section 3, I described intentions as

states that are subject to normative constraints concerning consistency and rationality (Bratman 1987; Holton 2009). This makes intentions cognitively demanding: representing an intention implies being sensitive to the fact that this intention should not conflict with many of one's intentions and beliefs. Now, it is plausible to assume that the complexity of a mental state imposes constraints on the ease of identification and ascription of such mental states (see, e.g., Butterfill and Apperly 2013). Working on this assumption, and on the assumption that intentions are relatively complex mental states due to the normative constraints applying to them, it is highly implausible that creatures such as infants ranging from 9 to 18 months of age should be able to represent and ascribe standardly conceived intentions, insofar as this would place too high demands on their inferential abilities. Call this line of reasoning *can't have*.

Can't have makes it worthwhile to explore mental states representing outcomes to which actions are directed that are different from standardly conceived intentions—different insofar as their representation does not impose as high demands on infants' inferential abilities as standardly conceived intentions. In other words, these mental states should be such that infants between 9 and 18 months of age can represent and ascribe them.

Another consideration in favour of exploring this option, which I shall call *needn't have*, is that even creatures such as human adults, who could plausibly represent and ascribe standardly conceived intentions, *do not need* to ascribe anything as complex as that when they have to, e.g., tell someone who is unwilling to perform a certain bodily action apart from someone who is unable to do so.

In short, under the assumption that we cannot rule out an interpretation of the above reported experiments in terms of mental state ascription, the *can't have* line of reasoning provides motivation for exploring the option that minimal forms of mental states could be ascribed in action understanding. Independent motivation for exploring this option is given by the *needn't have* line of reasoning.

6. Proto-intention: a minimal form of intention

In this section, I will present a minimal form of intention that I shall call *proto-intention*. Just like intentions, proto-intentions are mental states with a world-to-mind direction of fit and mind-to-world direction of causation. However, differently from intentions, they represent outcomes in a less cognitively demanding way, modelled on a kind of outcome identification that, following Tomasello and colleagues (2005), I shall call *tracking the choice of plans*. In section 6.1, I will say what tracking the choice of plans is. In section 6.2, I will show how ascribing proto-intentions enables tracking the choice of plans while posing inferior cognitive demands on a subject's inferential abilities compared to ascribing standardly conceived intentions.

6.1 Tracking the choice of plans

Tracking the choice of plans consists in identifying an outcome to which an action is directed while also telling it apart from the specific means with which it was achieved. Several experiments can be taken to indicate that their subjects have the ability to track the choice of plans. One of them is the Gergely and colleagues' (1995) experiment described in section 4, where subjects can be interpreted as able to tell apart the outcome to which the observed movements are directed (i.e., the large circle being reached) from the means with which this is achieved (straight-line approach vs. jump-like approach). This kind of understanding is to be contrasted with an even less demanding kind of outcome identification: *tracking the pursuit of outcomes* (see Tomasello et al. 2005).

Tracking the pursuit of outcomes, by contrast with tracking the choice of plans, merely involves understanding that there is an outcome that the observed individual is bringing about through certain means, and that they will persist in pursuing it until this is achieved, without any variation in the means employed to this end. Roughly, understanding choosing plans involves understanding at least two things, namely at least two different means in which the same

outcome may be achieved. Understanding the pursuit of outcomes, by contrast, involves understanding just one thing: that an outcome might be achieved by certain means.

Suppose that Gergely and colleagues' (1995) experiment, described in section 4, is interpreted as one in which action understanding involves some form of mindreading. I am leaving it open whether this is actually the case, but note that Gergely and colleagues themselves previously supported an interpretation of their own results in terms of mental state ascription (Gergely et al. 1995). Under an interpretation in terms of mental state ascription, what could the mental states ascribed by the infants be? My proposal is that they could be *proto-intentions*.

6.2 Proto-intention ascription enables tracking the choice of plans

In this section, I shall characterise proto-intentions as mental states partly analogous to intentions but subject to more local normative constraints concerning consistency and rationality.[6]

The way I shall characterise proto-intentions assumes that proto-intentions could be both *states one has*, i.e. that are part of someone's psychology, as well as *states one ascribes* to other individuals. This is one of the main differences between my proposed minimal forms of mental states and Butterfill and Apperly's registrations, described in section 5: registrations are *not* supposed to be part of anyone's psychology, but rather useful tools for explanation and prediction on the part of the individual that ascribes them. In other words, registrations are supposed to be *states one ascribes*, but not necessarily *states one has*.[7]

---

[6] I will focus on differences in normative constraints between proto-intentions and intentions, while leaving it open that they may differ also in other respects.

[7] A theory about how action understanding takes place that draws on states one has is amenable to being framed in terms of the Simulation Theory, in which the states one ascribes are precisely the states one has (I am here just pointing out the possibility of doing so, but I shall not pursue it in this paper). Note that, assuming that the states one ascribes are the states one has, there are deep and difficult questions about what enables subjects that have certain mental states to also ascribe them to other individuals (see, e.g., Tomasello 1999; Tomasello and Call 1997; Hurley 2003; Peacocke 2014).

Why think that proto-intentions could exist? Based on the idea that proto-intentions are both states one has and states one ascribes, support for the idea that some creatures might have proto-intentions comes from reflections on animal cognition made by Susan Hurley (2003).

Recall from section 3 that intentions are subject to characteristic normative constraints, such that one's intentions should (ideally) be consistent with the rest of one's beliefs and intentions. Notice that there is no principled boundary on the number of intentions and beliefs that one's intentions should not conflict with. Suppose, for example, that I intend to spend tomorrow writing a book chapter. Suppose that someone invites me to join them on a leisurely day out, walking in the countryside. Should I settle on that course of action, thereby forming an intention to spend tomorrow walking the countryside? A quick inference leads me to conclude that spending tomorrow walking the countryside means I will not do any writing. This conflicts with my intention to spend tomorrow writing a book chapter—which speaks in favour of not forming an intention to spend tomorrow walking the countryside. However, I also believe that it would be good for me to do some exercise—something that the doctor recently advised me to do, and I intend to follow his advice. Another inference leads me to conclude that spending tomorrow walking in the countryside would be the perfect way to follow my doctor's advice. But then I also believe that my book is overdue, which speaks in favour of my original intention… and so on. This is an illustration of how standardly conceived intentions presuppose in principle unbounded inferential abilities.

By contrast, Hurley (2003) pointed out that there is an interesting normative middle ground between the full-blown rationality that norms on intentions seem to require and, on the other hand, the complete absence of norms of rationality (which would be the case with mere responses to stimuli, for example). In particular, according to Hurley, "[n]on-human animals can occupy islands of practical rationality: they can have *context-bound* reasons for action" (2003, p. 231, my emphasis). To make things more concrete, Hurley considers the following possibility, based on observations by Cheney and Seyfarth (1992), Tomasello (1999) and Tomasello and Call

(1997). The possibility is that some animals (e.g., chimpanzees) could make transitive inferences in some contexts (e.g., social contexts) but not others (e.g., non-social contexts). For example, Hurley conjectures that a chimpanzee could make transitive inferences of the kind "A is dominant over B, B is dominant over C, therefore A is dominant over C" (where A, B and C are conspecifics), but not of the kind "A has more fruit than B, B has more fruit than C, therefore A has more fruit than C" (where A, B and C are trees). This would enable the chimpanzee to use the former, but not the latter, kind of information to guide her actions flexibly in relation to various goals (e.g., holding fixed the goal of retrieving food, she may do so following different paths depending on the circumstances). In other words, some animals' reasons for acting may be context-bound, that is, not generalise to all possible contexts. Inference to the best explanation would then make it plausible that, if actions whose reasons are context-bound could be driven by mental states, then these would have to be mental states that are subject to more local normative constraints than intentions. These are what I shall call *proto-intentions*: states with a world-to-mind direction of fit and mind-to-world direction of causation that are subject to a limited form of the strong consistency requirement. A limited form of this normative requirement merely prescribes that an individual's proto-intention should not conflict with:

(i) another proto-intention of that individual that is linked to the former via means-end reasoning, and

(ii) with information that the individual has about how one's end could be achieved in the circumstances (i.e., information about what Gergely and Csibra called *situational constraints*).

As an example, a proto-intention to reach another individual by following a straight-line path should be consistent both with one's intention to reach the other individual and with the information one has about the obstacles present on that path at the moment, and not with information spread across longer timescales (the latter illustrated in the previous example, concerning the intention to spend tomorrow walking the countryside).

At this point, I have introduced the notion of proto-intention to explain some cases of action production. Working on the assumption that proto-intentions are states that one has but also that one can ascribe, I shall now present the following possible way in which action understanding might occur:

3. *Proto-intention ascription.* One could identify an outcome to which an action is directed by ascribing to an observed individual a proto-intention representing that outcome. Given that proto-intentions are minimal forms of intentions, if action understanding consisted in proto-intention ascription, it would be a minimal form of mindreading.

The reason why ascribing a proto-intention would be a useful strategy for identifying outcomes is that proto-intentions, just like intentions, represent outcomes with a world-to-mind direction of fit and have a mind-to-world direction of causation. They differ in the normative constraints to which they are subject. Intentions are subject to the strong consistency requirement, and, as a result, representing and ascribing intentions presupposes in principle unbounded inferential abilities—or at least, rather cognitively demanding inferential abilities. By contrast, proto-intentions are subject to a limited form of the strong consistency requirement, and, therefore, in order to represent and ascribe proto-intentions, one need only have inferential abilities that enable the evaluation of different (typically two) potential means for achieving the same outcome. This makes proto-intentions in principle more widely available across development.

Up to this point, two different types of mental states with a world-to-mind direction of fit and mind-to-world direction of causation have been distinguished: standardly conceived intentions (which are subject to the strong consistency requirement) and proto-intentions (which are subject to a local form of the strong consistency requirement).

6.3 Proto-intentions are *not* intentions in action or proximal intentions

One clarification is now in order. In the philosophical literature on action production, occasionally it has been suggested there is a variety of intention that is supposed to trigger and guide the course of the action it represents. Intentions of this variety are known under various names, depending on different conceptions: *intentions in action* (Searle 1983), *proximal intentions* (Mele 1992), *present-directed intentions* (Bratman 1987, Pacherie 2006, 2008).[8] I am here clustering them together in virtue of their functional commonality: that of triggering and guiding the course of action they represent.

An interesting question is whether these states should be considered an additional variety with respect to standardly conceived intentions and proto-intentions. Answering this question relies on taking a stance on an issue that, I believe, so far has not received enough attention: whether and to what extent intentions in action and similar conceptions are subject to normative constraints concerning consistency and rationality (cf. Mylopoulos and Pacherie 2017). Depending on which stance is taken on this issue, intentions in action, present-directed intentions and proximal intentions may be either intentions standardly conceived, or proto-intentions. But, in the absence of any clarifications on this issue, it is wrong to assume that proto-intentions *just are* intentions in action, present-directed intentions or proximal intentions.

7 Conclusion

This article started with the observation that action understanding has mainly been interpreted in terms of two very extreme options: either as involving the ascription of standardly conceived intentions, which would make action understanding a form of full-blown mindreading, or as

---

[8] Pacherie (2006, 2008) has also introduced a variety of these intentions that she calls *motor intentions*. In her later work (see, e.g., Mylopoulos and Pacherie 2017), Pacherie no longer classifies them as intentions. I will therefore not consider them here.

involving no mental state ascription at all. I have given reasons for considering a middle ground between these two extreme options. Two considerations support the exploration of this middle ground. On the one hand, one may think that some creatures (e.g., infants of 9 to 18 months of age) *can't* represent standardly conceived intentions. On the other hand, some creatures capable of representing standardly conceived intentions sometimes *needn't* represent them, given the characteristics of the action they are in the position of understanding. Either consideration suffices to consider the following alternative: that action understanding might involve a minimal form of mindreading.

I have presented and explored one way in which action understanding to be minimal mindreading—one that involves the ascription of proto-intentions. By contrast with the posits of Minimal Theory of Mind, proto-intentions are states that one has, and not just states that one ascribes, and they are representations rather than relations.

To sum up, here are the options explored so far concerning how action understanding might occur, together with an indication as to whether they are a form of mindreading and, if so, which form:

In what does action understanding consist?	Is it mindreading?
Mere outcome identification	No
Proto-intention ascription	Yes (minimal)
Intention ascription	Yes (full-blown)

The notion of proto-intention can help interpret experiments, such as that by Behne et al.

(2005), that have the following characteristics: on the one hand, it would be explanatory advantageous to suppose that the experimental subjects ascribe some form of mental states to the observed individuals, but, on the other hand, we might be reluctant to think of these subjects as mastering cognitively demanding mental state notions, such as the standard one of intention. Proto-intentions, by contrast, are apt to be represented and ascribed more widely across development.

References

Andrews, K. (2020). Naïve normativity: The social foundation of moral cognition. *Journal of the American Philosophical Association*, *6*(1), 36-56.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological review*, *116*(4), 953.

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental psychology*, *41*(2), 328-337.

Borg, E. (2007). If mirror neurons are the answer, what was the question?. *Journal of Consciousness Studies*, *14*(8), 5-19.

Bratman, M. (1987). Intention, plans, and practical reason. Cambridge, MA: Harvard University Press.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind &*

Language, *28*(5), 606-637.

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). 'Unwilling' versus 'unable': chimpanzees' understanding of human intentional action. *Developmental science*, *7*(4), 488-498.

Cheney, D. L., & Seyfarth, R. M. (1992). *How monkeys see the world: Inside the mind of another species.* Chicago: University of Chicago Press.

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

Davidson, D. (1963). Actions, reasons, and causes. *The journal of philosophy*, *60*(23), 685-700.

Gallagher, S., & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. *The shared mind: Perspectives on intersubjectivity*, *12*, 17-38.

Gallese, V., & Goldman, A.I. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences, 2*, 493-501.

Gergely, G., & Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to Premack and Premack. *Cognition*, *63*(2), 227-233.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in cognitive sciences*, *7*(7), 287-292.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165-193.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* Oxford University Press.

Goldman, A.I. (2009). Mirroring, mindreading, and simulation. In J. A. Pineda (Ed.), *Mirror neuron systems: The role of mirroring processes in social cognition* (pp. 311–330). New York: Humana Press.

Gopnik, A., and Meltzoff, A. N. (1997). Words, Thoughts and Theories. Cambridge, MA: MIT Press.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, *7*(1-2), 145-171.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, *1*(2), 158-171.

Heal, J. (1986). "Replication and Functionalism", in *Language, Mind, and Logic*, J. Butterfield (ed.), Cambridge: Cambridge University Press.

Holton, R. (2009). *Willing, wanting, waiting*. Oxford University Press.

Hurley, S. (2003). Animal action in the space of reasons. *Mind & Language*, *18*(3), 231-257.

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*(3), 305-311.

Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. Oxford University Press.

Michael, J., & Christensen, W. (2016). Flexible goal attribution in early mindreading. *Psychological Review*, *123*(2), 219-227.

Mylopoulos, M., & Pacherie, E. (2017). Intentions and motor representations: The interface challenge. *Review of Philosophy and Psychology*, *8*(2), 317-336.

Nanay, B. (2013). *Between Perception and Action*. Oxford University Press.

Montefiore, ACRG & Noble. D. (Eds). (1989). *Goals, No Goals and Own Goals*. Unwin Hyman. Republished by Routledge, 2021.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science*, *308*(5719), 255-258.

Pacherie, E. (2006). Towards a dynamic theory of intentions. *Does consciousness cause behavior*, 145-167.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*(1), 179-217.

Peacocke, C. (2014). *The mirror of the world: subjects, consciousness, and self consciousness*. Oxford University Press.

Perner, J., & Roessler, J. (2010). Teleology and causal understanding in childrens' theory of mind. *Causing human action: New perspectives on the causal theory of action*, 199-228.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.

Sinigaglia, C. (2008). Mirror neurons: This is the question. *Journal of Consciousness Studies*, *15*(10-1), 70-92.

Sinigaglia, C., & Butterfill, S. (2015). Motor representation in goal ascription. *Conceptual and Interactive Embodiment. Foundations of Embodied Cognition*, *2*, 149-164.

Smortchkova, J. (2018). Seeing goal-directedness: a case for social perception. *The British Journal for the Philosophy of Science*.

Spaulding, S. (2013). Mirror neurons and social cognition. *Mind & Language*, *28*(2), 233-257.

Stich, S., & Nichols, S. (1992). Folk psychology: Simulation or tacit theory?. *Mind & Language*, *7*(1-2), 35-71.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard university press.

Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford University Press, USA.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states–the question is which ones and to what extent. *Trends in cognitive sciences*, *7*(4), 153-156.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). In search of the uniquely human. *Behavioral and brain sciences*, *28*(5), 721-727.

Whiten, A. (2013). Humans are not alone in computing how others see the world. *Animal Behaviour*, *86*(2), 213-221.

Wilkes, K. (1989). *Explanation – How Not to Miss the Point*. In Montefiore, ACRG & Noble, Denis (Eds), 194-211.