The Myth of Mindreading

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

Friday, 11th February 2022

Contents

1. The Question

For a process to track an attribute is for how the process unfolds to nonaccidentally depend, perhaps within limits, on the presence or absence of the attribute.

The ability to track a variety of instrumental actions and mental states is widespread not only in human adults but also infants and nonhuman animals.[1] Within limits, you can change how they respond to situations by changing some facts about someone's mental states.

This fact about tracking invites two questions.

1. Which processes are involved in tracking instrumental actions and mental states?
2. Which models are involved in tracking instrumental actions and mental states?

1.1. Processes and Limits

Although the first question is not very puzzling, it has received little sustained attention. Let me illustrate a strategy for answering it.

Consider first the abilities of infants, from around three months of age, to track the goals of instrumental actions (Sommerville et al. 2005). One hypothesis about the first question above is that this early-developing ability to track goals involves motor process (Woodward & Gerson 2014).

Support for this hypothesis comes from considerations about **limits**. First, limits on infants' abilities to track goals line up, roughly, with limits on their abilities to act (e.g. Kanakogi & Itakura 2011; Ambrosini et al. 2013). Second, intervening on infants' abilities to act—both enhancing them through training (Sommerville et al. 2008; Gerson & Woodward 2014) and impairing them through restraining (Bruderer et al. 2015)—has a corresponding effect on their abilities to track the goals of actions.

I am not suggesting that the evidence is decisive,[2] but I do think the focus on limits is fruitful.

In fact we can use limits in the same way to defend the hypothesis that some belief tracking involves motor processes. In testing this hypothesis, Jason

[1] See, for example, Kovács et al. (2010); Kano et al. (2019); Kaminski et al. (2009); Superman, 1978.

[2] Not all goal tracking in the first year of life can involve motor processes only (Butterfill 2021).

Low and Katheryn Edwards generously let me join them to follow up on some of their earlier work (Edwards & Low 2017, 2019). They had adapted a paradigm first introduced by Kovács et al. (2010) which builds on an object detection task. Simplifying,[3] Kovács et al. (2010) found that how long it takes for participants to respond to the objects' presence by pressing a key is influenced by a protagonists' task-irrelevant belief about whether it is present or absent. We wondered:

> Why do another's task-irrelevant false beliefs ever influence my reaction times?

Our conjecture was that the influence is a consequence of me tracking the other's *apparent* action possibilities. Not their actual action possibilities but the action possibilities they would have if the other's beliefs were true. Anticipating that the other could act speeds up my own action.

Since tracking another's action possibilities often involves motor processes, you can impair tracking by temporarily limiting both the other's ability to act (Costantini et al. 2011) and by temporarily limiting my ability to act (Ambrosini et al. 2012).

Our conjecture therefore generates the prediction that constraining either my own, or the other's, action possibilities will reduce or even eliminate the influence of the other's task-irrelevant false beliefs on me. And so far we found quite promising evidence for the second part of this prediction (Low et al. 2020).

We would not expect, of course, that all abilities to track mental states are explained by a single type of process. There may be variation between species or ages. And there may even be two or more processes for tracking mental states at work simultaneously in a single individual—or so the 'two systems' theory of mindreading postulates.

If there are multiple processes involved in tracking instrumental actions and mental states, then they probably also rely on different models.

1.2. Models

A model is a way that some part of the world could be. The point of specifying a model is to capture the point of view of the agent or process that is tracking mental states. How, from their point of view, does the world appear?

Or, to put this less metaphorically, the model is *the world as it would have to be for the tracking to be free of errors.*

[3] This is the 'P-A+ > P-A-' effect.

Specifying a model is a key part of providing a computational description of a mindreading process.

1.2.1. Why Not Representations?

Why ask about models instead of representations? A claim about what a mindreader represents answers, in effect, two questions simultaneously:

1. Which model characterizes this mindreading process?
2. What links this model to the mindreading process?

The second question (about links) can sometimes be answered by saying that the mindreader represents the model, or represents a theory that specifies the model. But this is not always the right answer because sometimes the model is merely implicit in constraints on how a process operates.[4]

This is why it is useful to focus on models rather than representations: doing so allows us to answer the first question without committing on how the second question will be answered.[5]

1.2.2. Why Not Theories?

To specify a model, we as theorists might use a theory or a set of equations; or we might specify the model less formally. Note that the theory or equations are distinct from the model. They are tools that we theorists use and might be entirely unknown to those who rely on the model.

1.2.3. Why Models?

The importance of models is easy to see in the case of the physical. Suppose we are interested in commonsense physical thinking. We notice that our subjects are able to track the movements of objects, but also that this ability is subject to some odd limits. For example, they are fine at tracking objects launched horizontally but struggle when objects are launched vertically. This and other limits on their tracking might move us to postulate that

[4] To illustrate, it is plausible that Spelke's Principles of Object Perception specify the model that characterizes object cognition in early infancy—but rather than being represented by infants, it seems that they characterise how a system of object indexes operates (Leslie et al. 1998; Carey & Xu 2001). Similarly, it may be that the Teleological Stance provides a model of goals from an infants' perspective even though infants do not represent any of the principles. Butterfill (2020) discusses both examples.

[5] This is probably over-cautious. It is unlikely that many readers will care at this stage. I should probably frame things in terms of representation (more familiar)?

their tracking involves an impetus model of the physical.[6]

It is important to notice that you can track things without having a very accurate model of those things. An impetus model of the physical has all kinds of limits but remains useful in a range of everyday situations. In some cases you can even track things without having any model of them at all. For example, you might track toxicity by with a model of the world which involves only odors; or you might track what others can perceive with a model of the world which involves only lines of sight.

1.3. The Question

In the past I focussed on the problem of identifying the models that are involved in the most basic forms of mindreading, those that are common to several species and occur early in development.

I took for granted that we are all acquainted with the models that are involved in the most sophisticated forms of mindreading.

Today I want to argue that this was a mistake. We face as significant problem in identifying the models involved in the most sophisticated forms of mindreading. And failure to recognize this problem is impairing even the most prominent recent debates about mindreading (see *We Lack a Shared Understanding* (section §2)). But it is a problem that we can work around, and I will attempt to outline how we can do so (in *But One We Can Work Around* (section §4)).

2. We Lack a Shared Understanding

The overall question for this talk is,

> Which models of instrumental action and mental states are involved in the most sophisticated forms of everyday mindreading?

I aim, first, to show that this question is a problem. That is the aim of this section. (The following sections are about why it matters and how to work around the problem.)

Here is a partial answer that I think almost all researchers would agree on (tho it would further my aims were there substantial disagreement):

[6] See Kozhevnikov & Hegarty (2001) and Hubbard (2013), for example. White (2012) offers an opposing view.

> They are models which involve intentional actions and mental states like belief, knowledge, desire, intention, anger and joy.

Of course, this is only a partial answer. Accepting it means that we need to say, further, what these states are. So we should ask,

> What anchors our understanding, as researchers, of intentional action, belief, knowledge and the rest?

Here I think there are three main options, none adequate. One is to invoke our own everyday expertise as mindreaders. Another is to involve philosophers' attempts to characterise these mental states.[7] And the third is to rely on attempts to operationalize mindreading.

In this section, I am going to explore these options with the aim of showing that none provides the basis for a shared understanding of what we're talking about when, as researchers, we are talking about knowledge, desire, intention, anger, joy and the rest.

In fact we lack any such shared understanding.

2.1. Option 1: The Researcher's Personal Expertise

As well as being researchers, you and I also live ordinary lives and in these ordinary lives we have gained much expertise as mindreaders. Could this expertise be what anchors our understanding, as researchers, of belief, knowledge and the rest?

This question almost answers itself. The problem is not simply that our expertise may differ in important ways, perhaps because we are at different points on the autistic spectrum or perhaps because of cultural differences between us (see, for example, Dixson et al. 2018). This is a problem, of course. But there is a deeper problem.

This everyday expertise we both have does not enable us to know what terms like 'knowledge' and 'belief' pick out. These words may not pick any one thing out—or there may be nothing at all that they pick out (compare Fiske 2020 on emotion: this would be an instance of what he calls the lexical fallacy).

It's possible to be blind to this problem because of a temptation to suppose that the workings of your own mind and the reasons for your own actions are somehow transparent to you.

[7] Am guilty of explicitly adopting this second option.

2.1.1. Myths about Folk Psychology

Consider Lewis (1972). He postulates that there are a set of platitudes concerning mental states which are common knowledge among us all. He also claims that if we assembled these platitudes, we could use them to define mental state terms like 'intention' and 'knowledge'.

If this were true it would mean that we can, after all, rely on our everyday expertise as mindreaders to anchor our understanding, as researchers, of knowledge, intention and the rest. But is it true?

To illustrate how his view works, Lewis imagines that some important platitudes have this form:

> 'When someone is in so-and-so combination of mental states and receives sensory stimuli of so-and-so kind, he tends with so-and-so probability to be caused thereby to go into so-and-so mental states and produce so-and-so motor responses.' (Lewis 1972, p. 256)

But what are these platitudes that are supposed to be common knowledge? Heider (1958, p. 12) offered what is probably still, more than half a century later, the most sustained, carefully developed attempt to 'make explicit the system of concepts that underlies interpersonal behavior'.[8] There isn't much in Heider's work that looks useful for defining 'intention' or 'knowledge'.

It is also striking that not very much of Heider's construction could plausibly be regarded as common knowledge among ordinary mindreaders. Heider relies on a mix of informal observation, imagination, guesswork as well as philosophers' ideas (Ryle and Satre, for example), my guess is that we should regard the principles he identifies not as articulating an understanding that we all share but rather as an imaginative take on possible strategies for everyday mindreading. In fact, Heider's approach is not that different from philosophers like Bratman or Alvarez.

But if Lewis were right about common knowledge of platitudes anchoring mental state terms, either Heider's work should have turned out very differently or else there should be a lot less disagreement among the philosophers. This is why I think Lewis must be wrong.

We might be able to use theories to specify models that help us characterise

[8] Heider did not share Lewis' assumption about being able to rely on common knowledge of platitudes alone. On Heider's view, 'If people were asked about these conditions they probably would not be able to make a complete list of them. Nevertheless, these assumptions are a necessary part of interpersonal relations; if we probe the events of everyday behavior, they can be brought to the surface and formulated in more precise terms' (Heider 1958, p. 60).

the expertise of ordinary mindreaders. But we are not in a position to identify those theories simply by virtue of posessing such expertise ourselves.

2.1.2. Comparison with Naive Physics

You can see that relying on each researcher's individual everyday expertise would be a nonstarter by comparison to the physical. The successful attempts to characterise folk physics do not rely on researchers' pre-theoretical understanding of notions like force and motion. Instead they anchor these terms by invoking fragments of physicists' theories.

Since we as ordinary folk do not have much in the way of common knowledge of detailed psychological theories about belief, knowledge, desire, intention and the rest, it is perhaps natural to rely on philosophers instead.

2.2. Option 2: Rely on Philosophical Accounts

What anchors our understanding, as researchers, of action, belief, knowledge and the rest? Could it be philosophical accounts of these mental states?

At first glance this may seem like a mad suggestion just because there is so much apparent disagreement among philosophers.

Take intention, for example. It is not just that they disagree on whether intentions are beliefs about the future (Velleman 1989), or belief-desire pairs (Sinhababu 2013), or something entirely distinct from both beliefs and desires (Bratman 1987). Nor is it just that some think of intentions as essentially components of plans (Bratman 1987 again) whereas others do not connect intentions with plans at all (Searle 1983). Nor is it even that there is much disagreement about how intentions relation to intentional action, to knowledge and to belief. Philosophers even disagree on whether intentions are mental states at all.[9]

There is similar radical disagreement concerning knowledge, and concerning emotions.

So yes, it would be understandable to despair of using philosophical accounts to anchor understanding just because there is such deep and widespread disagreement among the philosophers.

But there is another, deeper reason for thinking that we cannot use philosophical accounts to anchor our understanding, as researchers, of knowledge, intention and the rest.

[9] 'There is a deep opposition here between accounts that take intention to be a mental state in terms of which we can explain intentional action, and those that do not' (Setiya 2014).

Philosophers have different, mostly unarticulated aims. Some philosophers seem to be proposing new ways of thinking in the hope that we adopt them. Others appear to be attempting to make explicit principles that are implicit in a particular tradition of law or in the activities of a particular historical culture. And of course some are trying to make systematic things that seem so obviously true that we can accept them without having any reason to do so (e.g. Lewis 1969).

Further, in trusting philosophers, you do not avoid relying on individual researcher's personal expertise. Or so Nagel argues:

> 'Unless there is a special reason to think that knowledge attributions work quite differently when we are reading philosophy papers—and [there is] evidence against that sort of exceptionalism—we should expect to find that epistemic case intuitions [which are among the things that inform philosophers' views about what knowledge is] are generated by the natural capacity responsible for our everyday attributions of states of knowledge, belief and desire. This capacity has been given various labels, including 'folk psychology', 'mindreading', and 'theory of mind" (Nagel 2012, p. 510).

To be clear, let me distinguish two claims:

1. We could (mis)use philosophical accounts of minds and actions to characterise various models of mind.
2. Philosophical accounts of minds and actions anchor a shared understanding of what knowledge, belief, joy and the rest are.

I am rejecting the second claim only. (The first claim has been very good to me, and I hope to keep misusing philosophical accounts of minds and actions.)

2.3. Option 3: Rely on the Operationalization

You might object that it doesn't matter how we characterise the models of instrumental action and mental states involved in mindreading because we already have a solidly operationalized construct, Theory of Mind.

This would be a welcome objection if true. I very much favour working back from a solid operationalization to an understanding of the things operationalized. In fact I will suggest that we can do this to some extent.

But it is important to recognise that we currently have only very limited understanding of how to operationalise mindreading.

I say this for two reasons. First, we do not really know much about the structure of Theory of Mind, and different researchers use different taxonomies (Happé et al. 2017; Beaudoin et al. 2020).[10] Second, while there is some evidence that a wide range of false belief tasks appear to test for a single underlying competence (Flynn 2006, p. 650; Wellman et al. 2001), when we turn to theory of mind tasks more broadly we find that different theory of mind tasks appear to test for different things in the sense that an exploratory factor analysis fails to find that they load on a single factor (Warnell & Redcay 2019).[11]

This means that when faced with the question of what anchors our understanding, as researchers, of action, belief, knowledge and the rest it is not enough simply to point to an operationalisation. We need more.

This does not mean, of course, that operationalisations are irrelvant. Quite the opposite. Later I will suggest that both false belief tasks (Wellman et al. 2001; Flynn 2006) and Wellman & Liu (2004)'s theory of mind scale are useful starting points.

3. This Is a Practical Problem

My conclusion so far is that **nothing** adequately anchors our understanding, as researchers, of action, belief, knowledge and the rest. Not our everyday expertise as mindreaders, not the philosophical accounts of these states, and not the attempts to operationalize Theory of Mind.

The truth is that, as researchers, you and I probably do not have a shared understanding of what intention is, what knowledge is, or what desire is. And even if we do, there are probably plenty of researchers who neither share our understanding nor have any reason to adopt our way of thinking about it.

The overall question is,

> Which models of instrumental action and mental states are in-

[10] See Beaudoin et al. (2020, p. 15): 'The lack of theoretical structure and shared taxonomy in ToM definitions and its underlying composition impedes our ability to fully integrate ToM in a coherent and comprehensive framework linking it to various socio-cognitive abilities, a pervasive issue observed across the domain of social cognition.'

[11] It is important to be clear about why this is a problem. It is not a problem that Theory of Mind may involve a variety of different processes and models, so that no single factor will explain performance across a sufficiently diverse set of tasks. But if you want to say, independently of answering the question about models, that we have a solid operationalization of Theory of Mind, then you need statistics to show that your operationalization has some kind of internal coherence. And that is what appears to be missing.

> volved in the most sophisticated forms of everyday mindreading?

You can still say, if you like, that the most sophisticated forms of everyday mindreading involve models which involve mental states like belief, knowledge, desire, anger and joy. But this doesn't get you very far because we do not know what these mental states are.

This is a practical problem. Let me illustrate.

I will start by talking you through a practical problem that I have encountered when trying to do some research on the development of action understanding. It has troubled me for several years, and I hope you will be able to emphathize.

3.1. Illustration 1: Intention

3.1.1. Deceptive Intentions

According to Scott et al,

> 'infants in the 2nd year of life can understand deceptive intentions' (Scott et al. 2015, p. 50)

Here my problem is very simple and quite modest. I am unsure whether intention matters in this context or whether they would be just as happy to say 'actions influenced by some kind of decepitive motive', leaving aside claims about whether the particular mental states are intentions, desires or something else.

3.1.2. Residing Within

Woodward suggests that:

> 'infants understand intentions as existing independently of particular concrete actions and as residing within the individual. Each of these [...] is part of what it means to understand intention in psychological terms.' (Woodward 2009, p. 55)

When I read this I am first struck by 'residing within the individual'. It's so much the metaphor of a mental state having a residence (which suppose is supposed to be just a flourish); it is the idea that a mental state might have a location.

I guess there is room for disagreement on this, but personally I find it strange to think that mental state have locations. Also none of the evidence she cites bears on this as far as I can tell. For this reason, I am tempted to think that

what Woodward needs to say here is just about intentions having subjects. Some intentions are mine and others are yours; and our intentions may differ.

Similarly, 'existing independently of particular concrete actions' triggers some metaphysical concerns about whether mental states might be like events in being individuated by their causal relations (Davidson 1969). But I will spare you that.

I just do not understand what Woodward means when she writes that 'infants understand intentions'.

3.1.3. Unfulfilled Intentions

Moses (2001, p. 74) defends the claim that 'a child's concept of intention could not fully emerge before the concept of belief.'[12] He holds this for the reason that:

> 'an unfulfilled intention must be accompanied by at least one false belief.' (Moses 2001, p. 74)

And that claim is in turn justified on the grounds that:

> 'Part of what distinguishes intentions from other motivational states, such as desires, is that intentions must be consistent with beliefs.' (Moses 2001, p. 74)

Using these considerations to interpret a body of evidence, he arrives at the view that:

> 'children of age 3 and younger may not yet have differentiated their concept of intention from their concept of desire—that at this point in development they lack an understanding of the epistemic factors (and, possibly, of the causal factors) that distinguish intention from desire.' (Moses 2001, p. 78)

I am not sure what to make of this. Theoretically, I do not see why an intention could not go unfulfilled even though I lack any relevant false beliefs but just because I am very unlucky. But perhaps this is irrelevant and we should take Moses' claim as specifying the notion of intention he has in mind. My problem, then, is not just whether Moses is right; I am unsure what would count as a terminological dispute about different notions of intention and what would count as substantive disagreement.

[12] This claim is also defended by Davidson on different grounds.

3.1.4. How do Woodward and Moses and Scott et al relate?

Things go even worse for me when I attempt to relate Moses to Woodward. As I understand Moses, his position is incompatible with Woodward's in that he would not accept Woodward's interpretation of the evidence from infants' abilities concerning actions (because none of that evidence concerns whether infants are sensitive to how beliefs constrain intentions). But why are these positions incomaptible? I can see at least three posibilities:

1. Woodward and Moses are working with different notions of intention. Despite using the same word, they are talking about different mental states. (And these mental states may even belong in incommensurable models of minds and actions.) So the disagreement is merely terminological.

2. Woodward and Moses hold incompatible views about a single notion of intention. At most one of them is right. The other has based their interpretation of the evidence on an error about the features of this mental state.

3. Woodward and Moses hold compatible views about a single notion of intention but disagree on what is required to 'understand' intention or to 'differentite their concept of intention'.

I can find no good way to decide between these three possibilities.

I encounter the same difficulty if I attempt to relate Woodward to Scott et al. Woodward seems to have a higher standard than Scott et al for postulating that infants understand intentions. Again, I am unsure whether this is because they are saying compatible things about different notions or intention, whether they are saying substantially different things about a single notion of intention, or whether they have different views on what it takes to understand something.

3.2. Illustration 2: Knowledge

According to the 'knowledge-first' hypothesis:

> 'Rather than representing what others know by first representing what they believe, people may have a separate set of processes that give rise to some comparatively simple representation of what others know.' (Phillips et al. 2020; see also Nagel 2013)

An immediate difficulty in understanding this hypothesis is that we need a shared understanding of what knowledge is.

Recognizing the difficulty, Phillips et al. (2020) propose to rely on four 'signature features that are specific to knowledge':

> (i) it is factive
>
> (ii) it is not just true belief
>
> (iii) it allows for egocentric ignorance
>
> (iv) it is not modality-specific.' (Phillips et al. 2020)

A problem several commentators note is that these four features are not in fact specific to knowledge. It is easy to identify more than two mental states with these features.

This matters because the evidence that Phillips et al. (2020) and Nagel (2013) rely on concerns observations of mindreaders' abilities to track knowledge.

Following Starmans' commentary on Phillips et al. (2020), there is a dilemma:

> On some notions of knowledge, the evidence could not support the hypothesis.
>
> On other notions (e.g. encountering), the evidence might support the hypothesis but the hypotheses is trivial.

The lack of a shared understanding of knowledge prevents us from evaluating the 'knowledge-first' hypothesis that Phillips et al. (2020).[13]

4. But One We Can Work Around

We can divide the problem of identifying models of minds and actions into two: first, a characterisation of mental states generally; and, second, a characterisation of what distinguishes different attitudes link knowledge, intention, surpise and the rest.

It turns out that the two parts of the problem are to a significant degree independent of each other.

4.1. Part I: Mental States (Perner's Strategy)

Perner starts with a theory of mental states.

[13] Because Nagel offers a different way of characterising knowledge, her view does not face exactly this problem. Instead the problem facing her view is that the evidence concerning abilities to track knowledge could not support the hypothesis. This is because there are closely related hypotheses involving notions weaker than knowledge (such as encountering, facts or not being ignorant) which are equally well supported by the evidence.

> 'representation involves a representational medium that stands in a representing relation to its representational content.' (Perner 1991, p. 40)

Mental states are understood as a relation to a thing. But there are two distinct ways of understanding mental states corresponding to two different kinds of thing they can be understood as relations to.

Option 1: The thing can be a situation, that is an aspect of the world.

Option 2: The thing can be a representation of a situation.

Option 1 is simpler but also more limited. For on Option 1, there is no way to understand the possibility of *misrepresentation*, that is, a mental state which is supposed 'to describe the real situation (referent) and yet (mis)describes it as a quite different situation (sense)' (Perner 1991, p. 92).

So why bother with Option 2 at all? Actually Perner's view is that in everyday mindreading we rarely do bother with Option 2.[14] But there are some limits on Option 1. In particular, understanding actions based on false beliefs requires Option 2.[15]

4.1.1. Perner's Paradox

The following four claims cannot all be true:

1. Ancient philosophers were deeply puzzled about the possibility of speaking and thinking falsely.
2. Ancient philosophers could have passed false belief tasks.
3. To pass a false belief tasks is to understand a case of misrepresentation.
4. 'Explicit understanding of representation (mentally modeling the representational relationship = metarepresentation) [...] is necessary for understanding cases of misrepresentation.'

This motivates considering alternatives to Perner's theory. In particular, what would happen if we rejected either (3) or (4)?

---

[14] See Perner (1991, p. 120): 'our common sense is capable of taking a representational view of the mind but that, unless really necessary, it tries to get by without it.'

[15] Perner (1991, p. 178): 'with the ability to interpret certain thinking activities as mental representation the child gains new insight into aspects of mental functioning that are nearly impossible to comprehend without a representational theory. One such case is mistaken action, that is, action based on a misconception of the world or false belief.'

4.1.2. Davidson's Measurement-Theoretic Alternative

According to Davidson:

> 'Beliefs are true or false, but they represent nothing.' (Davidson 2001, p. 46)[16]

On Davidson's view, the sentences (or, better, utterances) we use to distinguish between different things someone might intend, know or believe function a bit like the numbers we use to distinguish temperatures.

Just as numbers play no physical role, so the sentences play no psychological role. Nor do either the numbers or the sentences have counterparts that play a psychological role.

This is a measurement-theoretic, non-representational theory of the nature of mental states. (Matthews 1994, 2007 develops the idea in detail).

4.1.3. How Do Mindreaders Model Mental States?

In philosophy, the focus is sometimes on how mental states actually are. That is not our concern.

We are concerned with how mental states are modeled in mindreading. Perner's (Fodor-esque) proposal provides one option, Davidson's proposal provides an alternative option. Each option can be used to generate a hypothesis about a particular mindreading ability. Because the hypotheses generate different predictions, they are testable.

It is possible that both models are used by mindreaders at different times. Perhaps different mindreading abilities involve different models.

4.2. Part II: Attitudes

Decision theory provides a way of characterising instrumental action as a consequence of two attitudes, subjective probabilities and preferences Jeffrey (1983).

We also know from the history of decision theory that it is possible to construct models that are less sophisticated. For example, there is a model which uses objective rather than subjective preferences (that is, there is just one preference ranking that applies in all cases regardless of which subject is the agent of the action).

[16] See also Davidson (2001, p. 184): 'we ought also to question the popular assumption that sentences, or their spoken tokens, or sentence-like entities, or configurations in our brains can properly be called 'representations', since there is nothing for them to represent.'

It is possible to map some of the tasks from the Theory of Mind Scale (Wellman & Liu 2004) on to these more and less sophisticated models. This enables us to use decision-theoretic notions to characterise which models are involved in mindreading.

The advantage is that we do have a shared understanding of subjective probabilities and preferences. After all, these are characterised by the theory. The limit is that few aspects of mindreading can be characterised in this way. These limits are quickly reached even within the Theory of Mind Scale (Wellman & Liu 2004): there is no way to capture what 'Knowledge-Ignorance' is measuring, for instance.

Other features that we would like a theory of mindreading to incorporate are also missing from decision theory. For example, we would like to know to what extent mindreaders are sensitive to the distinction between strength of justification and strength of confidence. Or how mindreaders model situations involving temporal constraints among actions, as when future action possibilities depend on how an agent acts now.

How could we overcome this limit? Useful formal models are probably too much to hope for. Attempts to model notions of knowledge that are relevant to predicting or explaining action face formidable problems (see, for example, (Stalnaker 1999, Chapters 13–14) on the problem of logical omniscience).

Instead we can characterise aspects of mindreading by identifying limits of the decision theoretic model. In the talk, this is illustrated by situations in which adopting shorter or longer temporal intervals in framing the available actions influences which action will be performed (or which action we would predict if deriving predictions using a decision-theoretic model of minds and actions). This limit of decision theory corresponds to one aspect of mindreading competence that sometimes is associated with the word 'intention' (for example, by Bratman 1987).

4.3. Conclusion

It is possible to characterise even sophisticated forms of mindreading without assuming what we do not have, namely a shared understanding of notions like knowledge, intention, surprise, anger and the rest.

As researchers we do not need a shared understanding of these notions. There are better alternatives to casting theories about mindreading in terms like 'knowledge', 'intention' or 'surprise'.

No research succeeds by unreflectively using the language of the targets of explanation in characterising physical cognition, colour cognition, or any

other cognitive domain. Except mindreading. But that is something that we could change.

Glossary

computational description A computational description of a system or ability specifies what the thing is for and how it achieves this. Marr (1982) distinguishes the computational description of a system from representations and algorithms and its hardware implementation. 5, 20

instrumental action An action is *instrumental* if it happens in order to bring about an outcome, as when you press a lever in order to obtain food. (In this case, obtaining food is the outcome, lever pressing is the action, and the action is instrumental because it occurs in order to bring it about that you obtain food.) You may encounter variations on this definition of *instrumental* in the literature. For instance, Dickinson (2016, p. 177) characterises instrumental actions differently: in place of the teleological 'in order to bring about an outcome', he stipulates that an instrumental action is one that is 'controlled by the contingency between' the action and an outcome. And de Wit & Dickinson (2009, p. 464) stipulate that 'instrumental actions are *learned*'. 3, 6, 11

lexical fallacy 'the lexical fallacy consists of reifying a vernacular lexeme as a psychological entity' (Fiske 2020, p. 3). 7

model A model is a way some part or aspect of the world could be. 3, 4, 6, 11

motor process A process featuring motor representations. 3

motor representation The kind of representation characteristically involved in preparing, performing and monitoring sequences of small-scale actions such as grasping, transporting and placing an object. They represent actual, possible, imagined or observed actions and their effects. 19

Principles of Object Perception These are thought to include no action at a distance, rigidity, boundedness and cohesion. 5

representations and algorithms To specify the representations and algorithms involved in a system is to specify how the inputs and outputs are represented and how the transformation from input to output is

accomplished. Marr (1982) distinguishes the representations and algorithms from the computational description of a system and its hardware implementation. 19

Teleological Stance To adopt the Teleological Stance is to exploit certain principles concerning the optimality of goal-directed actions in tracking goals (Csibra & Gergely 1998). 5

track For a process to *track* an attribute or thing is for the presence or absence of the attribute or thing to make a difference to how the process unfolds, where this is not an accident. (And for a system or device to track an attribute is for some process in that system or device to track it.)

Tracking an attribute or thing is contrasted with *computing* it. Unlike tracking, computing typically requires that the attribute be represented. 3, 6, 15

References

Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking Ahead: Anticipatory Gaze and Motor Ability in Infancy. *PLoS ONE*, *8*(7), e67916.

Ambrosini, E., Sinigaglia, C., & Costantini, M. (2012). Tie my hands, tie my eyes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 263–266.

Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic Review and Inventory of Theory of Mind Measures for Young Children. *Frontiers in Psychology*, *10*.

Bratman, M. E. (1987). *Intentions, Plans, and Practical Reasoning*. Cambridge, MA: Harvard University Press.

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, *112*(44), 13531–13536.

Butterfill, S. A. (2020). *The Developing Mind: A Philosophical Introduction*. London: Routledge.

Butterfill, S. A. (2021). Goals and targets: A developmental puzzle about sensitivity to others' actions. *Synthese*, *198*(1), 3969–3990.

Carey, S. & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition*, *80*, 179–213.

Costantini, M., Committeri, G., & Sinigaglia, C. (2011). Ready both to your and to my hands: Mapping the action space of others. *PLoS ONE*, *6*(4), e17923.

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

Davidson, D. (1969). The individuation of events. In *Essays on Actions and Events* (pp. 163–180). Oxford: Oxford University Press.

Davidson, D. (2001). *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.

de Wit, S. & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal–human translational models. *Psychological Research PRPF*, *73*(4), 463–476.

Dickinson, A. (2016). Instrumental conditioning revisited: Updating dual-process theory. In J. B. Trobalon & V. D. Chamizo (Eds.), *Associative learning and cognition*, volume 51 (pp. 177–195). Edicions Universitat Barcelona.

Dixson, H. G. W., Komugabe-Dixson, A. F., Dixson, B. J., & Low, J. (2018). Scaling Theory of Mind in a Small-Scale Society: A Case Study From Vanuatu. *Child Development*, *89*(6), 2157–2175.

Edwards, K. & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition*, *160*, 1–16.

Edwards, K. & Low, J. (2019). Level 2 perspective-taking distinguishes automatic and non-automatic belief-tracking. *Cognition*, *193*, 104017.

Fiske, A. P. (2020). The lexical fallacy in emotion research: Mistaking vernacular words for psychological entities. *Psychological review*, *127*(1), 95–113.

Flynn, E. (2006). A microgenetic investigation of stability and continuity in theory of mind development. *British Journal of Developmental Psychology*, *24*(3), 631–654.

Gerson, S. A. & Woodward, A. L. (2014). Learning From Their Own Actions: The Unique Effect of Producing Actions on Infants' Action Understanding. *Child Development*, *85*(1), 264–277.

Happé, F., Cook, J. L., & Bird, G. (2017). The Structure of Social Cognition: In(ter)dependence of Sociocognitive Processes. *Annual Review of Psychology*, *68*(1), 243–267.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. Hillsdale, N.J: Lawrence Erlbaum Associates.

Hubbard, T. L. (2013). Launching, Entraining, and Representational Momentum: Evidence Consistent with an Impetus Heuristic in Perception of Causality. *Axiomathes*, *23*(4), 633–643.

Jeffrey, R. C. (1983). *The Logic of Decision, second edition*. Chicago: University of Chicago Press.

Kaminski, J., Bräuer, J., Call, J., & Tomasello, M. (2009). Domestic dogs are sensitive to a human's perspective. *Behaviour*, *146*(7), 979–998.

Kanakogi, Y. & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*, 341.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences*, *116*(42), 20904–20909.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830 –1834.

Kozhevnikov, M. & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, *8*(3), 439–453.

Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing 'what' and 'where' systems. *Trends in Cognitive Sciences*, *2*(1).

Lewis, D. K. (1969). *Convention : a philosophical study*. Cambridge, MA.: Harvard University Press.

Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, *50*(3), 249–258.

Low, J., Edwards, K., & Butterfill, S. A. (2020). Visibly constraining an agent modulates observers' automatic false-belief tracking. *Scientific Reports*, *10*(1), 11311.

Marr, D. (1982). *Vision : a computational investigation into the human representation and processing of visual information.* San Francisco: W.H. Freeman.

Matthews, R. (2007). *The measure of mind: propositional attitudes and their attribution.* Oxford: Oxford University Press.

Matthews, R. J. (1994). The measure of mind. *Mind*, *103*(410), 131–146.

Moses, L. J. (2001). Some Thoughts on Ascribing Complex Intentional Concepts to Young Children. In B. Malle, L. J. Moses, & D. Baldwin (Eds.), *Intentions and Intentionality: Foundations of Social Cognition* (pp. 69–83). MIT Press.

Nagel, J. (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research*, *85*(3), 495–527.

Nagel, J. (2013). Knowledge as a mental state. *Oxford studies in epistemology*, *4*, 273.

Perner, J. (1991). *Understanding the Representational Mind.* Cambridge, Massachusetts: MIT press.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., & Knobe, J. (2020). Knowledge before Belief. *Behavioral and Brain Sciences*, *X*, 1–37.

Scott, R. M., Richman, J. C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, *82*, 32–56.

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

Setiya, K. (2014). Intention. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.). Metaphysics Research Lab, Stanford University.

Sinhababu, N. (2013). The Desire-Belief Account of Intention Explains Everything. *Noûs*, *47*(4), 680–696.

Sommerville, J. A., Hildebrand, E. A., & Crane, C. C. (2008). Experience matters: The impact of doing versus watching on infants' subsequent perception of tool-use events. *Developmental Psychology*, *44*(5), 1249–1256.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–B11.

Stalnaker, R. (1999). *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford Cognitive Science Series. Oxford: Oxford University Press.

Velleman, D. (1989). *Practical Reflection*. Princeton: Princeton University Press.

Warnell, K. R. & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, *191*, 103997.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, *72*(3), 655–684.

Wellman, H. & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541.

White, P. A. (2012). The impetus theory in judgments about object motion: A new perspective. *Psychonomic Bulletin & Review*, *19*(6), 1007–1028.

Woodward, A. L. (2009). Infants' Grasp of Others' Intentions. *Current Directions in Psychological Science*, *18*(1), 53–57.

Woodward, A. L. & Gerson, S. A. (2014). Mirroring and the development of action understanding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1644), 20130181.