Theory-Theory and the Direct Perception of Mental States

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

Theory-theory and the direct perception of mental states

Abstract

Philosophers and psychologists have often maintained that in order to attribute mental states to other people one must have a 'theory of mind'. This theory facilitates our grasp of other people's mental states. Debate has then focussed on the form this theory should take. Recently a new approach has been suggested, which I call the 'Direct Perception approach to social cognition'. This approach maintains that we can directly perceive other people's mental states. It opposes traditional views on two counts: by claiming that mental states are observable and by claiming that we can attribute them to others without the need for a theory of mind. This paper argues that there are two readings of the direct perception claims: a strong and a weak one. The Theory-theory is compatible with the weak version but not the strong one. The paper argues that the strong version of direct perception is untenable, drawing on evidence from the mirror neuron literature and arguments from the philosophy of science and perception to support this claim. It suggests that one traditional 'theory of mind' view, the 'Theory-theory' view, is compatible with the claim that mental states are observable, and concludes that direct perception views do not offer a viable alternative to theory of mind approaches to social cognition.

Key words: Direct Perception, Mindreading, Mirror neurons, Social cognition, Theory of mind, Theory-theory.

1. Introduction

Our ability to interact socially with others is central to the human species. Yet despite the prevalence of social interactions in our everyday lives, relatively little is known about the cognitive processes which facilitate such interactions. Many philosophers and psychologists believe that attributing mental states to others is

central to our ability to interact socially with them.[1] If one accepts that this is the case then the question arises: 'how do we attribute mental states to others?'

Until relatively recently the dominant thought was that attributing mental states to other people required having a theory, known as the 'theory of mind' (Premack & Woodruff, 1978). 'Theory of mind' enables you to recognise what another's mental states are, and without the theory one would be unable to attribute mental states to others. One is then left with the question 'What kind of structure does the theory of mind have?' 'Theory-theorists' believe that the theory of mind literally has the structure of a theory, and that 'mental state' is a theoretical term which can only be grasped from within the framework of the theory of mind (Botterill, 1996; Carruthers, 2009; Gopnik & Wellman, 1992; Gopnik & Meltzoff, 1997; Wellman, 1992). Thus, knowing about mental states in general requires having a theory of mind, regardless of whether those mental states are your own or other people's. 'Simulation theorists', on the other hand, believe that you can know your own mental states without a theory of mind by using introspective processes (Goldman, 2006, ch.9) but agree with Theory-theorists that knowing other people's mental states requires using a theory of mind. Unlike Theory-theorists, though, Simulationists do not believe that the theory of mind literally has the structure of a theory. Instead, they claim that we use simulative processes to project ourselves into the position of the other, ascertain what kind of mental states we would have in that situation, and then infer that the other person has these mental states (Goldman, 2006; Gordon, 1996; Heal, 1996).

The debate between Simulation and Theory-theorists has dominated the social cognition literature. Recently, however, an alternative to these views has been proposed: the 'Direct Perception approaches to Social Cognition' (henceforth 'direct perception' approaches). There are a number of different direct perception views, but they share a common claim: that we do not need to use a theory of mind to know other people's mental states because we can directly perceive them (Gallagher, 2008; Gallagher & Zahavi, 2008; Gallese, 2007;

[1] Many, but not all. Some philosophers believe that a large part of our social cognition takes place without having to attribute mental states to others, e.g. Daniel Hutto (2008; forthcoming), Adam Morton (2007) and Matthew Ratcliffe (2007).

Reddy, 2008). Thus direct perception approaches stand in opposition to both Theory-theory and Simulation theory.

This paper examines the challenges that the direct perception views raise for the Theory-theory.[2] The next section outlines Shaun Gallagher's direct perception view, and its purported differences to the Theory-theory. Section three suggests that there are two ways of interpreting the direct perception view: a weak and a strong way. The weak interpretation is supported by what I call the 'Inferential step' argument, and is compatible with the Theory-theory. The strong interpretation is supported by the 'Process' argument and is not compatible with the Theory-theory. Sections four and five examine the 'Process' argument from theoretical and empirical perspectives, and conclude that the argument is not a good one, and that therefore the strong interpretation of direct perception fails. Direct perception views thus do not offer a viable alternative to Theory-theory accounts: if one is to take a weak view of direct perception then one has a position that is compatible with the Theory-theory; and the strong view of direct perception is seriously challenged by empirical and theoretical worries. Finally section six clarifies a frequently misunderstood aspect of the Theory-theory, namely, its relation to the observation of mental states.

2. The 'Direct perception' of mental states

2.1. The conflict between Theory-theory and direct perception approaches

There are many ways of supporting the claim that we can directly perceive mental states, but I will be focussing primarily on Shaun Gallagher's arguments. Gallagher's thesis is this:

> The basic claim that I will defend is that in most intersubjective situations we have a direct understanding of another person's interactions because their intentions are explicitly expressed in their embodied actions, and mirrored in our own capabilities for action. For the most part this understanding does not require the postulation of some belief or desire that is hidden away in the other person's mind, since what we might reflectively or abstractly call their belief or desire is expressed directly in their behaviour. (Gallagher, 2005, p. 224)

[2] Whilst my focus is on the Theory-theory, many of the arguments discussed apply to the more general view that we need a theory of mind to know other people's mental states.

Direct perception views see themselves as opposed to the Theory-theory for a number of different reasons. To see why, it is useful to look at Theory-theory as understood by Gallagher and Zahavi (both of whom endorse direct perception views):

> Theory-theory holds that the understanding of minded beings (be it of oneself or others) is theoretical, inferential, and quasi-scientific in nature. It views the attribution of mental states as a matter of inference to best explanation and prediction of behavioral data and argues that mental states are unobservable and theoretically postulated entities. It consequently denies that we have any direct experience of such states. Many philosophers (phenomenologists included) would claim that we need concepts in order to extract and comprehend the informational richness of what is already given [...]. When Theory-theory claims that the attribution of mental states is theoretically mediated, however, it has something more radical in mind. The idea is basically that the employment of theory allows us to transcend what is given in experience. (Gallagher & Zahavi, 2008, p. 172)

The points of conflict between Theory-theory and Direct perception views, on this reading of Theory-theory, can thus be summarised as follows:

1. Theory-theory 'argues that mental states are unobservable'. Direct perception views claim that they are observable and 'transparent to perception' (Reddy, 2008, p. 15).

2. Theory-theory 'denies that we have any direct experience' of mental states, because we have to 'transcend' our experience using theory of mind to find out about mental states. Direct perception views maintain that we have a 'direct understanding of other people's interactions'.

In what follows I will maintain that it is not necessarily true that the Theory-theory is committed to mental states being unobservable, and that the Theory-theorists can accept that we have a direct experience of mental states. Before doing so, however, it is important to outline the commitments of the direct perception approaches, as proposed by Gallagher.

2.2 What is direct perception?

The first thing that needs to be ascertained is what is meant by 'direct perception'.

> What we mean by direct perception (or a direct perceptual grasp) is nothing more than perception itself, not in contrast to indirect perception, but in the sense that all perception is direct. [...]The relevant contrast is not between direct and indirect perception but between perception and something added to perception, e.g. an inference or interpretation that goes

> beyond what is perceived. If I directly see my car I do not ordinarily have to make an inference on the basis of what I see that it is my car. (Gallagher, 2008, p. 537)

In order for a perceptual experience to be direct, it has to be the case that there is 'nothing added' to the contents of that perceptual experience. Gallagher's perception of his car is direct because the knowledge that it is his car is immediately available in the contents of the perceptual experience he has when he looks at it. This is contrasted against an experience where 'something is added to perception'. An experience where something is added to perception is one where one has a perceptual experience, and has to engage in some kind of inferential process before one knows what is perceived. Gallagher offers the example of seeing a metal wreck and using a few visual clues before knowing that what is perceived is his car after an accident. In this case, Gallagher must 'add' some extra processes to his initial visual experience in order to know that he is looking at his car; because of this, his perception of his car is not a direct one. Direct perceptions are those perceptual experiences where knowledge of what one is looking at is immediately available in the contents of that experience; one does not have to engage in any additional process to gain knowledge about what is perceived.

Gallagher then goes on to distinguish between two different types of direct perception: 'smart' and 'not-so-smart' direct perception (ibid p. 536). A direct perception experience which is 'not-so-smart' is one where very little information about the object is available in the content of the perceptual experience. Gallagher offers the example of opening his eyes to see 'a certain unrecognized red mass with a specific shape just in front of me' (ibid). This is an instance of a direct perception which is not so smart: his perceptual experience of the object contains only the colour and shape of what is in front of him, and that is about all. He contrasts this with a 'smart' direct perception of the same object, when he opens his eyes and sees his car. His direct perception is smart because he can see a car, something which is more meaningful to him than the red mass of the previous perception. Although both perceptual experiences are direct, one contains more information about what is perceived than the other, and so is classified as a 'smart' direct perception.

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

2.3 'Smart' direct perception and mental states

Gallagher claims that we can directly perceive other people's mental states because our direct perceptions of other people are 'very smart' (ibid, p.538). Being very smart means that our perceptual experiences of other people contain a lot of information about them, including information about their mental states. A not-so-smart direct perception of another person would be one where you see their hand moving from a point near their body to a point on top of the cake on the table. A smarter direct perception of the same event would be one where you see the person reaching for the cake, in other words, you see their intention to grasp the cake.[3] The second type of perception contains more information about what is perceived than the first, and is therefore 'smarter'. Gallagher maintains that our perceptions of other people are nearly always smart perceptions of this second kind, and as a consequence we can directly perceive their mental states.

3. The 'Inferential' step and 'Process' arguments

Gallagher's claim that we see people through 'smart' perception is intended to offer an alternative to Theory-theory. My interpretation of his position is that he is running two slightly different but complementary arguments against Theory-theory: the 'Inferential step' argument, and the 'Process' argument.

3.1. The 'Inferential step' argument

The 'Inferential step' argument maintains that because our perceptions of other people are 'smart', we do not have to add anything to what is immediately available in the contents of our perceptual experience in order to know what the other person's mental states are. When Gallagher sees his car through smart perception he does not have to 'make an inference on the basis of what I see that it is my car', and analogously we do not ordinarily have to make an inference on the basis of what we see of other people to know their mental states. I've called this the 'Inferential step' argument because advocates of direct perception claim that Theory-theorists are committed to the existence of an 'inferential step' that allows

[3] You can only understand a movement as a reach if you already understand that the movement is caused by an intention to grasp the cake. If you didn't have this background knowledge you would not understand the movement as a 'reach'.

us to 'transcend what is given in experience' in order to know another's mental state.

The 'Inferential Step' argument.

1. Information about another's mental states is directly available in the contents of our visual experience: we do not need to 'transcend what is given in experience' to know what their mental states are (Gallagher & Zahavi, 2008, p. 172).
2. Theory-theory claims that knowing another's mental states requires 'transcending what is given in experience' (ibid)
3. If *1* is true, then the Theory-theorist's claim in *2* is false.
4. *1* is true.
5. Therefore the Theory-theorist's claim in *2* is false.

Phenomenological experience supports premise one. It's rare that we find ourselves watching another person's behaviour and trying to infer what their mental states are. Most of the time we see Bob reaching for the cupcake, rather than seeing Bob's body move in a certain way, then inferring that Bob wants the cupcake. It looks like the direct perception views have made a good point against the Theory-theorists.

Before going any further we need to note an important distinction: that between 'personal' and 'sub-personal' explanations of perception. This distinction is articulated in a number of different ways in the literature, but for present purposes I will borrow Gallagher's articulation of the distinction, namely, that 'sub-personal' explanations of perceptual experiences are those explanations that cite neural and cognitive processes that we are not aware of, and that such explanations should be distinguished from 'conscious perception itself'. He writes,

> We want to distinguish between a sub-personal explanation of perception, and conscious perception itself. At the personal, or conscious level, I do not have to perceptually piece together the shape and the color and the mass in order to get my car. Even if the sub-personal processes are complex (and I do not deny that they are), the perception that I have of my car is direct – I see it right there in front of me. I do not have to glue anything together, add an interpretation or add an inference. (2008, p. 537)

We are now in a better position to assess the scope of the 'Inferential step' argument. The direct perception view claims that Theory-theory requires that we 'transcend what is given in experience'. The Inferential step argument is thus

challenging the Theory-theory's account of what happens at a personal or conscious level of perception. On the direct perception view, information about another's mental states is part of our conscious perceptual experience, one need not add another process to the conscious experience in order to know it. According to the direct perception views, the Theory-theorist claims that we must add an 'inferential' or theory-like step once we have our conscious experiences of others in order to know their mental states.

The problem with the Inferential Step argument is that one would struggle to find a Theory-theorist who agreed with premise two. The Theory-theorist is not making a claim about what happens at the personal or 'conscious' level of experience; the Theory-theory is instead intended as a 'sub-personal level' explanation of our perceptual experiences. Theory-theorists do not believe that one has a conscious experience of another which does not include information about their mental states, before 'adding' an inferential step to gain this knowledge. Rather, their claim is that one needs a theory of mind in order for this information to be part of our conscious experience. The Theory-theorist maintains that the cognitive processes that facilitate our conscious perceptual experiences of others are theory-like. As such, the Inferential Step argument does not offer a challenge to the Theory-theory: both Theory-theory and direct perception views acknowledge that one does not need to 'transcend what is given in experience' or 'add' anything to our conscious perceptual experiences of others in order to know about their psychological states.

3.2 The 'Process Argument'

Unlike the 'Inferential Step' argument, the 'Process' argument challenges the Theory-theory's 'sub-personal' explanation of the processes which facilitate our grasp of other people's mental states. As just discussed, the Theory-theory and direct perception views agree that we do not use theory-like processes at a conscious or personal level to ascertain another's mental states. However, whilst the Theory-theorist maintains that the sub-personal processes which facilitate such conscious experiences are theory-like, the direct perception views claim that they are not, drawing on data from mirror neurons to support this claim.

Mirror neurons are a group of neurons which discharge both when we perform an action and when we perceive other people perform that same action. Importantly, the same neurons are active in both these cases. Philosophers and psychologists sympathetic to the direct perception approach have claimed that the activity of an observer's mirror neurons when she perceives someone acting is sufficient to provide her with knowledge about the other's intentions. There is some disagreement about how exactly mirror neurons provide this information (Gallagher, 2007; Gallese, 2009b), but what is not in dispute within this group is that mirror neurons facilitate the direct perception of mental states. Direct perception views argue that because the discharge of a neuron is not the kind of process that is inferential or theory-like, and mirror neurons enable information about another's mental states to be part of the 'smart' direct perceptual experience, mirror neurons support the claim that our knowledge of mental states is not provided through theory-like processes. I call this the 'Process' argument.

<u>The Process argument</u>

1. The discharge of mirror neurons provides information about another's intention.
2. The information provided by mirror neurons about other people's intentions becomes part of the contents of our 'smart' perceptual experiences of other people.
3. [from 2] Information about other people's intentions is part of our 'smart' perceptual experience of them.
4. [*Implicit*] The discharge of mirror neurons is 'automatically induced' (Rizzolatti & Craighero, 2004, p. 172) and therefore this process cannot be described as either 'theory-like' or 'inferential'.
5. Therefore, the processes that generate information about other people's intentions are not theory-like or inferential.
6. Theory-theory maintains that the processes that provide us with information about other people's intentions are theory-like and inferential.
7. [from 5 & 6] Therefore Theory-theory is false.

Premise 3 is where the 'Inferential step' argument comes in. The 'Inferential Step' argument is intended to support the claim that theory of mind has no role to play in detecting other people's mental states because information about these mental states is already available in the contents of experience. The 'Process' argument pushes the claim that the processes which enable such information to be part of the contents of visual experience are not inferential ones either.

3.3 Strong and weak direct perception

There are two versions of the claim that we can directly perceive mental states: a weak and a strong one. The weak version makes a descriptive claim about our conscious experiences of other people's mental states, by saying that information about other's mental states is available in our conscious experiences of them and we need not add an extra step to our experience in order to know such information. The weak version is silent on the processes facilitating our conscious experiences. The Inferential step argument is intended to support this weak version of direct perception. I have argued that the Theory-theory is compatible with the weak version of direct perception, because it does not accept premise two of the Inferential step argument.

The strong version of direct perception makes a descriptive claim about the sub-personal processes that facilitate our conscious experiences, by maintaining that these sub-personal processes should not be understood as theory-like. The strong version also claims, like the weak version, that information about another's mental states is immediately available in our conscious perceptions of them. The Process argument is intended to support the strong version of the direct perception view. Strong direct perception is not compatible with the Theory-theory, as Theory-theory believes that the sub-personal processes facilitating our conscious experiences of others are theory-like. In what follows I focus on the Theory-theorist's response to the Process argument.

4. The theory-laden nature of perception

4.1 'Epistemic' and 'Non-epistemic' seeing

'Seeing' is an ambiguous term which we use in everyday discourse to cover a whole range of visual phenomena. 'Seeing' can refer to a visual experience where one sees shapes and colours but has no beliefs about them, and it also refers to those visual experiences which we do have beliefs about, such as visual experiences of chairs or zebras. Fred Dretske (1969) disambiguates 'seeing' by distinguishing two kinds of visual experience: 'Epistemic' and 'Non-epistemic'

seeing.[4] 'Non-epistemic seeing' refers to those visual experiences where you see something but have no beliefs about it. For example, Bob may look into the night sky and see a point of light. This is a case of non-epistemic seeing. One could object that this is not a case of non-epistemic seeing because Bob does have a belief about his visual experience: Bob believes that he is looking at a light in the sky. The reason this counts as an instance of non-epistemic seeing is that the beliefs that Bob has about the specific content of the visual experience are not relevant for this experience. Contrast this with a case of epistemic seeing: Bob looks into the night sky and sees the North Star. In order for Bob to see the North Star it is necessary but not sufficient that his neural and visual systems detect properties of light. The detection of these light stimuli does not amount to epistemic seeing, nor do the sensations which are triggered in Bob once these reactions occur. (Bob could have a particular phenomenal sensation when he detects 'properties of light$^{A'}$ and can distinguish this sensation from the sensation he detects 'properties of light^{B}'. However, this ability to distinguish one kind of sensation from another does not amount to having beliefs about the objects which caused these sensations.[5]) A visual experience of seeing the North Star is an instance of epistemic seeing because it requires that Bob has certain beliefs about what he is looking at: that it is a star; that it marks north.

This distinction between two types of seeing has some interesting consequences. One is that two people can detect the same light properties in the same way, but one person may be having an epistemic seeing experience whilst the other is having a non-epistemic seeing experience. Another consequence is that someone could have the same epistemic seeing experience of an object for years until a new theory causes them to have a different one. Astronomers saw Uranus epistemically for years before Herschel discovered it. The difference was that the theory shaping those astronomers' epistemic seeing was that the point of light they were looking at was a star. It wasn't until Herschel spotted that the point of light did not move in a stellar way that astronomers had an epistemic

[4] In later writings Dretske prefers to use the terms 'simple seeing' and perception (1979/2000). Dretske is not the only philosopher to have noted this distinction. Norwood Russell Hanson distinguishes between seeing and 'seeing as'; the distinction also corresponds with Gallagher's 'smart' and 'not-so-smart' perceptions.

[5] This distinction is also made by Paul Churchland (1979, p. 14)

seeing experience of that point of light as a planet. The astronomers' change in theoretical commitments affected the contents of their epistemic perceptions.

4.2 'Epistemic' seeing and 'Smart' perception

The case of epistemic seeing shows us how the organisation of our theoretical commitments plays an essential role in determining the contents of our perceptual experiences. Both 'epistemic' and 'non-epistemic' seeing are instances of 'direct perception' in Gallagher's sense. They are perceptual experiences with certain contents: in the case of 'non-epistemic' seeing the subject has no beliefs about the contents of the visual experience, whereas in the case of 'epistemic' seeing the subject has some beliefs about the contents of the visual experience. One's beliefs about what is perceived affects the contents of the epistemic perception, and the theories one holds also affect these contents. When it comes to mental states, one's beliefs about mental states and how these beliefs are organised will affect the epistemic contents of your visual experience. Theory-theorists maintain that it is the theory of mind which facilitates our epistemic seeing experiences of mental states, and that such a theory is thus required in order to perceive mental states. Direct perception views, on the other hand, need some explanation of how information about another's mental states can become part of the epistemic, or 'smart' visual experience in the absence of a theory of mind. This they do by appeal to mirror neurons, a move we will examine presently. For now all that is important is the Theory-theorist's line that, in order for epistemic perceptual experiences of another's mental states to be possible, the subject must have a theory of mind. The contents of an epistemic perception depend on the beliefs and theoretical commitments of the subject, and in the case of mental states, the theoretical commitments in question are those of the theory of mind.

4.3. Epistemic seeing and inference

The strong formulation of direct perception is not new, although Gallagher's application of it to mental states is. James Gibson (1979) proposed that there are certain things which we can see without the use of any theory-like process. Gallagher is clearly influenced by Gibson's work (Gallagher and Zahavi, ch.5) when making his claims about the direct perception of mental states.

However, Gibson's arguments for the possibility of direct perception have been challenged, most prominently by Jerry Fodor and Zenon Pylyshyn (1981). The many criticisms they raise against direct perception are comprehensively damaging to the view, however, I wish to focus on only one of their arguments here, concerning their defence of the claim that perception necessarily involves an inferential process. They refer to their position as the 'Establishment' position.

At the heart of Fodor and Pylyshyn's concern is the problem of under-determination. Organisms detect properties of light, and these properties of light have in turn been caused by properties of the layout that the organism is looking at. Both the Establishment and direct perception views concur on this. What is directly 'picked up' by the organism's perceptual mechanisms are properties of light, and not properties of the layout which caused that light. However, whilst the direct perception view claims that the information contained within the light properties that are detected is 'almost invariably sufficient to specify the ecologically relevant properties of the layout', the Establishment claims that the properties of light detected 'in general significantly *underdetermine* the percepts to which they give rise' (Fodor & Pylyshyn, 1981, pp. 169, 171). In other words, if one takes the direct perception view, one is committed to the claim that each set of light properties detected by an organism uniquely correlates with a particular layout. Thus, the information contained within the light properties detected is sufficient to furnish the organism with knowledge about the layout it is perceiving. In contrast, the Establishment view maintains that there is not a one-to-one correlation between light properties and the layouts that they correlate with, and that therefore the detection of light properties alone is insufficient to provide information about the layout which the organism is looking at. The detection of light properties is insufficient for 'epistemic' or 'smart' perceptual experiences. As Fodor and Pylyshyn write:

> Properties of the layout are inferred from properties of the light, and [...] the directly detected properties of the light generally underdetermine features of the layout. [...] So long as the organism is detecting light patterns and not layout patterns, there is no way to avoid the conclusion that perceptual knowledge of the latter is inferred. (Ibid, p. 186)

According to the Establishment view, 'Epistemic Seeing' just isn't possible without a process of inference. An organism can only detect properties of light,

and these are not sufficient to determine what the layout causing those light properties is. Bob's visual system can detect properties of light which are sufficient to tell him he is looking at something red and black. But the detection of these light properties is not sufficient to inform Bob that he is looking at his car. In order for Bob to have a perceptual experience of a car, other knowledge has to be added to the light information detected by his visual systems. This is an inferential process. The light detections are one 'premise' in this process, Bob's additional knowledge forms other premises, until the inference to the best explanation for the light detections is formed: Bob is looking at a car. All this happens sub-consciously and is cognitively impenetrable. Bob will have a perceptual experience of a car unless something happens for his experience to change, e.g. someone picks up the car with one hand. When new knowledge like this is gained Bob's perceptual experience will change from one of looking at a car, to looking at a model car. Fodor and Pylyshyn argue convincingly that there is no way one can get from light detections to an epistemic seeing experience without an inferential process. Therefore, epistemic seeing involves inferential processes.

The problem of underdetermination becomes more pressing when one considers the claim that we can directly perceive mental states. We have already examined the claim that the mapping between distal layouts and the light properties they cause are many-to-one. In the case of behaviour this is compounded because the same movement can be caused by a number of different mental states. Bob might pick up the cup because he wants to drink from it, because he wants to throw it at someone, because he wants to tidy it away. Each of these mental states will cause the same behaviour – that of picking up the cup. So the direct perception view has to explain how we can directly perceive mental states without any inferential processes when...

a. The mapping between distal-layout to light properties is many to one, and
b. The mapping between the mental states that caused the movements and the movements themselves is many to one.

Even if we grant that we can directly perceive the movement (which, in the light of the arguments just given is unlikely), advocates of the direct perception view

still need to explain how we can get round the problem that a movement underdetermines the mental state that caused it.

Fodor and Pylyshyn's problem of underdetermination is a crippling one for the claim that we can directly perceive anything. When it comes to mental states, the problem appears twofold: one has to explain how we directly perceive that it is a person moving, as well as how we directly perceive their mental states, when in both cases the information available to direct perception underdetermines the contents of the perceptual experience. However, strong direct perception views have a way of countering the problem that movements underdetermine the mental states that caused them in the form of the mirror neuron data. Mirror neurons appear to provide very detailed information about another's intention, moreover, they provide this information in a non-inferential way. So perhaps the strong direct perception view can be salvaged with this data. It may be the case that inferential processes are required to generate a smart/epistemic perceptual experience of the other person, but one aspect of this experience, namely, the perception of their mental state, is provided in a non-inferential way. There is still hope for the strong direct perception view.

5. Mirror neurons and the 'Process' argument

The discharge of mirror neurons is said to provide us with 'a direct form of understanding others from within' (Gallese, 2009a, p. 494). Moreover, their discharge is thought by the direct perception view to be sufficient to provide the subject with information about another's mental states 'without the benefit of anything that on the sub-personal level could be considered an extra cognitive step, a simulation or an inference' (Gallagher 2008, p. 542). Mirror neurons thus support the Process argument, and in doing so, the strong direct perception view. The strong direct perception interpretation of the mirror neurons data has been criticised by a number of researchers, (Csibra, 2005; Jacob, 2008; Southgate, Johnson, Osbourne, & Csibra, 2009) and although their critiques are different they share a common worry: that the discharge of mirror neurons is not capable of providing sufficient information about another's intentions. In other words, the discharge of mirror neurons underdetermines the knowledge we have about another's intentions.

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

5.1 The problem of context

An important discovery in the mirror neuron field is that of 'Parietal mirror neurons'. Parietal mirror neurons are said to provide not only information about 'what' an agent's action is, e.g. that it's a 'grasp', or a 'push', or a 'bite', but also information about 'why' the agent is acting as she does, e.g. 'she's grasping the apple in order to give it to her sister'. The 'why' of an action is referred to throughout this literature as the 'intention' of that action; in the case just discussed the agent's intention was to give the apple to her sister.[6] The discovery of parietal mirror neurons led researchers to claim that the mirror neuron system 'does not simply provide an action recognition mechanism ("that's a grasp") but rather it is critical for understanding the intentions behind other's actions' (Iacoboni et al, 2005, p. 531).

The most well-known study with parietal mirror neurons is Iacoboni and colleagues' 'tea' study. In this experiment, subjects watched the same motor act – picking up a cup – performed in two different contexts. Subjects watched three film clips. One was a *context* condition, where participants saw either a scene of tea things (teapot, cups, milk jug etc) set out for tea, or a scene of the tea things after tea, with crumbs on the saucer and empty cups. The second was an *action* condition, where participants saw a cup against a blue background, and someone's hand reached to pick it up. The third was an *intention* condition, where participants saw one of the scenes described in the context condition, and a hand reaching in and picking up a cup. The results showed that subjects' parietal mirror neuron activity was different when they watched the cup being picked up in the during- and after- tea conditions, suggesting that the parietal mirror neurons were sensitive to the intention of the action: in the 'during tea' condition the cup was being picked up with the intention of 'drinking', whereas in the 'after tea' condition the cup was being picked up with the intention of 'clearing away'. This study thus supports the hypothesis that parietal mirror neurons enable us to grasp another's intentions.

Can parietal mirror neurons really provide us with all the information necessary to know another's intentions, and, moreover, do so in a 'direct', non-

[6] 'Intention' is thus used in this literature slightly differently to its use in the philosophy of mind.

inferential way? One of the problems facing these data is that of context.

In our society, the etiquette at the start of teatime is to drink from the cup. Hence, when I perceive Bob grasping a cup of tea, my parietal mirror neurons should discharge in the same way as they would if I had the intention to drink from a cup. But now imagine someone from a society where it is customary to throw the first cup of tea over your shoulder, and that this person is unaware of Western tea etiquette. When this person perceives Bob grasping the cup of tea, their parietal mirror neuron discharge will be the same as when they have the intention to throw the cup of tea over their shoulder. Clearly cultural norms, and our assumption that others will adhere to them, will affect the discharge of our parietal mirror neurons. But this raises the question - Is the contextual information (including background norms as well as immediate factors) required to ascertain an intention generated by parietal mirror neurons, or is it produced elsewhere in the cognitive system and fed into the mirror neuron system?

When a parietal mirror neuron system discharges in the 'tea drinking way', and is correct, then this means that the parietal mirror neuron system has ascertained that what is being perceived is a tea party, has information about all the cultural norms that tea parties entail, as well as an expectation that the person perceived will abide by these cultural norms. But it is mysterious how the parietal mirror neurons could contain and process such information. In response to this concern, Gallese says that it is just as mysterious to claim that this information processing takes place elsewhere in the cognitive system, and then used as input to the mirror neuron system. fMRI scanning, he argues, reveals no evidence that this information is processed somewhere other than the parietal mirror neuron system:

> We stick to the extant empirical evidence and our claim is that we don't need to suppose an over-arching top-down influence in order to have a neural mechanism that maps the goal. We already have it in the premotor system. We don't need to imply a further mechanism that maps the goal. (Gallese, 2006, p. 15)

But the problem of contextual information persists. If one accepts Gallese's claim that the mirror neurons provide the 'neural mechanism that maps the goal' of an action, then one must be able to explain how the mirror neuron system organises and accesses the relevant contextual information. This claim simply moves the

problem of how one attains the contextual information required for successfully recognising a goal from processes outside of the mirror neuron system to within the mirror neuron system. Now one has to explain how the mirror neuron system accesses the correct contextual information and how that information is organised. Given the problems of underdetermination already discussed, it seems reasonable to think that contextual information is provided through an inferential process. But this undermines the claim that mirror neurons provide direct information about another person's goals. Mirror neurons may provide this information once contextual information about situations and norms has been processed, and perhaps, as Gallese claims, such information is processed within the mirror system. But these processes are surely inferential.

One possible way of rescuing the claim that mirror neurons can provide non-inferential information about another's action is to maintain that mirror neurons are formed through an associative process (Heyes, 2001, 2005, 2010). On Heyes' view, mirror neurons are an evolutionary by-product formed through associative learning, and as such their discharge is not governed by inferential processes. But, as Heyes observes, if mirror neurons are a by-product of evolutionary processes then it is unlikely that they have a specific function (e.g. facilitating action understanding). This does not preclude the possibility that their discharge may contribute to action understanding, but it does challenge the claim that their activation is necessary (Heyes, 2010). Although Heyes' associative hypothesis offers the possibility that mirror neuron discharge is not governed by inferential processes, she readily acknowledges that, on this account, the contribution of mirror neurons to action understanding is as yet unknown. Careful empirical research is required to ascertain whether the associative account can be used to support the claim that mirror neurons facilitate the direct perception of mental states. This opens up an important new debate, which it is not within the scope of this article to address. The claim of this section is simply that, as the data currently stand, mirror neurons are not able to resolve the problem of underdetermination facing the strong direct perception view.

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

5.2 Mirror neurons and visual inputs

According to the direct perception approaches, mirror neuron activity is caused by the observation of another's action (as well as your own execution of that action). The observation of another's action is necessary for mirror neuron activity, because this is the primary source of input to the mirror neuron system. Mirror neurons take visual input and provide information about the intentions of the agent perceived. Thus, Gallese writes that 'mirror neurons require, in order to be activated by visual stimuli, an interaction between an action's agent (human being or monkey) and its object' (Gallese, 2001, p. 35).

Two empirical findings challenge this hypothesis. The first is the discovery of 'canonical mirror neurons', which discharge when an object which is usually used in actions is perceived. For instance, a monkey's canonical mirror neurons discharge in the same way both when the monkey grasps an object and when the monkey perceives a 'graspable' object (Iacoboni et al 2005, p. 79). Secondly, it has also been shown that an infant's mirror neuron system is active even when she has not perceived an interaction between an agent and an object (Southgate, Johnson, Osbourne, & Csibra, 2009). These data are problematic for Gallese's hypothesis that the observation of an action alone induces mirror neuron activity, as he must explain how such activity can take place in the absence of the appropriate observation.

However, if the claim to be defended is simply that mirror neurons provide information about another's intentions, then these data do not provide an insurmountable challenge to the direct perception view. Whilst it may be the case that some mirror neurons discharge even when there is no intention to be perceived, this does not preclude the possibility that mirror neurons are responsible for our understanding of intentions. It may be that some mirror neurons are highly sensitive, and discharge even when no intention is present.[7] But, one could argue that these mirror neurons are not sufficient to yield an understanding of another's intention, and that it is only when more specific mirror neurons discharge that such an understanding comes about. This is an area for further empirical study.

[7] I'm grateful to an anonymous reviewer who suggested this possibility.

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

5.3 Mirror neurons do not support the 'Process' argument

The 'Process' argument supports the strong direct perception view and turns on the premise that 'the discharge of mirror neurons is not the kind of process that we could describe as theory-like'. However, there are reasons to believe that mirror neurons do not support the 'Process' argument. I have mentioned two here, and others have been discussed elsewhere (Csibra, 2007). The empirical worries follow the same underdetermination concerns discussed in section four: visual stimulus alone does not give the mirror neuron system sufficient material to ascertain an intention. The additional information required is not 'direct' visual information, but information about culture, context, and expectations. This information cannot be described as 'direct' as it will be accessed through inferential, theory-like processes. Therefore, the discharge of mirror neurons does rely on inferential processes, because the information required before the discharge of mirror neurons can take place is accessed through inferential processes. Mirror neurons do not support the strong direct perception claim that the sub-personal processes that facilitate our perception of mental states are neither theory-like nor inferential. This, in conjunction with the underdetermination criticism discussed in section four, provides good ground for rejecting the strong direct perception claim.

6. Are mental states observable?

In this final section I challenge the claim, made by Gallagher and Zahavi (2008, p. 172) amongst others (e.g. Reddy, 2008), that Theory-theory is necessarily committed to mental states being unobservable.

6.1 Theoretical ≠ unobservable

There is a common, but false, implicit assumption within the direct perception literature that 'theoretical' entails 'unobservable'. If X is a theoretical entity, then X is unobservable. But this assumption is contentious, as debates in the philosophy of science show. Many philosophers would argue that theoretical terms can refer to observable things (e.g. Churchland, 1979; Lewis, 1970,). 'Shingles' is an example of a theoretical term that does not refer to something invisible. 'Shingles' is a theoretical term; we can only grasp its meaning by

understanding its role within a broader network of beliefs about viruses, their effects and causes. One could not properly understand the term 'shingles' unless one also understood something about viruses. 'Shingles' is also a term that conceptually informs our observations: we see the red spots not just as a series of red spots but as part of the Shingles virus.

However, whilst theoretical entities need not be invisible, they are not visible to those who do not already possess the theory. Bob initially does not observe the shingles on his arm, because he does not know enough about the theory in which that term is embedded. Bob's doctor, on the other hand, does possess a theory of the shingles, and as a consequence is able to observe the shingles on his arm. The moral is that while theoretical entities need not be unobservable, one requires a theory in order to observe them.

It should also be noted that those theoretical entities that are made observable by a theory can be part of a subject's 'epistemic' or 'smart' perception. A trained observer can see the shingles immediately; she need not add an extra step to her perceptual experience in order to see them. Theories conceptually inform our observations, and there is no need to think that this conceptual information is added after a perceptual experience. In other words, this view of theories fits with the 'Process' argument. Theoretical entities come to be part of our 'smart' perceptual experiences because the theory which makes them observable informs these 'smart' perceptual experiences.

6.2 Theory-theory and the observation of mental states

Theory-theorists need not be committed to the view that, because mental state terms are theoretical terms, mental states are unobservable. This has not always been clear, even amongst Theory-theorists. For instance, the Theory-theorist Susan Johnson writes,

> Mental states, and the minds that possess them, are necessarily unobservable constructs that must be inferred by observers rather than perceived directly. (Johnson, 2000, p. 22).

My claim is that mental states are not necessarily unobservable. It is perfectly possible for Theory-theorists to claim that some mental state terms refer to

observable entities, and that once one has a theory of mind one can observe mental states.

What does it mean to say that we can observe some mental states? We perceive certain behaviours as actions, and thus as intentional. I see Bob's movement *as* a reach for the cupcake, which entails seeing it as caused by a particular mental state. 'Reaching' is a psychological term as it implies some kind of pro-attitude for the object being reached for, and it seems perfectly reasonable to say that seeing Bob's movement as a 'reach' amounts to seeing his psychological state. Without a theory of mind I would not be able to see Bob's movement as a reach, I would only see it as a movement of his arm. Theory of mind could thus make certain mental states observable, and this is a position that is perfectly compatible with the Theory-theory.

This possibility does not seem to have been entertained by the Theory-theorists, who frequently claim that mental states are not observable:

> Mental states, such as beliefs and desires, are private, internal and not observable in others. (Wellman, 1992, p. 107)

> Mental states are unobservable, and have complex logical properties...if anything, we should have expected that mental state concepts should be bafflingly difficult to acquire, and yet even the most unremarkable child seems to understand them – without any explicit teaching. (Baron-Cohen & Swettenham, 1996, p. 158)

Implicit within these views is the idea that we can only see the effects of mental states, in the form of behaviour, and that this is insufficient for seeing the mental state itself. It parallels an argument in the philosophy of science which claims that there are certain theoretical entities, like neutrinos, which cannot be observed, even though their effects can be observed. One must infer the presence of a neutrino from what is observable, e.g. Cherenkov radiation in heavy water, and the observation of the radiation is not sufficient to be considered the observation of the neutrino, even though it is caused by the neutrino. Neutrino theory cannot make neutrinos observable. The parallel argument in the philosophy of mind is that seeing Bob's arm move in a certain way does not amount to seeing Bob's psychological state of wanting the cupcake. This is the case even when one has a theory of mind informing one's observations. Theory of mind may allow you to

infer that Bob wants the cupcake, but this does not amount to observing that Bob wants the cupcake.

More debate is needed to ascertain whether the mental states are theoretical entities in the same way that shingles are, or whether they are more akin to neutrinos. The stance one takes on this issue will depend considerably on one's position regarding theory and observation in the philosophy of science; for instance, there are those who suggest that seeing the effects of neutrinos amounts to observing neutrinos, for the current laws of physics dictate that the only particles which can yield these effects can be neutrinos (Hanson, 1958; Shapere, 1982). A parallel argument may present itself for the theory of mind and the observation of mental states. Furthermore, there can be radically different ontological accounts that ground the view on direct perception (van Riel, 2008). My aim here is modest: I want simply to point out that Theory-theorists need not be committed to the claim that mental states are necessarily unobservable (contrary to the portrayals given by the direct perception views, and indeed some Theory-theorists), and that the parallel debates concerning observation and theory in the philosophy of science offer a fruitful new way of addressing the question of whether we can observe another's mental states.

7. Conclusion

Direct perception views have grown in popularity as an alternative to theory of mind approaches to social cognition. They can be understood in one of two ways: as making the weak claim that information about another's mental states is part of our conscious experience of them; or as making the strong claim that the sub-personal processes facilitating our experiences of others' mental states are not theory-like (and agreeing with the weak claim too). The Theory-theory is compatible with the weak version of direct perception, and the strong version does not work. Furthermore, it is possible to understand the Theory-theory in such a way that it makes mental states observable. Further work is needed to secure this last point, but as things currently stand the Theory-theory need not be committed to the unobservability of mental states.

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

References

Baron-Cohen, S., & Swettenham, J. (1996). The relationship between SAM and ToMM: two hypotheses. In P. Carruthers, & P. Smith (Eds.), *Theories of Theories of Mind* (pp. 158 - 168). Cambridge University Press.

Botterill, G. (1996). Folk psychology and theoretical status. In P. Carruthers, & P. Smith (Eds.), *Theories of Theories of Mind* (pp. 105-119). Cambridge University Press.

Botterill, G., & Carruthers, P. (1999). *The Philosophy of Psychology.* Cambridge: Cambridge University Press.

Carruthers, P. (2009). How we know our own minds: the relationship between mindreading and metacognition. *Behavioural and Brain Sciences , 32*, 121 - 182.

Churchland, P. (1979). *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Cohen, A. and German, T. (2009). Encoding of others' beliefs without overt instruction. *Cognition, 111,* 356 - 363.

Csibra, G. (2007). Action mirroring and action understanding: an alternative account. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition: attention and performance XXII* (pp. 435 - 459). Oxford: Oxford University Press.

Csibra, G. (2005, January 2). *Mirror neurons and action observation. Is simulation involved?* Retrieved from http://www.interdisciplines.org/mirror/papers/4.

Dretske, F. (1969). *Seeing and Knowing.* London: Routledge & K.Paul.

Fodor, J., & Pylyshyn, Z. (1981). How direct is visual perception? *Cognition , 9*, 139-196.

Gallagher, S. (2008). Direct perception in an intersubjective context. *Consciousness and Cognition , 17*, 535-543.

Gallagher, S. (2005). *How the Body Shapes the Mind.* Oxford: Clarendon.

Gallagher, S. (2007). Simulation trouble. *Social Neuroscience , 2*, 353 - 365.

Gallagher, S., & Zahavi, D. (2008). *The Phenomenological Mind.* Oxford: Routledge.

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

Gallese, V. (2009a). Motor abstraction: a neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research , 73*, 486 - 498.

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical transactions of the royal society of the biological sciences , 362*, 659 - 669.

Gallese, V. (2006). Embodied simulation: from mirror neuron systems to interpersonal relations. *Empathy and Fairness* (pp. 3 - 20). Novartis Foundation.

Gallese, V. (2001). The 'shared manifold' hypothesis: from mirror neurons to empathy. *Journal of Consciousness Studies , 8*, 33-50.

Gallese, V. (2009b). We-ness, embodied simulation and psychoanalysis: reply to commentaries. *Psychoanalytic Dialogues , 19*, 580 - 584.

Gallese, V., Eagle, M., & Migone, P. (2007). Intentional attunement: mirror neurons and the neural underpinnings of interpersonal relations. *Journal of the American Psycholanalytic Association* , 131 - 176.

Gibson, J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

Goldman, A. (1989). Interpretation Psychologized. *Mind and Language , 4*, 161 - 185.

Goldman, A. (2006). *Simulating Minds.* Oxford: Oxford University Press.

Gopnik, A. (1996). The scientist as child. *Philosophy of science , 63*, 485 - 514.

Gopnik, A., & Meltzoff, A. (1997). *Words, Thoughts and Theories.* Cambridge, Mass.: M.I.T. Press.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind and Language , 7*, 145-71.

Gordon, R. (1996). 'Radical' simulationism. In P. Carruthers, & P. Smith (Eds.), *Theories of Theories of Mind* (pp. 11-22). Cambridge: Cambridge University Press.

Hanson, N. (1958). *Patterns of Discovery.* Cambridge: Cambridge University Press.

Heal, J. (1996). Simulation, theory and content. In P. Carruthers, & P. Smith, *Theories of Theories of Mind* (pp. 75-90). Cambridge: Cambridge University Press.

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

Heyes, C.M. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences, 5*, 245–261.

Heyes, C.M. (2005). Imitation by association. In S. Hurley & N.Chater (Eds) *Perspectives on Imitation: From Mirror Neurons to Memes.* (pp.157 - 176). Cambridge: MIT Press.

Heyes, C.M. (2010). Where do mirror neurons come from? *Neuroscience and Behavioural Reviews, 34,* 575 – 583.

Hutto, D. (forthcoming). Elementary mindreading, enactivist style. In A. Seeman (Ed.), *Joint Attention: New Developments in Philosophy, Psychology, and Neuroscience.* Cambridge, Mass: MIT Press.

Hutto, D. (2008). *Folk Psychological Narratives: The sociocultural basis of understanding reasons.* Cambridge, Mass: MIT Press.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology , 3*, 529 - 535.

Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind and Language* , 190 - 223.

Johnson, S. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences , 4*.

Lewis, D. (1970). How to define theoretical terms. *The Journal of Philosophy , 67*, 427 - 446.

Morton, A. (2007). Folk psychology does not exist. In D. Hutto, & M. Ratcliffe (Eds.), *Folk Psychology Re-assessed* (pp. 211-223). Dordrecht: Springer.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioural and Brain Sciences , 1*, 515 - 26.

Ratcliffe, M. (2006). 'Folk psychology' is not folk psychology. *Phenomenology and the cognitive sciences , 5*, 31 - 52.

Reddy, V. (2008). *How Infants Know Minds.* Cambridge, Mass: Harvard University Press.

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience , 27*, 169 - 192.

Lavelle, J.S. (2012). 'Theory-theory and the direct perception of mental states'. *Review of Philosophy and Psychology, 3*, 213 – 230.

Shapere, D. (1982). The concept of observation in science and philosophy. *Philosophy of Science 49*, 485 - 525.

Southgate, V., Johnson, M., Osbourne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology letters , 5*, 769 - 772.

Van Riel, R. (2008). On how we perceive the social world. *Consciousness and Cognition, 17*, 544 - 552.

Wellman, H. (1992). *The Child's Theory of Mind.* Cambridge, Mass: M.I.T. Press.