Ali Boyle, University of Cambridge
Draft – comments welcome

Two More Logical Problems for Animal Mindreading[1]

Abstract: Mindreading is the ability to ascribe mental states to others. Research into whether any animals are mindreaders faces a 'logical problem', according to which the methods used for detecting mindreading cannot detect it, in principle. Existing solutions to this problem succeed relative to a permissive 'behavioural equivalence' conception of mindreading. But I show that at least two more demanding conceptions of mindreading are prevalent in the literature. These generate two more logical problems, requiring us to empirically discriminate stronger and weaker conceptions of mindreading. I argue, against some indications to the contrary, that these problems can be empirically solved.

0. Introduction.

Mindreading is the ability to ascribe mental states to others, in order to predict, manipulate or explain their behaviour. The animal mindreading research programme aims to determine whether any animals have this ability. The 'logical problem' for animal mindreading claims that the methods used to detect mindreading in animals cannot detect it, as a matter of principle. These methods investigate whether subjects are able to predict a target's behaviour on the basis of the target's mental states.[2] According to the logical problem, the subject could predict behaviour in these situations by 'behaviour reading' – that is, predicting behaviour solely on the basis of observable features of the situation. 'Experience projection' tasks aim to solve this problem, by exploiting the thought that mindreaders ought to recognise that the same mental state may be present in superficially different situations – and so predict similar behaviour in these situations. But, it has been argued, these tasks can provide evidence of mindreading only if grouping situations into 'behavioural equivalence classes' is sufficient for mindreading.

In this paper, I show that there are at least two more restrictive conceptions of mindreading in the literature, endorsed by both advocates and opponents of the logical problem. On the 'causal view' of mindreading, mindreaders must represent the unobservable states which intervene between observable situations and behaviour. On the still more restrictive 'conceptual view' of mindreading, they must represent these unobservables as

[1] Thanks to …

[2] Throughout this paper, I use 'target' to refer to the individual whose mind is being read, and 'subject' to refer to the individual who is (potentially) doing the mindreading.

mental states, using mental state concepts. I argue that each of these conceptions of mindreading generates its own logical problem.

Solving these problems requires discriminating between more and less restrictive conceptions of mindreading, as well as between mindreading and behaviour reading. Making these fine-grained discriminations seems worthwhile, insofar as we are interested in what animals know about other minds – whether, for instance, they have the concept SEEING. But since the problem for causal mindreading appears intractable, and it has been argued that mental state concepts are not behaviourally discriminable from non-mental ones, it is tempting to think that these are not, in fact, empirical problems.

Against this, I argue that both of these logical problems can be solved empirically. Because conceptual mindreaders are a species of causal mindreaders, solving the logical problem for conceptual mindreading would thereby solve the problem for the causal view. I show that the logical problem for conceptual mindreading can be solved, since possession of mental state concepts confers two distinctive features on conceptual mindreaders. First, conceptual mindreaders represent mental states as intentional – as attitudes directed upon objects. Second, conceptual mindreaders appreciate that mental states bring about behaviour through their role in practical reasoning. These features, I argue, can be empirically detected.

I begin in §1 by describing the logical problem, and the 'experience projection' tasks offered in response. In §2, I argue that the experience projection response succeeds relative to a very permissive conception of mindreading – but that two more restrictive conceptions of mindreading are routinely endorsed, each of which generates its own version of the logical problem. In §3, I offer a solution to these problems. §4 concludes.

1. Background: The Logical Problem

The logical problem is so called because it identifies a problem with the 'logic' of the standard experimental approach to mindreading (Penn & Povinelli, 2007; Povinelli & Vonk, 2004). As Marta Halina (2015) explains, this approach involves applying Mill's methods of agreement and difference. Here, the method of difference involves manipulating a target individual's mental state, whilst holding background conditions as fixed as possible, and determining whether a test subject's behaviour is sensitive to the changing mental state. The method of agreement involves holding the target's mental state fixed whilst background conditions are

manipulated, to determine whether the subject is sensitive to the target's mental state across a range of background conditions. In one representative study a dominant and subordinate chimpanzee compete over two food items (Hare, Call, Agnetta, & Tomasello, 2000). The dominant's visual access to one piece of food is manipulated using an opaque barrier. Experimenters observe whether the subordinate preferentially targets food which the dominant cannot see. Various control conditions are used, to determine whether the subordinate preferentially targets unseen food across a range of observable conditions.

On the basis of positive results in this and similar studies, those I'll call 'optimists' propose there is good evidence that chimps represent others' states of seeing. 'Sceptics' claim that this approach fails to distinguish the mindreading hypothesis from the deflationary 'behaviour reading hypothesis' – according to which animals merely exploit knowledge of behavioural regularities. In the study above, chimps might know that when there is an opaque barrier between a dominant and some food, it is safe to approach the food.

This problem arises because mindreaders are not literally telepathic (Halina, 2015). They predict behaviour on the basis of mental states – but can attribute mental states only on the basis of observable features of the situation. Successful mindreading therefore always involves exploiting the very regularities a behaviour reader exploits. This makes it possible to generate what Robert Lurz (2011) calls a 'complementary behaviour reading hypothesis' for every mindreading scenario. The complementary behaviour reading hypothesis says that, rather than predicting behaviour on the basis of a mental state, the subject predicts behaviour on the basis of whatever collection of observable cues she would, according to the mindreading hypothesis, have used as the basis for ascribing a mental state.

In response to this problem, an 'experience projection' methodology has been proposed for distinguishing mindreading and behaviour reading. In experience projection tasks, a subject is supposed to attribute a mental state by 'projecting' its own experience of the target's situation. For instance, in the 'goggles test', subjects are familiarised with two sets of goggles which look identical except that one is red-rimmed and the other green-rimmed. The red-rimmed pair can be seen through, and the green-rimmed pair cannot. The subjects then interact with a target individual wearing one of these pairs of goggles. Success requires making the appropriate predictions about how the target will behave – as though she can see, if wearing the red goggles, and as though she cannot, if wearing the green (Heyes, 1998). Both chimpanzees (Karg, Schmelz, Call, & Tomasello, 2015) and ravens (Bugnyar, Reber, & Buckner, 2016) have passed variants on the goggles test.

It has been argued that the goggles test and its variants fail as solutions to the logical problem. The goggles test, in particular, could be solved by a behaviour reader recognising that the green goggles are an opaque barrier and that it is safe to approach food which is behind opaque barriers relative to a dominant individual (Lurz, 2011). The same problem arises for any variant of the goggles test which varies the target's visual state by using barriers to manipulate the target's line of gaze. To address this, Lurz (2018) has argued that a valid test for mindreading must meet four conditions. Subjects are able to learn that a particular observable cue indicates the presence of a particular mental state (the 'learning condition') – but they must have no prior reason to think that this cue is associated with a particular behaviour ('novel cue condition'). They should have prior knowledge of what behaviour tends to follow that mental state (the 'prior knowledge condition'). If these three conditions are met, the subjects should – on the mindreading hypothesis – accurately predict behaviour in the presence of the novel cue. As long as no other observable cue is present in the test condition which might enable them to predict the relevant behaviour (the 'no confounding cue condition'), Lurz claims this will be a valid test for mindreading.

As an example, Lurz and colleagues (Lurz, Krachun, Mahovetz, Wilson, & Hopkins, 2018) recently implemented the following experience projection task. Chimpanzees were familiarised with mirrors – from which they learned that the mirror enabled them to see what was behind them (the learning condition). They had no prior experience interacting with agents facing mirrors (the novel cue condition). But they had experience begging from humans who could see them (the prior knowledge condition). They were given the opportunity to beg from humans in a range of conditions, including one in which the human was facing away from them, but toward a mirror which afforded a view of the chimpanzee. Subjects displayed a preference for begging from mirror-facing humans over humans who were facing away from them without a mirror. There was no other observable cue beside the mirror present in the mirror test but not the control tests which they might have used to make different predictions about the human's behaviour (the no confounding cue condition). The experimenters concluded that the chimpanzees displayed this preference because they knew that the human could see them in the mirror.

However, tests meeting these conditions face another problem. Subjects can 'pass' such a test without mindreading – that is, without representing any unobservable mental state, or understanding that the target *sees* them. This is because in the learning condition, the subjects may learn something else about the novel cue which enables them to make the correct

predictions: that the novel cue affords certain opportunities for behaviour (Andrews, 2005; Buckner, 2014). On this basis, they may recognise that the novel situation forms a 'behavioural equivalence class' with another situation with which they are familiar – and predict, on this basis, that the target individual will behave as they would in this other situation (Buckner, 2014). Applied to the test just described, subjects may learn that the mirror affords, with respect to the objects behind them, the same behavioural opportunities as having direct line of gaze to those objects. This, together with the knowledge that humans with direct line of gaze to them can supply them with food and a preference for begging from humans who can supply food over those who cannot, would be enough to induce a preference for begging from the mirror-facing human over the wall-facing human.

In places, advocates of the logical problem indicate that the ability to group perceptually distinct situations into behavioural equivalence classes of this kind is sufficient for mindreading (e.g. Penn & Povinelli, 2007). If this were accepted, then experiments meeting Lurz' four conditions would solve the logical problem (Andrews, 2016; Buckner, 2014). But as I argue in the next section there are at least two conceptions of mindreading in the literature, frequently endorsed by sceptics and optimists, according to which mindreading requires more than grouping perceptually disparate situations into behavioural equivalence classes.

2. Two Conceptions of Mindreading

2.1 The Causal View

The first conception of mindreading I call the 'causal view'. On this view, grouping perceptually disparate situations into behavioural equivalence classes is insufficient for mindreading. To be a mindreader is to represent the unobservable *state* which is responsible for the behavioural equivalence of these situations – the unobservable state which is generated by the observable situations and which generates the same behaviour in each case.

A prominent defender of the causal view is Andrew Whiten, who characterises mindreading in terms of the representation of 'intervening variables'. Whilst behaviour readers, in his view, represent individual connections between observable situations and behavioural outputs (as in Figure 1), mindreaders represent these connections as being underpinned and mediated by an unobserved 'intervening variable' (as in Figure 2). On this view, mindreading amounts to 'an insight into the causal patterns' connecting inputs and outputs, 'a little theory, one might say, about what can unify those elements, *viz.*, the intervening variable' (Whiten,

1996, p. 290). This suggests that the idea of an intervening variable is that of a theoretical postulate. Representing these intervening variables requires representing their role in a theory of the causal patterns characterising the target's behaviour. So, to represent an intervening variable is to represent something as the unobservable occupant of a particular causal role.

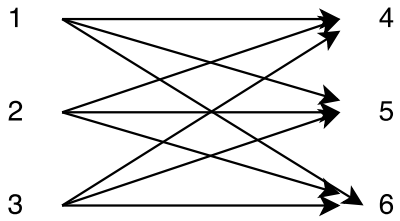

Fig. 1

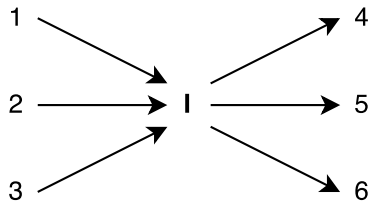

Fig. 2

In this vein, Daniel Povinelli and Jennifer Vonk (2004) contrast behaviour reading with the representation of 'unobservable mental states'. Similarly, Derek Penn and Daniel Povinelli have recently claimed that "the essence of an ToM [mindreading] […] is the ability to explicitly represent (i.e. predicate) and reason about the causal role played by a given mental state' (Penn & Povinelli, 2013). And Marta Halina (2015) characterises mindreading as involving the representation of regularities of the form 'some situation *s* will lead an agent to have the cognitive state cs, which will in turn lead that agent to produce a particular behaviour *b* (*s* → cs → *b*)'.

It is worth emphasising the distinction between the causal view of mindreading and the behavioural equivalence strategy, since the two are frequently equated (Andrews, 2016; Penn & Povinelli, 2007). They are equated because both involve introducing an additional 'inferential step' (Halina, 2015) between the representation of an observable situation and behaviour. In the behavioural equivalence strategy, the additional step involves representing that the two situations are behaviourally equivalent. In the causal mindreading strategy, it involves representing that an observable situation gives rise to a causally efficacious unobservable state in the target. But these are distinct, since representing two things as falling into a behavioural equivalence class is not to represent that there is an unobservable state, with the causal role of a mental state, responsible for their behavioural equivalence.

Despite this difference between representing behavioural equivalence classes and the unobservable states by which they are underpinned, it is difficult to see how the two abilities might be empirically distinguished in a behavioural test. Distinguishing causal mindreading

from the behavioural equivalence strategy requires telling a story about the 'unique causal work' causal mindreading is supposed to do (Povinelli & Vonk, 2004). But it is unclear what this unique causal work could possibly be. Causal mindreaders, in representing the unobservable states responsible for behaviour, thereby group situations into behavioural equivalence classes. So, causal mindreaders are a subspecies of what we might call 'behavioural equivalence mindreaders' – wherever causal mindreading is happening, behavioural equivalence classes are being formed. The difference between causal mindreading and behavioural equivalence mindreading is just that causal mindreaders represent the unobservable state responsible for the behavioural equivalence. But this is simply to represent an additional element in pattern behavioural equivalence mindreaders already represent – a pattern connecting observable situations to behaviour. Representing this additional state does not improve their ability to predict behaviour.

Given that causal mindreading always involves representing behavioural equivalence classes, and that causal mindreading appears to imbue agents with no additional mindreading powers, there remains a logical problem for causal mindreading – one which, on its face, looks intractable.

2.2 The Conceptual View

The second, still more restrictive conception of mindreading I will call the 'conceptual view' of mindreading. On this view, mindreaders do not simply represent the unobservable states causally underpinning behavioural regularities – but they represent these *as* mental states, using mental state concepts.

This conception of mindreading is evident in the way researchers typically characterise mindreading by reference to representations of attitudes like belief, perception, desire and so on. For instance, Premack and Woodruff characterise a theory of mind as one in which mental states are imputed to others, where mental states include 'all those designated by the italicised term in each of the following statements: John *believes* in ghosts; he *thinks* he has a fair chance of winning; Paul *knows* that I don't *like* roses; she is *guessing* when she says that; I *doubt* that Mary will come; Bill is only *pretending*' (1978, their italics). Whiten asks 'does the ape wonder, while looking quizzically at another individual, What does he really *want*? What does he *believe*? What are his *intentions*?' (2001, p. 199, emphasis his). And Povinelli and Vonk ask whether humans are alone in having the ability to interpret behaviour in terms of 'things like feelings,

beliefs, desires, emotions, and intentions' (2004, p. 1). These remarks indicate that mindreaders do not merely represent the unobservable causes of behaviour – they represent them as instances of belief, perception, knowledge and so on.

To crystallise the distinction between conceptual and causal mindreading, consider the following. Causal mindreaders group situations into superordinate causal classes by representing the mental state they have in common. For instance, a number of perceptually disparate situations have in common the state 'seeing food', which tends to lead to certain behaviours. Conceptual mindreaders group mental states themselves into superordinate classes. They represent the states 'seeing food' 'seeing me' and 'seeing a threat' as instances of the same kind: seeing. So, conceptual mindreaders recognise commonalities between episodes involving *distinct* mental states – even states which generate very different behaviour, and consequently occupy quite distinct causal roles. Causal mindreaders need not appreciate this. They need not recognise anything in common between the state 'seeing bananas and 'seeing a bear' – two states occupying quite different causal roles. These states could, for a causal mindreader, be represented using syntactically simple representations, employing no common representational components – the causal mindreader might simply label them 'S1' and 'S2'. For this reason, the causal view of mindreading and the conceptual view of mindreading are non-equivalent. According to both, mindreading involves representing unobservable mental states occupying certain causal roles. But according to the causal view, this is *all* that it must involve. On the conceptual view, these unobservable mental states must be represented *as* mental states, using mental concepts.

Although causal and conceptual mindreading are non-equivalent, the two are closely related. Just as causal mindreading always involves forming behavioural equivalence classes, conceptual mindreading always involves causal mindreading. On the conceptual view, mindreaders represent mental states using mental state concepts – but the states they represent are not observable, and are uniquely characterised by a causal role. They must appreciate this causal role in order to make the behavioural predictions diagnostic of mindreading. They can recognise an instance of seeing only by recognising the observable events which precede it – and on the basis of ascribing a particular instance of seeing, they are able to predict what

behaviour will follow. By extension, they must also group situations into behavioural equivalence classes.[3]

This relationship between the causal and conceptual views of mindreading reveals the existence of yet another logical problem. Let 'MC' denote a representation of a mental state which employs mental concepts, and 'C' a representation of an unobservable state occupying the relevant state's causal role. In any case where it appears that an animal represents the regularity (*s* → *MC* → *b*), the claim that it represents the regularity (*s* → *C* → *b*) is equally well supported. Representing C is simply a matter of representing an unobservable state occupying the right causal role. Anyone who represents the regularity (*s* → *MC* → *b*) represents the occupant of the relevant causal role – they simply represent it under a particular concept. Since anyone who represents (*s* → *MC* → *b*) also represents (*s* → *C* → *b*), it is not clear that any behavioural test can distinguish a conceptual mindreader from a merely causal one.

Existing methods seem primarily to be directed at determining subjects' understanding of mental states' causal roles. By manipulating a mental state like seeing food in several ways, researchers test whether the test subjects understand the variety of observable situations which generate that mental state. By looking for appropriate behaviour across conditions, they test whether the subject knows what type of behaviours the mental state generates in turn. So, notwithstanding the logical problem for causal mindreading, a pass on these tests seems at *best* evidence that they represent a state occupying the causal role of the relevant mental state. It is hard to see what support this approach can provide for the hypothesis that the animal uses mental state concepts to represent the mental state. Again, to solve this problem requires telling a story about the 'unique causal work' done by the possession of mental state concepts.

At this point, is worth making explicit the general shape of these problems, and how they are related. The logical problem arises in its original form because mindreaders are a species of behaviour readers: whenever mindreading occurs, the subject must be representing the regularities connecting observable situations to behaviour. Thus, the challenge is to show that subjects belong to the restricted category of mindreaders, rather than the broader category of behaviour readers. Each of the two stronger logical problems I have outlined in this section

[3] Lurz (2011) disputes that conceptual mindreading involves forming behavioural equivalence classes or representing unobservable causes. But it is unclear how it can fail to involve the former, whilst supporting success in experience projection tasks. His denial that it involves the latter turns on the claim that we don't take mental states to be 'literally internal'. But as Whiten writes, on the causal view, mindreaders represent 'states which others are classed as being in: it is not necessary that they are seen as "internal" to others' (1996, p. 286).

arises in precisely the same way: causal mindreaders are a species of behavioural equivalence class mindreaders, and conceptual mindreaders are a species of causal mindreaders. Thus, in each case, the challenge is to show that animals belong to the species, rather than to the genus, by articulating the 'unique causal work' species membership performs.

Because the problems are related in this way, it is possible in principle to solve them all at a single stroke by solving the logical problem for conceptual mindreading. Since conceptual mindreaders are a species of causal mindreaders and behavioural equivalence class mindreaders, to show that animals are conceptual mindreaders would be to show that they were mindreaders in these weaker senses, too. This means that, despite its apparent intractability (§3.1), the logical problem for causal mindreading can be solved, so long as it is possible to solve the problem for conceptual mindreading. However, it has been argued that conceptual mindreading is indistinguishable from behaviour reading and weaker construals of mindreading by empirical means. Kristin Andrews (2016, 2017) claims that for any mental state concept, there is a non-mental concept which is functionally equivalent. A conceptual mindreader would not differ from a user of this non-mental concept in any detectable way. If so, then the logical problem for conceptual mindreading, along with the one for causal mindreading, will be intractable – and so, perhaps, can be safely ignored.

In the next section, I argue that this is not the case. Conceptual mindreading can be empirically distinguished both from behaviour reading and weaker construals of mindreading. So, the logical problem for conceptual mindreading is an empirical one. Insofar as mindreading research concerns 'what chimpanzees know about seeing' (Povinelli, Eddy, Hobson, & Tomasello, 1996) and other mental states, it is a problem worth solving .

3. Conceptual Mindreading

Experience projection tasks attempt to solve the logical problem by exploiting the thought that mindreaders should recognise the same mental state in a range of 'perceptually disparate' situations. In this section, I argue that unifying perceptually disparate situations is only one element of the 'unique causal work' conceptual mindreading does. Possession of mental state concepts also enables conceptual mindreaders to unify *behaviourally* disparate situations, in two ways. First (§3.1), conceptual mindreaders appreciate the intentionality of mental states. That is, they represent mental states as falling into higher order attitude types, and recognise that these attitudes can be directed on different things on different occasions – with different

behaviour as a result. Consequently, they recognise that the same *type* of mental state can generate distinct behaviours. Second (§3.2), conceptual mindreaders appreciate that mental states generate behaviour through their role in practical reasoning. So, they should anticipate that their targets will exhibit the behavioural flexibility characteristic of minded creatures.

3.1 Intentionality

As noted in §2, conceptual mindreaders group mental states into higher-order categories, according to the attitude type to which they belong – as instances of seeing, hearing, knowing and so on. When they represent particular instances of these attitudes, they must represent them as directed on particular objects. Appreciating that mental states exhibit intentionality, 'the mind's direction on its objects' (Brentano, 2015), is plausibly part of what it is to possess a concept like SEEING. Moreover, representing attitudes together with the objects upon which they are directed is required for making the behavioural predictions diagnostic of mindreading – since behaviour is a function not only of the type of attitude, but also the object upon which that attitude is directed.

That conceptual mindreading constitutively involves a representation of mental states with this intentional, object-directed structure, is frequently emphasised in the literature on animal mindreading. For instance, Hutto and colleagues write that it is part of the 'standard view' of mindreading that a mindreader 'represents mental states with their intentional […] content, i.e. that which they are "about" or "directed toward"' (Hutto, Herschbach, & Southgate, 2011, p. 376). Povinelli and Eddy question whether chimpanzees understand seeing 'as a mental event' – which, they explain, involves appreciating 'how visual perception (seeing) *links organisms to the external world*' (Povinelli et al., 1996, p. v, my italics). And Lurz (2011) emphasises that our own mental state attributions 'describe relations […] that the subject bears to various external items'.

The idea that conceptual mindreaders represent mental states as intentional states provides an initial foothold for distinguishing conceptual mindreaders from causal mindreaders. Causal mindreaders, as noted in §2, can use syntactically simple representations to represent unobservable states like 'seeing food'. But to represent this state as intentional requires a representation which is compositionally structured – that is, one composed out of multiple discriminable conceptual components. When a conceptual mindreader represents that

A sees food, her representation of A's mental state factors into two conceptual components: SEEING, and FOOD.[4]

Since conceptual mindreaders' representations of mental states are compositional and systematic, one route to solving the logical problem for conceptual mindreading involves determining whether animals' mindreading behaviour exploits such compositional, systematic representations. This proposal has limited applicability when considering individual mindreading scenarios – since, in it is impossible to tell, by looking at a single representational episode, whether a subject's representation is simple or compositionally structured. But when representations are compositional, they are also said to be systematic – that is, the concepts out of which they're composed can be flexibly recombined in various ways to produce many *other* complex representations (Evans, 1982, p. 104). So, we might investigate whether animals' representations are compositionally structured by considering whether they are systematic.

As Carruthers (2009) has argued, the connection between compositionality and systematicity is an epistemic one. If the concepts we ascribe to a subject, in ascribing compositionally structured thought, do not feature in at least some other representations – then there is no reason to ascribe a structured thought at all. To put it another way, when we ascribe compositionally structured thoughts, we expect our ascriptions of conceptual competence to be 'projectable' (Bermúdez, 2003). So, if a creature represents seeing food using the concept SEEING, we should predict that they will use this concept in other situations as might reasonably be expected of someone competent with that concept. The nature of the concept SEEING should make ascriptions of seeing projectable across a great many situations. This is because competence with SEEING involves appreciating that seeing is an intentional attitude which can be directed on any visible object. So, users of this concept should be able to ascribe states like 'seeing a bear', 'seeing the sea', and so on, when presented with evidence that others are in these states.[5]

---

[4] I treat concepts as the 'building blocks' out of which thoughts are composed (Carruthers, 2009). Another approach treats compositional thoughts as drawing on 'distinct, systematically interacting representational abilities' (Camp, 2009). The difference is not significant for my purposes – those who prefer the 'abilities' approach can paraphrase.

[5] It should be noted that systematicity in this sense is distinct from domain generality – so, to claim that users of SEEING should exploit it systematically to represent seeings of any visible object is consistent with the view that there may be 'islands of social understanding' (Buckner, 2014) – that is, certain contexts in which this systematic competence with SEEING might manifest.

These considerations indicate that, on the mindreading hypothesis, there should be a certain probabilistic relationship between succeeding in one mindreading task, and succeeding in another – for certain pairs of tasks. Elliot Sober (2015) proposes a methodology for investigating such probabilistic connections. He describes a two-wing test, in which subjects face two distinct mindreading tasks drawn from a study by Alicia Melis and colleagues (Melis, Call, & Tomasello, 2006). On each trial, subjects are required to choose between taking food through an opaque or transparent tunnel, and then through a silent or noisy trapdoor, in the presence of a human competitor. The question is 'whether the two chimpanzee behaviours – choosing the opaque tunnel and choosing the silent trapdoor – are correlated, conditional on the human competitor's being present' (Sober, 2015, p. 222). Sober's claim is that the mindreading hypothesis and the behaviour reading hypothesis make different predictions. The mindreading hypothesis predicts that these should be positively correlated, because having a belief about what the competitor can see should raise the probability of having a belief about what the competitor can hear. By contrast, the behaviour reading hypothesis predicts 'screening off' – that is, probabilistic independence – because there is no reason that having a belief about getting food through the opaque tunnel should raise the probability of having a belief about getting food through the silent trapdoor.

A problem for this approach, as presented by Sober, is that there is no reason to suppose that success in the tasks described should be positively correlated in a mindreader. Sober simply states that 'a mindreader has the resources to apprehend these connections'. But it is unclear why this should be accepted. Here, it's worth highlighting that in mindreading research, questions about whether animals represent one type of attitude are treated as insulated from questions about other attitude types. The consensus among optimists is that chimpanzees represent seeing and knowing. This consensus is not undermined at all by chimpanzees' history of failing (for the most part)[6] false belief tests. This is an entirely separate issue. Similarly, it would not be undermined by the discovery that chimpanzees were incapable of mindreading with respect to hearing. So, a subject might be competent with the concept SEEING, but not with HEARING. And even if they were competent with both concepts, their responses in the two tasks might nevertheless be independent, since the two tasks exploit no common conceptual ability.

[6] See (Buttelmann et al., 2017; Krupenye, Kano, Hirata, Call, & Tomasello, 2016) for evidence that chimpanzees' are sensitive to others' false beliefs.

However, the fact that conceptual mindreaders employ compositional representations of mental states, in which mental state concepts are deployed systematically, provides the necessary support for the idea that certain mindreading tasks should be probabilistically related in a conceptual mindreader. These will be tasks which draw on a common concept – the concept SEEING, for instance. Given their possession of the concept SEEING, we can predict that a conceptual mindreader's succeeding in one visual mindreading task should raise the probability that they will succeed in another involving a distinct visual state. A conceptual mindreader passes a visual mindreading task involving the state 'seeing food' by exploiting the concept SEEING. They should consequently be able to use SEEING to represent other instances of seeing, directed on other objects, when presented with evidence of their occurrence.

Importantly, the claim here is not merely that subjects should pass other tests involving the same state of seeing (seeing food, for instance) across scenarios in which the observable cues for seeing are perceptually distinct.[7] It is that they should pass tests involving *distinct* states of seeing, directed on different objects – including novel objects, generating states of seeing with which they are not familiar. Moreover, given that behaviour is a function not only of attitude type, but also of the object upon which the attitude is directed, they should predict that the behaviour in each case will vary – that seeing an angry bear has very different behavioural consequences from seeing bananas. So, in this sense they should anticipate that the same type of mental state (seeing) can be present across *behaviourally* disparate situations. None of this is the case for merely causal mindreaders. Passing a test involving 'seeing food' does not raise the probability of passing tests involving other, non-food directed states of seeing for causal mindreaders – since causal mindreaders are not credited with systematic mental state concepts.

In this way, the idea that conceptual mindreaders possess systematic mental state concepts motivates the two-wing methodology Sober proposes – but indicates that the tasks chosen must meet certain conditions. First, in order for the conceptual mindreading hypothesis to predict probabilistic dependence, the two tasks must exploit a common mental state concept – so, for instance, both should be visual mindreading tasks, in which a competent user of SEEING could be expected to ascribe a visual state. Call this 'the common concept condition'. Second, the mental states involved in the two tasks should be directed on different objects. These objects should be sufficiently different that the states generate distinct behaviour and

[7] This thought underpins the distinct unifying hypothesis strategy, due to (Tomasello & Call, 2006).

consequently occupy distinct causal roles. This condition is required so that no weaker conception of mindreading will predict probabilistic dependence – the states occupy different causal roles and generate distinct behaviour, so only for a conceptual mindreader in possession of the concept SEEING should the two tasks be connected. Call this the 'behavioural difference condition'. Finally, the two tasks must differ sufficiently in their observable features that no common behaviour reading strategy can be recruited to solve both tasks. In particular, they should avoid the familiar objection that the behaviour reading concept DIRECT LINE OF GAZE supports success in the majority of visual mindreading tasks – since, otherwise, the behaviour reading hypothesis will also predict probabilistic dependence. To satisfy this 'perceptual difference condition', one test might manipulate the target's state using opaque barriers, as in Hare et al (2000), whilst another manipulates it using mirrors, as in Lurz et al (2018).

Andrews argues that a two-wing methodology cannot provide evidence for the use of mental state concepts, since behaviour readers can group perceptually disparate mindreading tasks under common 'behaviour reading' concepts. She offers the concept DETECTION – responsiveness to the presence or absence of a certain feature in the world. For instance, litmus is a detector, because it is responsive to acid. She claims that DETECTION can be applied to the target in both wings of Sober's task, so the hypothesis that a subject is competent with DETECTION predicts probabilistic dependence. It is safe to assume she would say the same thing about any pair of visual mindreading tasks.

In response, it should be noted that it is not sufficient for a prediction of probabilistic dependence that there is some concept which can be applied to the target in both tasks. The concept must be projectable across the two tasks, such that a competent user of the concept could be expected to exploit it in the second task, given that they applied it to the first. The concept SEEING is projectable across visible mindreading tasks, because competence with seeing involves appreciating that any visible object can be seen. DETECTION is not like this. That something is a detector for one feature does not indicate that it is a detector for any other feature – litmus detects acid, but it makes for a poor smoke alarm. To classify somebody as a detector of X, one must already know that they are responsive to X – which means that that representing that someone detects food, for instance, does not by itself raise the probability of recognising that they can 'detect' rocks. This would require some independent information about their responsiveness to rocks – information which would suffice by itself for success in a task involving rock-detection.

Andrews might respond that some other concept could be constructed to be projectable in the right way. But it could not be a concept of a behavioural tendency to respond to visible things – since the majority of visible stimuli elicit no reliable response. The concept would instead need to be something like VISUAL DETECTION, a *capacity* to respond to any visible object. It is not clear that this qualifies as a behaviour-reading concept. It picks out an unobservable state of targets, which relates them to visible things in such a way that they can respond to them, thereby playing a role in the production of their behaviour. There is a story to be told about why a concept like this might be 'non-mental', as I discuss in §3.2. But if it is not a behaviour reading concept, then even if the two-wing methodology does not rule out the use of this sort of concept, it is more powerful than Andrews accepts. It rules out behaviour reading, the behavioural equivalence strategy, and merely causal mindreading.

Given a pair of tasks meeting these conditions, the hypothesis that a subject possesses the concept SEEING predicts probabilistic dependence, but neither the behaviour reading hypothesis nor any weaker mindreading hypothesis does. Suitable tasks can be produced by varying both the object the target sees and the way the target's visual access is manipulated. One task might use an opaque barrier to manipulate a dominant individual's access to their mate being approached by a subordinate; another using a mirror to manipulate an individual's visual access to an approaching predator. The conceptual mindreading hypothesis, and no other, predicts that if the subject can predict behaviour in one task, she should be more likely to predict the correct behaviour in the other. Consequently, a two-wing task meeting these three conditions can distinguish conceptual mindreading both from behaviour reading and from weaker construals of mindreading. If the result is probabilistic dependence, this indicates that the subject has compositional representations of intentional states. On the other hand, if the result is screening off, it disconfirms the conceptual mindreading hypothesis.

3.2. Practical Reasoning

A two-wing task meeting the three conditions above could provide evidence that animals represent unobservable states with an intentional, object-directed structure. But as acknowledged in §3.1, some 'non-mental' concepts might fit this description. Of course, what constitutes a mental state – and by the same token, a mental state concept – is likely to be a matter of interpretation. It is possible that some of those working in the mindreading debate

would be happy to conclude animals were mindreaders if they could be shown to represent unobservable, object-directed states of any kind. But Tyler Burge (forthcoming) has argued that to represent mental states, one must distinguish them from the more general category of sensory states – which in his view includes non-mental states which respond to stimuli, triggering conative states and behaviour. Non-mental sensory states include states of the inner ear, by means of which we sense the direction of the ground, or states in a plant which detect the direction of light, facilitating light-sensitive growth-patterns. There does seem something odd about the thought that ascribing sensory inner-ear or plant states might qualify one as a conceptual mindreader – a possessor of mental state concepts. So, the question is what the difference between ascribing non-mental and mental intentional states might amount to.

Burge (forthcoming) proposes that to represent a mental state, one must represent an object-directed state as exhibiting one of the two 'marks of the mental' – either as conscious or as representational, in the distinctively mental sense of being apt for accuracy or inaccuracy. But there are reasons to resist this claim. Many animal mindreading researchers take it to be a live possibility that animals possess non-representational mental state concepts. That is, they think it is possible that animals are mindreaders, with concepts of perception and knowledge, but that they fail to appreciate that perceptions can be illusory and beliefs false. Instead, they treat perception and knowledge as relations to actually existing objects and situations (Call & Tomasello, 2008; Martin & Santos, 2016). Moreover, there is no suggestion in these proposals that animals compensate for their non-representational mental state concepts by representing the conscious character of these mental states. It is unclear how representing conscious character could be relevant in most mindreading scenarios – which require behaviour to be predicted as a function of the target's representational states. So, to require conceptual mindreaders to represent the marks of the mental seems to involve a departure from what mindreading researchers take themselves to be investigating.[8]

A remark from Call & Tomasello (2008) suggests an alternative strategy for distinguishing mental from non-mental states. In support of their optimist view about chimpanzee mindreading, T&C claim that chimpanzees employ a 'perception goal psychology', one component of which is understanding "how these psychological states work together to produce intentional action; that is, they understand … [that] the other acts in a certain way

[8] More generally, one need not represent the mark of X-ness to be credited with the concept X: I possessed the concept BEAR long before I had any inkling about the mark of bear-ness.

because she perceives the world in a certain way and has certain goals of how she wants the world to be.' This indicates that to be a mindreader involves appreciating the *way* in which mental states combine to produce behaviour: by contributing to practical reasoning. One consequence of the role of practical reasoning is behavioural flexibility. Correspondingly, as Call and Tomasello point out, appreciating the role of practical reasoning provides a 'more flexible way' to predict behaviour.

Although the relationship between mental states and practical reasoning is not commonly construed as a mark of the mental, practical reasoning and the behavioural flexibility to which it gives rise have been proposed as marks of mental *systems*. It is not uncommon to appeal to behavioural flexibility in order to distinguish minded from non-minded creatures. To take a well-known example, the *sphex* wasp has been said to exhibit profoundly inflexible behaviour. When the wasp returns to its burrow with a paralysed cricket, it leaves it on the threshold whilst inspecting the burrow. The story goes that if the cricket is moved away from the threshold during the inspection, the wasp will move it back to the original spot and repeat the procedure – and that this can be repeated up to forty times with the same result.[9] The inflexibility in the wasp's behaviour indicates that it is not genuinely minded: as Carruthers (2004) writes, 'if this were the full extent of the flexibility of insect behaviours, then there would be no warrant for believing that insects have minds at all'.

Although the wasp has states which respond to environmental stimuli and trigger certain behaviours, they do not do so by contributing to a process of practical reasoning. Its sensory states do not produce behaviour by providing the wasp with information relevant to the pursuit of its goals. In fact, the wasp appears to have 'no conception of the overall goal of the sequence' (Carruthers, 2004). Consequently, the wasp's behaviour does not respond appropriately to reasons: it does not take its sensory states to provide mounting evidence that the burrow is cricket-ready, or that something has been interfering with its cricket, and it is consequently not motivated to bring the sequence to a close.

Genuinely minded creatures possess a more complex cognitive architecture. Perception provides information which feeds into practical reasoning; together with some goals or desires, the result is intentional action. The behaviour thus generated is flexible – it is not simply triggered by sensory states, but is sensitive to the informational import of those states, and the way in which this information bears upon the subject's goals. This provides for

[9] This story about *sphex* is almost certainly not true – see (Keijzer, 2013) for a history. I use it for illustrative purposes.

a way in which mindreaders might distinguish mental from non-mental sensory states, without representing the marks of mentality. If a creature represents genuinely *mental* states, she should appreciate that they produce behaviour by contributing in this way to practical reasoning. As a result, she should anticipate that her targets will exhibit the flexible behaviour characteristic of genuinely minded creatures.

Of course, here one might reply that practical reasoning involves representational states – beliefs and desires – and consequently, that this construal of conceptual mindreading implies that mindreaders must ascribe representational states after all. But whilst it is true that, as a matter of fact, practical reasoning involves representational states, a creature might have a minimal grasp on practical reasoning whilst failing to appreciate this. For instance, they might take perception to generate knowledge, which figures in practical reasoning, and yet fail to appreciate that it could generate belief, a state which is apt to be true or false. They should therefore anticipate that targets will act in accordance with the knowledge perception provides – and, since the same perceptual state may provide or fail to provide knowledge on different occasions, they should anticipate behavioural flexibility.

Similarly, whilst an appreciation of practical reasoning involves some representation of a subject's goals or desires, as Eric Schwitzgebel (1999) argues, desires are representational only in a weak sense. They can be directed on non-actual situations, but they are not representational in the sense of being apt for accuracy or inaccuracy. Thus, it is possible to represent a desire without representing an accuracy-apt representation – for instance, by representing a target's functional relationship to a hypothetical situation (see Perner, 1991). So, the claim that conceptual mindreading involves an appreciation of the role of practical reasoning in behaviour does not conflict with the claim, common in animal mindreading research, that one might possess mental state concepts, yet fail to appreciate that any mental states are representational in the sense of being apt for truth and falsity.

The idea that conceptual mindreaders are sensitive to the role of practical reasoning in the production of behaviour provides another response to Andrews' (2016) claim that mental and non-mental concepts are functionally indistinguishable – since the distinction drawn above, between mental and non-mental systems, is a functional one. In mental systems, practical reasoning plays a mediating role in the production of behaviour. Conceptual mindreaders appreciate this: so, their competence with SEEING involves appreciating something about the functional role of seeing: that it produces behaviour by contributing to practical reasoning. So, its role in producing behaviour is mediated, and the behaviour

produced is flexible: the behaviour resulting from seeing X, for instance, may vary depending upon whether seeing X is epistemically significant. Moreover, given the connection between seeing and practical reasoning, to learn that someone sees X can generate predictions in the absence of specific knowledge about how they tend to respond to X, given some limited information about the epistemic and motivational significance of seeing X.

Non-mental sensory states are not hooked up to practical reasoning. Correspondingly, representing them does not involve appreciating the role of practical reasoning in the production of behaviour. This means, first, that non-mental concepts do not license predictions of behavioural flexibility, and second, that they license few behavioural predictions in the absence of knowledge about how targets tend to respond to the stimulus. Given these differences between SEEING and its non-mental analogues, the claim that a non-mental concept can be made functionally equivalent to any mental one appears unmotivated. If a concept functioned like SEEING in these ways, there would be no reason to deny it was a mental state concept.[10]

One way to investigate whether animals appreciate the role of practical reasoning would be to determine whether their predictions about a target's behaviour are sensitive to both the target's perceptual states and their goals. One issue here, though, is that it is difficult to provide subjects with information about a target's goals regarding an object without providing information about how they tend to behave when they see the object – information which could support non-mental predictive strategies. A recent study adopted a different approach, encouraging subjects to 'reverse engineer' a target's goal in attempting to open an empty box. Whilst subjects correctly interpreted the behaviour of an agent with a false belief, their failure to distinguish knowledgeable and ignorant agents indicated that their ability to infer goals in this way was, at best, limited: they could not figure out a knowledgeable agent's goal in opening a box known to be empty (Buttelmann, Buttelmann, Carpenter, Call, & Tomasello, 2017). So, whilst this study may indicate a limited grasp of practical reasoning, more research is needed. In particular, this does not speak to the question of whether subjects'

[10] Andrews (2017) indicates that there might be 'qualitative' differences, so a user of functionally equivalent non-mental concepts would be a 'chimpanzee zombie'. But since mindreading is a cognitive capacity, not typically thought to involve distinctive qualia, it is difficult to know what to make of this suggestion. There is no reason to deny that a chimpanzee zombie could be a *bona fide* conceptual mindreader.

predictions about behaviour track the target's reasons, rather than what targets with certain mental states tend to do – that is, whether subjects anticipate behavioural flexibility.

A promising avenue would be to exploit the fact that perceptual states feed into practical reasoning by providing information. In different circumstances, the very same perceptual state can provide different information, or no information at all. As such, the same combination of perceptual and goal states can generate different behaviours in different contexts. The following set-up illustrates what I have in mind. Subjects observe as two individuals compete for food, which is accessible via either of two trapdoors. Both of these trapdoors are connected to a lightbulb, such that opening either trapdoor causes the light to switch on. There is also a switch next to the light, by means of which it can be switched on. The dominant individual, A, faces one of the trapdoors and the red light, and has her back to the second trapdoor, and to a subordinate, B, who can access the second trapdoor. After being familiarised with this set-up, subjects observe the following scenarios.

1) No door opens, and nobody flicks the switch, but the light comes on.
2) B opens his trapdoor and the light switches on.
3) A third individual comes in and flicks the switch, tuning on the light.
4) A opens her trapdoor and the light switches on.

Anticipatory looking or violation of expectation measures might be used to determine the subject's predictions about A's behaviour in each case. In all four situations, A sees the light come on. But in situations 3 and 4, the cause of the light switching on is known to A. The light thus provides her with no new information, and no reason to turn around. In situations 1 and 2, the cause of the light turning on is unknown. Consequently, seeing the light provides her with new information, and a reason to turn and investigate. If the subject represents 'seeing the light' as a sensory state which triggers or tends to produce food-protection behaviour, then she should expect the subject to turn and investigate the second trapdoor in all four conditions. On the other hand, if she represents 'seeing the light' as a mental state, feeding into practical reasoning, she should predict that the subject will turn only when seeing the light provides her with new information, and a reason to turn.

Because conceptual mindreaders understand how mental states bring about behaviour through practical reasoning, conceptual mindreading can fulfil what Hayley Clatterbuck (2016) calls the 'limiting role' of theories. Theories play a limiting role by explaining observed regularities – enabling their users to anticipate certain violations of those observed regularities. Non-theorists, working with knowledge of observed regularities alone, will 'over-extend' their

predictions in these cases – predicting that the regularity will hold, when the theory says it will not. Clatterbuck puts this thought to use in discussing how to discriminate theorists, who represent unobservable variables, from non-theorists, who do not. But it also provides the means to differentiate users of distinct theories – in this case, it can be used to discriminate theorists whose unobservable variables stand in a more or less rigid relationship to behaviour. Those who represent non-mental sensory states standing in a less flexible relationship to behaviour will tend to overextend their predictions. They will predict similar behaviour in response to similar sensory states, even in cases where conceptual mindreaders would not make such a prediction. Because conceptual mindreaders understand *how* mental states bring about behaviour, their predictions will be more circumspect. In this way, conceptual mindreaders are discriminable from users of non-mental sensory state concepts.

4. Conclusion

The logical problem may have been solved for a permissive 'behavioural equivalence' conception of mindreading. But stronger conceptions of mindreading are frequently endorsed both by advocates of the logical problem and their opponents. These generate stronger logical problems, challenging us not only to discriminate mindreading from behaviour reading, but also to make fine-grained discriminations between more and less restrictive conceptions of mindreading. The first of these problems, corresponding to causal mindreading, demands evidence that animals represent the unobservable causes of behaviour, and not merely behavioural equivalence classes. The second, corresponding to conceptual mindreading, demands evidence that animals represent unobservable states *as* mental states, using mental state concepts.

The logical problem for causal mindreading looks intractable on its face. And it has been suggested the logical problem for conceptual mindreading is empirically intractable – since a behaviour reading concept can be made functionally equivalent to any mindreading concept. So, it is tempting to dismiss both problems as not empirically concerning. I have argued that this is a mistake. The logical problem for causal mindreading is subsumed by the problem for conceptual mindreading; and the problem for conceptual mindreading can be solved empirically. In virtue of possessing mental state concepts, conceptual mindreaders have two distinctive, empirically detectable features. First, because they represent mental states as

intentional, their representations of mental states are systematic across situations in which the same attitude is directed on different objects, which are both perceptually and behaviourally disparate. Second, conceptual mindreaders appreciate that mental states generate behaviour through practical reasoning, and so that behaviour is flexible. These features are empirically detectable – so, the logical problem for conceptual mindreading is an empirical one. Insofar as we are interested in determining what chimpanzees know about mental states, it is a problem worth taking seriously.

References

Andrews, K. (2005). Chimpanzee theory of mind: Looking in all the wrong places? *Mind & Language*, *20*(5), 521–536. http://doi.org/10.1111/j.0268-1064.2005.00298.x

Andrews, K. (2016). Snipping or editing? Parsimony in the chimpanzee mind-reading debate. *Metascience*, *25*(3), 377–386. http://doi.org/10.1007/s11016-016-0119-0

Andrews, K. (2017). Chimpanzee mind reading: Don't stop believing. *Philosophy Compass*, *12*(1), 1–12. http://doi.org/10.1111/phc3.12394

Bermúdez, J. L. (2003). *Thinking Without Words*. Oxford: Oxford University Press.

Brentano, F. (2015). *Psychology from an Empirical Standpoint*. London: Routledge.

Buckner, C. (2014). The semantic problem(s) with research on animal mind-reading. *Mind & Language*, *29*(5). http://doi.org/10.1111/mila.12066

Bugnyar, T., Reber, S. A., & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications*, *7*, 10506. http://doi.org/10.1038/ncomms10506

Burge, T. (forthcoming). Do infants and non-human animals attribute mental states? *Psychological Review*.

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). Great apes distinguish true from false beliefs in an interactive helping task. *PLoS ONE*, *12*(4). http://doi.org/10.1371/journal.pone.0173793

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192.

Camp, E. (2009). Putting thoughts to work: concepts, systematicity and stimulus independence. *Philosophy and Phenomenological Research*, *78*(2), 275–311.

Carruthers, P. (2004). On Being Simple Minded. *American Philosophical Quarterly*, *41*(3), 205–220.

Carruthers, P. (2009). Invertebrate concepts confront the generality constraint (and win). In R. W.

Lurz (Ed.), *The Philosophy of Animal Minds*. Cambridge: Cambridge University Press.

Clatterbuck, H. (2016). The Logical Problem and the Theoretician's Dilemma. *Philosophy and Phenomenological Research*, 1–29. http://doi.org/10.1111/phpr.12331

Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.

Halina, M. (2015). There is no special problem of mindreading in nonhuman animals. *Philosophy of Science*, *82*(3), 473–490. http://doi.org/10.1086/681627

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, *21*(1), 101–114.

Hutto, D. D., Herschbach, M., & Southgate, V. (2011). Editorial: social cognition: mindreading and alternatives. *Review of Philosophy and Psychology*, *2*(3), 375–395. http://doi.org/10.1007/s13164-011-0073-0

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015). The goggles experiment: can chimpanzees use self-experience to infer what a competitor can see? *Animal Behaviour*, *105*, 211–221. http://doi.org/10.1016/j.anbehav.2015.04.028

Keijzer, F. (2013). The Sphex story: How the cognitive sciences kept repeating an old and questionable anecdote. *Philosophical Psychology*, *26*(4), 502–519. http://doi.org/10.1080/09515089.2012.690177

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. http://doi.org/10.1126/science.aaf8110

Lurz, R. W. (2011). *Mindreading Animals: The Debate over What Animals Know about Other Minds*. Cambridge, Massachusetts: MIT Press.

Lurz, R. W. (2018). Animal mindreading: The problem and how it can be solved. In K. Andrews & J. Beck (Eds.), *The Routledge Handbook of Philosophy of Animal Minds*. Abingdon: Routledge. http://doi.org/10.4324/9781315742250

Lurz, R. W., Krachun, C., Mahovetz, L., Wilson, M. J. G., & Hopkins, W. (2018). Chimpanzees gesture to humans in mirrors: using reflection to dissociate seeing from line of gaze. *Animal Behaviour*, *135*. http://doi.org/10.1016/j.anbehav.2017.11.014

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences*, *20*(5), 375–382. http://doi.org/10.1016/j.tics.2016.03.005

Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (Pan troglodytes) conceal visual and auditory information from others. *Journal of Comparative Psychology*, *120*(2), 154–162.

http://doi.org/10.1037/0735-7036.120.2.154

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind." *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 731–744. http://doi.org/10.1098/rstb.2006.2023

Penn, D. C., & Povinelli, D. J. (2013). The comparative delusion: the behavioristic/mentalistic dichotomy in comparative theory of mind research. In J. Metcalfe & H. S. Terrace (Eds.), *Agency and Joint Attention* (pp. 62–78). New York: Oxford University Press.

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

Povinelli, D. J., Eddy, T. J., Hobson, R. P., & Tomasello, M. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, *61*(3), i+iii+v-vi+1-189.

Povinelli, D. J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, *19*(1), 1–28. http://doi.org/10.1111/j.1468-0017.2004.00244.x

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *4*, 515–526.

Schwitzgebel, E. (1999). Representation and desire: a philosophical error with consequences for theory-of-mind research. *Philosophical Psychology*, *12*(2), 157–180.

Sober, E. (2015). *Ockham's Razors: A User's Manual*. Cambridge: Cambridge University Press.

Tomasello, M., & Call, J. (2006). Do chimpanzees know what others see - or only what they are looking at? In S. Hurley & M. Nudds (Eds.), *Rational Animals?* (pp. 371–384). Oxford: Oxford University Press.

Whiten, A. (1996). When does smart behaviour-reading become mind-reading? In P. Carruthers & P. K. Smith (Eds.), *Theories of Theories of Mind* (pp. 277–292). Cambridge: Cambridge University Press.

Whiten, A. (2001). Theory of mind in non-verbal apes: Conceptual issues and the critical experiments. In D. M. Walsh (Ed.), *Naturalism, Evolution and Mind* (pp. 199–224). Cambridge: Cambridge University Press.