The Unobservability Thesis

Overgaard, Søren

Published in:
Synthese - An international journal for epistemology, methodology and philosophy of science

DOI:
10.1007/s11229-015-0804-3

Publication date:
2017

Document version
Early version, also known as pre-print

In *Synthese*, Online First. DOI: 10.1007/s11229-015-0804-3. Please cite the published version.

The Unobservability Thesis

Søren Overgaard

Abstract

The Unobservability Thesis (UT) states that the mental states of other people are unobservable. Both defenders and critics of UT seem to assume that UT has important implications for the mindreading debate. Roughly, the former argue that because UT is true, mindreaders need to infer the mental states of others, while the latter maintain that the falsity of UT makes mindreading inferences redundant. I argue, however, that it is unclear what 'unobservability' means in this context. I outline two possible lines of interpretation of UT, and argue that on one of these, UT has no obvious implications for the mindreading debate. On the other line of interpretation, UT may matter to the mindreading debate, in particular if we think of it as a thesis about the possible contents of perceptual experience. The upshot is that those who believe UT has implications for the mindreading debate need to be more specific about how they understand the thesis.

1. Introduction

Psychologists and philosophers sometimes claim that the mental states of others are 'unobservable'. In a recent overview of research on mindreading (or theory of mind), Alvin Goldman formulates 'the core question in this domain' in the following way:

> How do [people], or their cognitive systems, go about the task of forming beliefs or judgments about others' mental states, states that are not directly observable? (Goldman 2012, 402)[1]

[1] See Saxe, Carey & Kanwisher (2004, 88), and Apperly (2008, 267) for examples of psychologists making similar statements.

I shall refer to the claim in question as the Unobservability Thesis (UT).[2] Some believe the supposed truth of UT has implications for what must be going on when we mindread each other. Perhaps Goldman, by choosing the above formulation, implies a view along these lines. In any case, as we shall see shortly, others are more explicit. Yet other theorists claim that UT is false, and that *this* puts important constraints on how we should think of mindreading. Roughly, the former argue that because UT is true, mindreaders need to infer the mental states of others, while the latter maintain that the falsity of UT makes mindreading inferences redundant, in at least a range of cases. As I argue in this paper, however, it is not clear what 'unobservability' is supposed to mean in this context. I outline two possible lines of interpretations of UT, and argue that on one of these, it has no obvious implications for the mindreading debate whether we endorse or reject UT. On the other line of interpretation, UT may matter to the mindreading debate, depending on how the thesis is specified. I suggest that one useful way to specify UT is to think of it as a thesis about the permissible contents of perceptual experience. I also offer some suggestions as to how we might go about deciding whether UT, thus understood, is true or false. The most important upshot of this paper, however, is that those who believe UT has important implications for the mindreading debate need to be much more specific about how they understand the thesis and its supposed implications.

In the next section, I outline two arguments that seek to show that it is important to the mindreading debate whether we affirm or deny UT, because how we settle this question determines whether or not all mindreading must be inferential. In section 3, I try to specify the relevant notion

[2] Krueger (2012) speaks of 'the unobservability principle', but this is supposed to designate the following conjunction of theses: 'minds are composed of exclusively intracranial phenomena, perceptually inaccessible … to anyone but their owner' (Krueger 2012, 149).

of ‘inferential’. The task of section 4 is to outline two ways of understanding what it might mean to say that the mental states of others are unobservable. In section 5, I proceed to argue that on one of these interpretations of UT, it does not seem to have the implications it is believed to have. Section 6 examines the other way of understanding UT. I suggest that, if UT is conceived of as a thesis about the permissible contents of experience, then it does seem to be important to the mindreading debate whether it is true or false. Finally, in section 7, I briefly summarize the results of the paper.

2. Unobservability and mindreading

As I use the term here, ‘mindreading’ refers to any understanding (or misunderstanding) of other people in terms of their (supposed) mental states. ‘Mental states’ I use in a similarly broad fashion to cover such heterogeneous things as emotions, sensations, perceptions, thoughts, intentions, desires, and beliefs. Thus, detecting anger in someone else’s face, wondering if someone prefers broccoli to cookies, and judging that someone thinks a particular tube contains pencils, are all cases of mindreading, as I understand it.[3] As mentioned in the introduction, a number of theorists seem to maintain that the mental states of other peoples are unobservable, and that this supposed fact puts important constraints on how we must think of what goes on when we mindread. Alan Leslie, for example, expresses UT and its supposed implications as follows:

[3] Some have a narrower understanding of mindreading, but my use accords, I think, with that of Nichols and Stich (2003, 1-2), among others.

> One of the most important powers of the human mind is to conceive of and think about itself and other minds. Because the mental states of others … are completely hidden from the senses, they can only ever be inferred.[4] (Leslie 1987, 139)

Susan Johnson concurs:

> Mental states, and the minds that possess them, are necessarily unobservable constructs that must be inferred by observers rather than perceived directly. (Johnson 2000, 22)[5]

The thought is that if we cannot observe the mental states of others, we must infer their presence somehow. Call the claim that we must infer the mental states of others the Inference Thesis (IT). We can then represent the line of thought in the following valid (modus ponens) argument:

UT

UT → IT

∴IT

[4] Note that, in maintaining that the mental states of others are '*completely*' hidden, Leslie goes beyond a simple rejection of behaviourism. His statement also commits him to rejecting mereological views, according to which some mental states may be composites, parts of which are hidden and other parts of which are straightforwardly perceivable (e.g. Green 2007).

[5] Similar views are found in Ickes (2003, 43) and Epley & Waytz (2009, 499), Tooby and Cosmides (1995, xvii), and Wellman (1990, 107). See also Jacob (2011, 522).

A number of writers have taken issue with this argument.[6] Shaun Gallagher is arguably its most prominent critic. According to Gallagher, in many accounts of mindreading:

> The supposition is precisely that the other person's mental states are hidden away and are therefore not accessible to perception. I cannot see into your mind; hence I have to devise some way of inferring what must be there, based on evidence that is provided by perception. (Gallagher 2008a, 536)

But UT is false: Gallagher 'rejects … the Cartesian idea that other minds are hidden away' (Gallagher, 2008b, 164). So we are not compelled to accept IT. In fact, Gallagher suggests that the falsity of UT gives us reason to think IT is false. As Gallagher expresses it, 'there is no puzzle to solve, no inference to make, since everything is just out there and obvious' (2008b, 165).

At first blush, it might look as if Gallagher commits the fallacy of denying the antecedent. Even if Gallagher is right that UT is false, this is no reason to think IT is. But I think this response misrepresents Gallagher's views. He, and the other dissenting voices, should be understood, rather, as offering an argument of their own. The suggestion is that we only need to make inferences if UT is true; and since it is not, there is no need for inferences. This is in effect a modus tollens:

-UT

IT → UT

∴-IT

[6] For an early critical voice, see Hobson 1991. More recent writers include Ratcliffe 2007, Reddy 2008, and Zahavi 2011.

So we have two arguments to suggest that UT has some importance. If the first argument goes through, it seems we must infer the mental states of others. If the second argument goes through, it follows that we don't need to infer others' mental states: there simply is, as Gallagher puts it, 'no inference to make'.

3. The inference thesis

I take it that IT constrains the possible shape of an acceptable account of mindreading in some way or other. But what those constraints are depends on what, in IT, is meant by 'inference'. My aim in this section is to fix on an interpretation of IT on which IT meets the following three conditions: (i) It is a thesis that defenders of the mentioned modus ponens are likely to accept, and defenders of the mentioned modus tollens are likely to reject. (ii) IT is not obviously true or obviously false. And (iii) IT should be of some potential relevance to the discussion of how we go about reading each other's minds.

On a couple of interpretations, IT fails to meet conditions (i) and (ii). Suppose just any bit of 'cognitive processing' counts as 'inferential' in the sense of IT. In that case it would pull the rug from under the mindreading debate if IT turned out to be false: the debate is, after all, about the precise nature of the processing involved. But surely no one would *deny* that mindreading has to be inferential in *that* way. IT, interpreted in this way, is obviously true.

Going to the opposite extreme, it might be suggested that 'inference' means the same as 'conscious reasoning'. If IT is the claim that we must consciously reason our way to what others are thinking or feeling, then IT places significant constraints on any account of mindreading. But hardly anyone would *affirm* IT understood in *that* way. Surely it is agreed on all sides that we *sometimes* use conscious reasoning to arrive at judgements about the mental states of others, but at

other times arrive at such judgments in a more automatic or spontaneous way. IT, understood in this way, is obviously false.

A less extreme option is to think of IT as involving inferences that people (not their cognitive systems) make, but that are so fast and routine as to not (or no longer) be *conscious*. Understood in this way, IT seems to meet condition (ii). It is not obviously either true or false. Moreover, on this interpretation, IT may perhaps have its advocates.[7] But it is still not clear that it meets condition (i). Presumably, fast and reliable mindreading abilities would have been extremely useful to our hominid ancestors. If so, human evolution might be expected to have furnished us with fast, automatic and mandatory ('modular') systems for at least some basic mindreading tasks, thus relieving *us* of the need to infer the mental states of others in at least in range of cases. Irrespectively of whether or not we think of others' mental states as 'unobservable', such systems could be highly advantageous. Absent a compelling reason to think the unobservability of mental states somehow rules out the evolution of such modular mindreading systems, it is hard to see why anyone would feel the need to insist that mindreading must involve personal-level inferences. In any case, some of the writers whom I have cited as defenders of IT do locate the relevant inference in subpersonal systems, and so would not accept the current interpretation of IT (e.g. Tooby and Cosmides 1995).

Fortunately, a less committal interpretation of IT is available. On this interpretation, IT states the following: mindreading must involve *extra-perceptual cognitive elements*. Conscious reasoning is inferential in this sense. So are the largely non-conscious, habitual inferences that we sometimes make (upon hearing the doorbell, we immediately form the belief that there's someone

[7] Epley and Waytz, for example, suggest that the inferential 'leap from observable behavior to unobservable mental states' that is mindreading requires 'is so common and routine that people often seem unaware that they are making a leap' (2009, 499).

at the door). And so are the 'subpersonal' cognitive processes underpinning much of our conscious and unconscious mental activity.

On this interpretation, then, IT is the following thesis:

> IT: Extra-perceptual cognitive machinery is needed to identify the mental states of others.

By 'identifying the mental states of others' I mean simply classing or categorizing other people as, for example, 'angry', 'intending to φ', or 'desiring a beer'.[8] IT, on the current interpretation, states that this can never be achieved by perceptual cognitive processing on its own.

Is this interpretation too permissive to meet condition (i)? It seems it is not. In fact, Gallagher explicitly characterizes the view to which he takes his 'direct perception' proposal to be opposed in terms of 'posit[ing] something more than a perceptual element as necessary for our ability to understand others', i.e., to '"mindread"' (2008a, 535). Or again, Gallagher takes his opponents to be committed to the view that 'extra-perceptual cognitive elements' are 'required' for mindreading (Gallagher 2008a, 536). Gallagher's own position, by contrast, is that perception can be 'smart enough on its own … to deliver some sense that [another] person is … angry and motivated to walk away' (ibid.).

Does this interpretation trivialize the debate between Gallagher and his opponents, by rendering IT obviously true or false? It would if we had opted for a narrow view of what qualifies as mindreading. Thus, if one thinks of mindreading as necessarily involving an explicit attribution

[8] 'Identify' is (I suppose) normally used as a success term. As I use it here, however, it includes wrongly classifying someone as being angry. What matters in the present context is not whether or not a mental state attribution is correct, but whether it must be underpinned by extra-perceptual cognitive elements.

of a mental state to a person, then it is trivially true that more than perception is required.[9] At least that is the case if, as seems natural, one thinks of such attributions as judgments, or as the formations of beliefs. To perceive is not (yet) to make a judgment or form a belief, though we usually believe what we see, hear and so on. To make a judgement, then, must involve something else, or more, than perception. So, on this narrow understanding of mindreading, it is trivially true that it must involve non-perceptual cognitive resources.

But given my broader notion of mindreading, no such conclusion can be drawn. It might be, for example, that identifying another person's anger – or categorizing another person as angry – could in some cases be a purely perceptual achievement, which we might (or, e.g. in the theatre, might not) subsequently endorse in judgements. Conversely, it might also be that non-perceptual cognitive elements are required for any mentalistic (mis-)understanding of others. We have so far left it entirely open which of these views is the correct one.

The third condition is trickier. No doubt one might question whether the current interpretation of IT meets condition (iii) for a number of different reasons. Since I want to move on to a discussion of UT, I will not attempt to catalogue and respond to the various possible sources of scepticism. But I will say this. It is at the very least not obvious that the question of whether some

[9] I must note two complications here. First, Gallagher in fact tends to associate the term 'mindreading' with the making of judgments (as does Goldman; cf. the quote in the introduction). Consequently, Gallagher tends not to present his proposal as offering an account of how we sometimes mindread, but rather as explaining why we often don't have to. However, nothing important for what follows hangs on my more liberal use of the term 'mindreading'. If one prefers the narrower use, one can just read 'understanding in terms of mental states' whenever I write 'mindreading'. Second, Gallagher also tends to use the term 'mental state' in a more restricted way than I do. He seems to use it to refer to what I call mental states, but thought of in a particular way, namely as 'hidden'. Again, nothing important hangs on my more liberal use of the term. Anyone who prefers Gallagher's narrower use can read 'understanding in terms of [emotion, sensation, intention, belief, or…]' whenever I write 'mindreading'.

mindreading could be a purely perceptual achievement is without relevance to the mindreading debate. Vision scientists seem to grant that some classification of objects into known categories (e.g. 'dog', 'house') is a visual and hence perceptual achievement (Palmer 1999, ch. 9). If so, it does seem legitimate to ask whether some classification into mental categories (e.g. 'anger', 'fear') might also be perceptual. This may be partly a terminological issue, but it is not clear that it can only be settled by arbitrary choice. Presumably, considerations of speed, automaticity, and the degree to which such classifications are mandatory bear on the issue.

Anyone who remains sceptical about whether IT, as I have interpreted it, has any significant implications for the mindreading debate can regard this paper as being concerned with a hypothetical question: if IT *did* matter to the mindreading debate, would the truth or falsity of UT matter? The answer to this question depends on the plausibility of the two mentioned conditionals: UT → IT and IT → UT. And before we can gauge their plausibility, we need to know what it means to claim that the mental states of others are *unobservable*. This is the topic of the next section.

4. Two notions of (un)observability

'Observable' is clearly a relational property. Something can be (or fail to be) observable only in relation to a potential observer or group of potential observers. Many mammals, such as dogs, bats and whales can pick up high-frequency sounds – ultrasounds – human beings cannot hear. Ultrasound is unobservable relative to us (at least by unaided audition), but not relative to bats. When theorists are debating whether the mental states of other people are unobservable, the question is whether they are unobservable to adult human beings with normal mindreading skills.

It is not clear what it means to say of mental states that they are unobservable relative to us. In this section, I outline two different things that might be meant by this.[10] In order to do so, I first have to consider two different things one might mean by saying that somebody sees something.

Again, part of my task is to think of observation and observability in such a way as not to trivialize this part of the debate between Gallagher and his opponents. Obviously, if UT collapses into IT, both of our conditionals are trivially true, and so there is no interesting question to ask about the relevance of UT to the mindreading debate: I have already granted its relevance. So what we are looking for is an interpretation of UT such that there may be entailment relations between it and IT, but without collapsing the two theses.

When we say that somebody sees *x*, [11] we might mean this to be either a transparent or an opaque context.[12] You might see something – visually pick out some object from its surroundings – without seeing what sort of object it is. You might see a cuttlefish, for example, without seeing that it is a cuttlefish. This is a transparent context: we can substitute terms for what you see and preserve the truth of the statement that you see it.

In other cases when we say someone sees *x*, the context is opaque. This is clearly the case when an abstract noun is inserted in *x*'s place. It is not sufficient for seeing *the problem*, for example, that one sees the thing that is the problem. Suppose the problem is the clogged drain, and

[10] I do not claim that these exhaust the options. But they are sufficient to establish that there is more than one way to think of (un)observability, and hence, depending on the way one thinks about it, UT may or may not have clear implications for the mindreading debate.

[11] For the sake of simplicity, I will restrict my discussion to the visual modality. If there is such a thing as observing others' mental states, the auditory modality is likely to be as important as the visual. I believe the points I will make also hold (*mutatis mutandis*) for that modality. As for the other modalities, my guess is the tactile may also be of some importance, while the olfactory (notwithstanding talk of 'smelling fear') and the gustatory play no significant role.

[12] See Jackson (1977, ch. 7) and McNeill (2012).

suppose it is true that you see the clogged drain. It still need not be true that you see the problem. For that, you need to see *that* the drain is clogged (Dretske 2000, 118).

The distinction also seems to apply to at least some properties of objects. Suppose Jack, who has no knowledge of geometry, looks at pentagon-shaped object. If Jack has normal vision, there is clearly a sense in which he sees the shape of the object – that is why it is possible to exploit this situation to teach Jack a new geometrical concept. If we resist the idea that Jack sees the shape, this is because we understand this as an opaque context: for Jack to see the shape (pentagon), he must somehow be aware of the thing *as* pentagon-shaped, be aware *that* the thing is a pentagon. The same goes for an event, such as that of a cuttlefish emerging from behind a rock. Jill can see that event even if she knows nothing about marine biology. But she cannot see *that* the cuttlefish emerges from behind the rock unless she is able to identify the creature as a cuttlefish.

Following Dretske (2000), call seeing *x* in the transparent sense 'simple seeing' or 's-seeing'. And call seeing *x* in the opaque sense 'seeing-that' ('t-seeing').[13] What I now want to suggest is that the distinction between 's-seeing' and 't-seeing' corresponds to a distinction in the way philosophers sometimes use the term 'observing'.

Robert Brandom discusses the case of a particle physicist reporting the presence of mu-mesons in situations involving hooked vapour trails in bubble chambers. According to Brandom, such a physicist may literally observe mu-mesons:

[13] I wish to avoid the labels 'non-epistemic' and 'epistemic' seeing, as my topic is not an epistemological one. It is thus not important to my notion of t-seeing whether the seen 'fact' obtains or not. If it makes sense to think of 'seeing-as' as a non-factive form of seeing-that (cf. Smith 2015), then seeing-as is the relevant notion of t-seeing in the present context. Moreover, if we want to think of the debate between Gallagher and his opponents as a dispute over whether the mental states of others are t-observable, we must not require that the subject believes what she seems to see. I will return to the latter point in section 6.

> … coming to be disposed reliably to respond to the vapor trail, and *hence* to the presence of mu-mesons, by asserting or acknowledging a commitment to the presence of a mu-meson is learning to *observe* mu-mesons, to report them *non*inferentially. (Brandom 1994, 223)

Brandom thinks inferences from observations of vapour trails to the presence of mu-mesons may be 'part of the training process that leads to becoming a reliable observer of mu-mesons (in bubble chambers)' (ibid.). The novice by the physicist's side, then, does not observe mu-mesons; only the specialist does, though they both observe the vapour trail. If so, there is no transparent observing of mu-mesons, no 's-seeing' them. Things that would be observable in the sense of s-seeing would be things that one could find oneself confronted with, without knowing what they were. Our situation with respect to mu-mesons, and elementary particles in general, is not like that. One does not find oneself confronted with a mu-meson or a positron, saying to oneself, 'What on earth is that?' 'Observing mu-mesons', then, is really shorthand for the observation of certain facts involving mu-mesons, then, such as their presence in the bubble chamber.

While Brandom understands 'observing *x*' in terms of seeing *that x* is *F*, van Fraassen seems to understand 'observing *x*' in terms of 's-seeing' *x*. He stresses the importance of not confusing '*observing* (an entity, such as a thing, event or process) and *observing that* (something or other is the case)' (van Fraassen 1980, 15). He offers the following illustration:

> Suppose one of the Stone Age people recently found in the Philippines is shown a tennis ball or a car crash. From his behaviour, we see that he has noticed them; for example, he picks up the ball and throws it. But he has not seen *that* it is a tennis ball, or *that* some event is a car crash, for he does not even have those concepts. He cannot

> get that information through perception; he would first have to learn a great deal. To say that he does not see the same things and events as we do, however, is just silly; (ibid.)

On van Fraassen's notion of observing *x*, then, it is sufficient that an observer perceptually picks out *x*, differentiates it from its immediate surroundings. But this is also necessary: an entity that is not picked out in this simple way is not observed, even if we may observe certain facts involving it, e.g. that it is present in the bubble chamber.

So have two notions of observation, corresponding to our two notions of seeing: Observing *x*, and observing that *x* is *F*. This gives us two corresponding notions of the observability of something, *x*. We might have in mind *x*'s s-observability: *x* is the sort of thing a normal human being may visually (or in some other sensory modality) stumble upon, irrespectively of the person's conceptual or recognitional capacities with respect to things of *x*'s kind. We might also have in mind the t-observability of certain facts about *x*, such as its presence in the bubble chamber.

Now we can ask: when some theorists assert, and others deny, that the mental states of others are observable, which of our two notions of observability is in play? In other words, is UT the claim that there is no such thing as s-seeing mental 'things' (Jack's anger) or the claim that there is no such thing as seeing mental 'facts' (that Jack is angry)? In the next section, I argue that it had better not be the former.

5. *Simple unobservability*

Call the Unobservability Thesis, understood in terms of s-observation, UT_s. According to UT_s, mental states just are not the sort of thing that can be s-seen (transparently seen) the way tennis balls and car crashes can.

Here are the two conditionals connecting UT_s and IT:

$UT_s \rightarrow$ IT:

If we cannot *s-see* the mental states of others, then we need extra-perceptual cognitive machinery to identify them

IT $\rightarrow UT_s$

If we need extra-perceptual cognitive machinery to identify the mental states of others, then we cannot *s-see* them

Unfortunately, neither conditional is very plausible.

Let me first consider $UT_s \rightarrow$ IT. Just to get us started, here is a bit of philosophical science fiction due to Dretske (1973). Mars and Earth are at war. Martians happen to look just like ordinary human beings. Unlike us, however, they have the ability to make themselves invisible, but at a cost: when they are invisible, they are also strongly magnetic. One thing that is immediately evident is that it will be much easier to identify an invisible Martian than a visible one. The latter is likely to be mistaken for an Earthling, whereas the former's presence is revealed by the paperclips and cutlery that outline a moving, human-like gestalt.

I do not think this example, as it stands, drives the point home. Unless these wars have been going on for a very long time, and people have gotten as used to spotting invisible Martians as we are used to seeing dogs, it seems highly likely that extra-perceptual cognition is needed to identify an invisible Martian. But we can at least imagine that over time, given the right circumstances, categorizing moving clusters of metallic objects as Martians might become as much a visual affair as categorizing certain animals as dogs might perhaps already be. So the example

ought to cast doubt on the idea that if something is s-unobservable, we need extra-perceptual cognitive elements to identify it.

Some less speculative examples reinforce the point. One of van Fraassen's examples of unobservables is light (van Fraassen 2001, 152). Illuminated objects and surfaces are of course paradigmatic cases of observables, but that does not mean that light itself is observable. We can see light beams, but only in dirty air; what we see are illuminated specks of dust and the like. Van Fraassen mentions a demonstration devised by physicist Arthur Zajonc, in which a light shines directly into a box without illuminating either the sides of the box or any object inside the box. The box is fitted with a viewport, but if one looks into it one sees only darkness. The moment an object is moved into view inside the box, however, it appears brightly illuminated (ibid.). This is supposed to show that light is not s-observable, only illuminated objects are. If this is true, the example of light effectively undermines the idea that if something is s-unobservable, then we need extra-perceptual cognitive resources to identify it. In normal situations, we do not have to go beyond (basic) vision to register the presence of light. Its presence is immediately obvious from the illuminated surfaces.

Wind might also be a case in point. Wind is moving air. Presumably, moving air cannot be s-seen. Suppose part of the air in front of you is moving, and part of it remains still. Can you visually pick out the moving part? Presumably not. If not, it seems wind cannot be s-seen.[14] Yet wind might be quite easily identified by visual means alone: you see the wind shaking the branches of trees, whirling loose leaves around, and so on. Classifying a situation involving loose leaves

[14] One can imagine the moving mass of air having dyed dust particles. In this sort of case, it would (I take it) be possible to pick out the moving mass of air from the stationary one. But one would do so by s-seeing the moving dust particles, not the moving air as such. In other words, this case is in principle similar to detecting wind by s-seeing moving leaves.

whirling around as one involving wind could be a purely perceptual achievement, even if wind itself is not s-observable.

Collectively, these examples suggest that even if Jack's anger is not the sort of thing one might find oneself visually confronting without knowing what it is (so even if UT_s is true), it does not follow that one needs extra-perceptual cognitive elements to identify Jack's anger. $UT_s \rightarrow$ IT, then, is at the very least doubtful.

Some of the examples we have considered in this and the previous sections already cast doubt on the other conditional: IT $\rightarrow UT_s$. Presumably, we would need to employ cognitive resources that go beyond simple perception to identify those *visible* Martians. As easy as it is to pick them out visually from their environment, as difficult it is to tell, simply by looking, that they are Martians. Bas van Fraassen's Stone Age person needs extensive instruction before he can make out what he is confronted with when he sees a car crash; yet he clearly observes – s-sees – that event.

In general, the fact that something is s-observable tells us nothing about the cognitive processes that are involved in identifying it as the sort of thing it is. S-observability just means that the thing, which is actually a car crash, a colour, a cuttlefish, or (perhaps) an emotion or an intention, can be perceptually picked out from its immediate surroundings. It does not tell us anything about how observers (or their cognitive systems) manage to make sense of what is s-observed.

Hence the rejection of UT_s only entails that the mental states of others may be s-observable. This leaves it open how observers are able to identify what they see *as* the mental state of another person. Nothing has been said about whether or not we 'can get that information through perception' alone, or whether extra-perceptual cognitive resources must be involved. Clearly, then, IT $\rightarrow UT_s$ is false.

I conclude that on the current interpretation of UT, it has no obvious implications for the mindreading debate whether the thesis is true or false. Contra the views of Leslie, Johnson and others, we can affirm UT_s without being committed to IT. And contra critics such as Gallagher, rejecting UT_s does nothing to render IT implausible. But perhaps talk about mental states being (or not being) 'unobservable' is not to be interpreted in terms of s-observability, but in terms of t-observability. Perhaps the question of whether Jack's anger is observable is really the question of whether one can observe *that* Jack is angry. I consider this possibility in the next section.

6. *Unobservability and perceptual content*

Call the Unobservability Thesis, understood in terms of observing-that, UT_t. According to UT_t, we cannot see facts involving the mental states of others the way can see physical facts. I can see that Jack is bald or tall, but I cannot see that he is angry or that he is thinking about football. Here are the two conditionals connecting UT_t and IT:

> $UT_t \rightarrow$ IT:
>
> If we cannot *t-see* the mental states of others, then we need extra-perceptual cognitive machinery to identify them

> IT $\rightarrow UT_t$:
>
> If we need extra-perceptual cognitive machinery to identify the mental states of others, then we cannot *t-see* them

The problem here is not that these conditionals are implausible. Rather, the problem – or better: challenge – is to specify the relevant notion of t-observation without trivializing the debate between

Gallagher and his opponents. I want, very briefly, to indicate two ways in which one might trivialize the debate, before suggesting a way to articulate the relevant notion of t-observation that steers clear of this problem. (There are no doubt other ways of both trivializing the debate and rescuing it than the ones I will be considering.)

First of all, to repeat a point made previously, the relevant notion of seeing-that must not involve believing or judging. Normally, we would not say of a person that she sees that Jack is bald unless she believes him to be. But if seeing-that involves believing or judging, then it involves something that requires us to posit extra-perceptual cognitive elements. To think of t-observing in this way, then, would make Gallagher's conditional – IT → UT_t – trivially false. To see this, suppose Gallagher is right that UT_t is false. That is, suppose we *can* t-observe the mental states of others. On the current understanding of it, t-observing another's mental state would, by definition, involve extra-perceptual cognition, because it involves believing or judging that the other is in that state. So unless we had other perceptual ways of mindreading besides t-observing, IT would be true.

We need a notion of seeing-that that does not (yet) involve believing or judging. The relevant notion of seeing-that can be loosely characterized in terms of it perceptually seeming to someone as if something is the case. This perceptual seeming would then be one the subject could endorse (or refuse to endorse) in belief and judgement.[15] But what does it mean to say the seeming is 'perceptual'? This had better not mean that t-observing is a purely perceptual achievement, for then we would collapse UT into IT. If UT_t is the thesis that mindreading cannot be a purely perceptual achievement, then there is no daylight between UT_t and IT. And this would make the two conditionals linking UT_t and IT entirely vacuous. It would hence be trivially true that UT_t has implications for the mindreading debate.

[15] Again, perhaps 'seeing-as' fits the bill here. I see one Müller-Lyer line *as* longer than the other, but I don't believe it to be.

Fortunately, there are ways of making the new interpretation more precise. One way is to think of it as making a claim about the possible representational *contents* of perceptual experience. For present purposes, we can understand this talk of 'representational content' simply in terms of the accuracy (or veridicality) conditions of an experience (See Siegel 2010, ch. 2).[16] So, roughly, if I have a perceptual experience as of something round and red, for example, that experience is veridical if and only if the thing (if any) I see is round and red. My experience, we can say, attributes certain colour and shape properties to an object, and the experience is only veridical if the object has those properties.

We could put the same point in terms of the non-factive notion of identification that I introduced above. [17] My experience *identifies* (categorizes or classes) an object as being round and red. If the thing is not round, the experience is illusory with respect to the object's shape; if there is no candidate object there, the experience is completely hallucinatory, and so on. Now, if we think of mindreading as involving the attribution of a mental state to something (typically a person) – that is to say, as identifying someone as having a mental state – we can express the new interpretation of UT_t as follows:

> UT_{tc}: Perceptual experience cannot have a content that identifies another person's mental state

[16] I am sidestepping a host of difficult issues here. Philosophers disagree about whether or not the contents of perceptual experiences are (or may be) object-dependent, singular, non-conceptual, rich (see below), and so on and so forth. In fact, although most philosophers think perceptual experiences have representational content, some variants of disjunctivism reject this idea altogether (See e.g. Brewer 2011; Travis 2004). Although he is sympathetic to the general idea of perceptual content, Pautz (2009) raises some worries about common ways of conceiving of the contents of experience.

[17] See section 3, especially footnote 8.

This gives us the following two conditionals:

UT_{tc} → IT:

If perceptual experience cannot have a content that identifies another person's mental state, then we need extra-perceptual cognitive machinery to identify the mental states of others

IT → UT_{tc}

If we need extra-perceptual cognitive machinery to identify the mental states of others, then perceptual experience cannot have a content that identifies another person's mental state

Given these formulations, someone might worry that UT_{tc} collapses into IT. This worry can be put to rest, however. Consider the qualitative or phenomenal character of a perceptual experience – 'what it is like' to have that experience. It is controversial whether the phenomenal character of an experience can be reduced to its representational content. That is, it is not obvious that the phenomenal character boils down to an aspect of how the experience represents the world as being. While some philosophers do think that phenomenal character either reduces to, or 'supervenes' upon, representational content (e.g. Harman 1990; Tye 2000), this is a controversial view.

Suppose I have just taken off my glasses. Everything is blurred. Is 'blurriness' a property my experience attributes to the coffee mug and laptop before me, so that the experience is illusory to the extent that these objects have no such property? That is at the very least not obvious

(Crane 2006). As we might put it, when I am having this experience, it does not seem to me that the mug is blurred, but only that I see it 'blurrily'. Suppose we grant that this is the right way to look at the case: blurriness is not a feature of how the experience represents the world as being. Does it follow that I need to employ extra-perceptual cognitive resources to identify the blurriness? Surely not. My experience without my glasses immediately gives itself as less clear and distinct than my experience before I took them off. This is a purely visual matter if anything is.

Arguably, then, a perceptual experience may have qualitative features over and above its representational content; yet identifying those features need not take us beyond perception. If so, it cannot be maintained that UT_{tc} reduces to IT. To say that others' mental states cannot be represented in perceptual experience is not yet to say that we need extra-perceptual cognitive machinery to identify those states.

While these considerations suggest that UT_{tc} does not collapse into IT, they also seem to suggest a possible objection to $UT_{tc} \rightarrow$ IT. We have raised doubts about the *general* claim that if something is not (or cannot be) represented in the representational content of the experience, then we need extra-perceptual cognitive resources to identify that something. Applied to the case of mindreading, it might be suggested, these considerations undermine $UT_{tc} \rightarrow$ IT.

However, such a conclusion seems premature. For intuitively, it is not easy to see how mindreading could be a purely perceptual achievement if others' mental states could not be represented in perceptual experience. If Jack's anger, or his being angry, cannot be part of the way my perception represents Jack to me, then intuitively, it seems as if we are going to need more than perception to get Jack's anger into the picture. (Certainly, *Jack's anger* cannot be a non-representational qualitative feature of *my* experience.) $UT_{tc} \rightarrow$ IT, then, seems plausible.

Interestingly, IT $\rightarrow UT_{tc}$ seems plausible too. If Jill's being afraid is part of the content of an experience, then presumably no extra-perceptual processing is needed to identify it.

Perception takes us all the way there; it already classifies Jill as afraid. Thus, if extra-perceptual processing *is* needed to identify her fear, then her being afraid cannot be part of the content of the experience.

It seems, then, that UT_{tc} is of some potential importance to the mindreading debate, at least to the extent that IT is. This would be one reason to think it important to settle whether or not UT_{tc} is true.[18] If it is true, it seems likely that mindreading must involve extra-perceptual cognition, precisely as Leslie and others have claimed. If it is false, it seems Gallagher may be right in affirming that perception alone can sometimes do the job. If the question of whether that is so is important to the mindreading debate, then so is the question concerning the truth of UT_{tc}. Thus, one way to rescue the debate between Gallagher and his opponents over the supposed unobservability of the mental – not necessarily the only way – is to think of it as a debate about the permissible contents of perceptual experience.[19]

Is UT_{tc} true? I cannot settle that question in this paper. However, before concluding, I will briefly outline a way in which one might go about settling it, viz. Siegel's (2010, ch. 3) 'method of phenomenal contrast'. Suppose we have a pair of cases involving experiences of Jack's scowling, flushed face and clenched fists such that one case (A), but not the other (B), involves recognizing that Jack is angry. It should be uncontroversial to hold that the two *overall* experiences, or total experiential situations, would differ phenomenally: 'what it is like' to see Jack's scowling etc. when one recognizes that he is angry is not the same as what it is like to see his scowling when

[18] Silins (2010) discusses the epistemological significance of the debate about the scope of perceptual content.

[19] If we think about the unobservability thesis in this way, it is not only of significance to the mindreading debate. In the philosophy of perception, there is currently much discussion about the permissible contents of perception. The main focus has been on the prospects of including natural kind properties in the content of experience, but causal and mental properties, too, have been mentioned as potential candidates for inclusion. See Bayne (2009) and Masrour (2011) for permissive (or 'rich', or 'thick') views, and Brogaard (2013) for a more restrictive ('poor', 'thin') view.

one is oblivious of his anger. The method of 'phenomenal contrast' consists in evaluating whether the best way to explain this contrast is by denying UT_{tc} – i.e. maintaining that Jack's being angry may be represented in the content of the visual experience in (A) – or whether there are rival hypotheses that fare better.

Siegel suggests that there are three main rivals to the thesis that explains the phenomenal contrast in terms of Jack's anger being represented in the content of the perceptual experience in (A), but not in (B) (see Siegel 2010, 90-96). First of all, one could deny that the phenomenal difference between the overall experiences in (A) and (B) has anything to do with a phenomenal difference in the *perceptual* experiences involved. Instead one might insist that the difference is due to a phenomenally conscious non-sensory state that is present in (A) but not in (B) – for example, a conscious thought or judgement that Jack is angry. Alternatively, one might grant that the difference is a difference in perceptual phenomenology, but deny that it is a difference that should be located in the contents of the contrasting experiences. Perhaps the visual experience in (A) has a *non-representational* phenomenal feature – a 'raw feel' or '*quale*' – that is absent from the visual experience in (B). Finally, one could accept that the phenomenal contrast is indeed due to a difference in the representational contents of the perceptual experiences involved in the two cases, but deny that the difference in question has anything to do with anger being attributed to Jack in either of the two perceptual experiences. Instead, one could maintain that the two perceptual experiences in (A) and (B) differ in terms of the *non-mental* properties or states they represent. [20]

[20] Spaulding (2015) argues that phenomenal contrast arguments fail because they do not distinguish between 'causally relevant and constitutive properties of perception' (ibid.). Spaulding contrasts the experiences of an expert art historian and a novice looking at the same impressionist painting. The undeniable phenomenal difference between their experiences, she claims, could be due to high-level properties (e.g. 'impressionist') in the content of the expert's, but not the novice's, perceptual experience. But it could also be that the expert's knowledge of art causally influences which features of the work she attends to, or finds interesting, or her expectation about the work. Consider the

Before concluding, let me briefly note some *prima facie* worries about (simple versions of) the alternatives to abandoning UT_{tc}. First of all, suppose the subject in (A) is convinced that Jack is not really angry at all. It seems possible that the overall experience of Jack in (A) could nevertheless still differ phenomenally from the experience in (B). And if so, it cannot be the presence of a conscious thought or judgment that Jack is angry that explains the difference. Secondly, it is not entirely clear how the 'raw feels' proposal can be made plausible. Some feeling of 'recognition' or 'familiarity' is supposed to be part of the perceptual experience of Jack in (A), but it is not a representational feature of that experience – nothing is being represented *as* familiar.

---

explanation in terms of attentional differences. The expert's knowledge may lead her to focus her attention on the special character of the brush strokes (say), which renders her experience different from that of the novice (whose attention might be grabbed by the depicted scene, say). But if this is the sort of thing Spaulding has in mind, then it is not clear that her criticism affects the method of phenomenal contrast as outlined above. For the obvious question to ask is *how*, precisely, the two experiences are supposed to differ phenomenally. The obvious candidates seem to be precisely the options Siegel highlights. Either the two experiences differ in terms of low-level contents (one, but not the other, represents short and broken brush strokes, say). Or else they differ in terms of non-representational features. Or, finally, the difference is in terms of non-sensory states. (Similar points apply to explanations in terms of different interests or expectations. The latter, for example, seems most naturally thought of in terms of different non-sensory states.) It is of course possible that Siegel's alternatives do not exhaust the options; but for all Spaulding has shown, they may well be exhaustive.

Spaulding goes on to argue that the best explanation of the phenomenal contrast between the two experiences is one that refers simply to differences in attention. She concedes that this second argument presupposes the controversial view that attention cannot itself be a constitutive element of perception (see Mole [forthcoming] for arguments against this view), but she maintains that the first argument stands independently of the second. The reverse is not the case, however. If Spaulding's explanation in terms of differences in attention is captured by one of Siegel's options, then Spaulding's second argument does not target the contrast method as such. Rather, it amounts to the suggestion that, at least in the particular case of the impressionist painting, the contrast method leads to a result that does not involve high-level content.

But, one might suggest, this corresponds to nothing in our experience. Even a *déjà vu* is the experience of a certain place or scene as familiar (Siegel 2010, 110). Finally, there is the idea that the phenomenal contrast between the overall experiences in (A) and (B) is due to a difference in the non-mental properties represented in the perceptual experiences involved. An initial worry here is that it seems that someone – say, a high-functioning autistic person – could have a perceptual experience representing the exact same non-mental properties (e.g., shapes, colours, and movements) that are supposed to be distinctive of the perceptual experience in (A), but without recognizing Jack as angry. And intuitively, that person's overall experience would still contrast phenomenally with the overall experience in (A). Conversely, it seems someone could have an experience with the non-mental representational content supposed to be distinctive of the perceptual experience in (B), yet recognize Jack as angry. Intuitively, such a person's experience would still differ phenomenally from the experience in (B). If this is right, then it seems unlikely that the contrast between (A) and (B) is exclusively a matter of differences in the non-mental properties represented in the perceptual experiences involved.

If worries of these or similar sorts prove well-founded, and if Siegel's four strategies exhaust the options, then UT_{tc} seems to be in trouble.[21] And if UT_{tc} is problematic, it looks as if Gallagher is right to deny that mindreading must involve extra-perceptual cognitive elements. It is, however, beyond the scope of this paper to attempt to settle these issues.

7. *Conclusion*

I have distinguished two ways of interpreting the Unobservability Thesis. On one interpretation, the thesis states that the mental states of others are s-unobservable. That is, they are not the sort of thing

[21] For a more fully developed discussion of the various explanatory strategies and their respective merits – focusing on 'kind properties' rather than mental states or properties – see Siegel (2010, ch. 4).

one may perceptually stumble upon, not knowing what one is seeing. I argued that, if this is the correct articulation of UT, the thesis has no clear implications for the mindreading debate. In particular, one can deny that Jack's anger is observable in this sense and still think of some mindreading as non-inferential (i.e., as involving no extra-perceptual cognitive steps); and one can affirm that Jack's anger is s-observable and still maintain that we need more than perception to identify it as anger.

On the other line of interpretation, what UT states is that it is not possible to observe *that* Jack is angry or Jill is scared. I suggested that if this we thought of this as a claim about the possible contents of perceptual experience, we could preserve the relevance of UT for the mindreading debate. If Jill's being afraid cannot be part of the way my experience represents Jill, then it does seem as though I will need more than perception to mindread her. On the other hand, if her being afraid *can* be among the contents of my perceptual experience, then it seems mindreading could be a purely perceptual achievement. I have suggested a way in which one may go about determining whether that is indeed so, and I have sketched some *prima facie* reasons for scepticism about UT, understood as a thesis about the possible contents of perceptual experience.

The most important upshot of this paper, however, is the following: those who claim UT has important implications for the mindreading debate need to specify what they mean by '(un)observability'. Depending on what this means, UT may or may not be a very interesting or important thesis as all. [22]

References

[22] Previous versions of this paper were presented at conferences in Copenhagen, Vienna, Bochum, and Kirchberg am Wechsel. I am grateful to all audiences for helpful comments. I owe a particular debt of gratitude to two referees at *Synthese* for pressing me to make a number of significant improvements to the paper.

Apperly, I. (2008). Beyond simulation-theory and theory-theory. *Cognition* 107: 266-283.

Bayne, T. (2009). Perception and the reach of phenomenal content. *The Philosophical Quarterly* 59: 385-404.

Brandom, R. (1994). *Making it Explicit*. Cambridge, MA: Harvard University Press.

Brewer, B. (2011). *Perception and its Objects*. Oxford: Oxford University Press.

Brogaard, B. (2013). Do we perceive natural kind properties? *Philosophical Studies* 162: 35-42.

Crane, T. (2006). Is there a perceptual relation? In T. S. Gendler and J. Hawthorne (eds.), *Perceptual Experience* (pp. 126-146). Oxford: Oxford University Press,

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

Dretske, F. (1973). Perception and other minds. *Noûs* 7: 34-44.

Dretske, F. (2000). *Perception, Knowledge and Belief.* Cambridge: Cambridge University Press.

Epley, N., and Waytz, A. (2009). Mind perception. In S. Fiske, D. Gilbert, and G. Lindzey (eds.), *The Handbook of Social Psychology*, 5th edn. (pp. 498-541). New York: Wiley.

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

Gallagher, S. (2008b). Inference or interaction: social cognition without precursors. *Philosophical Explorations*, 11, 163-174.

Goldman, A. I. (2012) Theory of mind. In E. Margolis, R. Samuels, and S. P. Stich (eds.), *The Oxford Handbook of Philosophy of Cognitive Science* (pp. 402-424). Oxford: Oxford University Press.

Green, M. (2007). *Self-Expression*. Oxford: Clarendon Press.

Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives* 4: 31-52.

Hobson, R. P. (1991). Against the theory of 'theory of mind'. *British Journal of Developmental Psychology* 9: 33-51.

Ickes, W. (2003). *Everyday Mind Reading*. Amherst: Prometheus Books.

Jackson, F. (1977). *Perception*. Cambridge: Cambridge University Press.

Jacob, P. (2011). The direct-perception model of empathy: a critique. *Review of Philosophy and Psychology* 2: 519-540.

Johnson, S. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences* 4: 22-28.

Krueger, J. (2012). Seeing mind in action. *Phenomenology and the Cognitive Sciences* 11: 149-173.

Leslie, A. M. (1987). Children's understanding of the mental world. In R. L. Gregory (ed.), *The Oxford Companion to the Mind* (pp.139-142). Oxford: Oxford University Press.

Masrour, F. (2011). Is perceptual phenomenology thin? *Philosophy and Phenomenological Research* 83: 366-397.

McNeill. W. E. S. (2012). Embodiment and the perceptual hypothesis. *The Philosophical Quarterly* 62: 569–591.

Mole, C. (Forthcoming). Attention and cognitive penetration. In J. Zembeikis and T. Raftopoulos (eds.), *Cognitive Penetration*. Oxford: Oxford University Press.

Nichols, S. and Stich, S. P. (2003). *Mindreading*. Oxford: Clarendon Press.

Palmer, S. E. (1999). *Vision Science: Protons to Phenomenology*. Cambridge, MA: MIT Press.

Pautz, A. (2009). What are the contents of perceptual experiences? *The Philosophical Quarterly* 59: 483-507.

Ratcliffe, M. (2007). *Rethinking Commonsense Psychology*. Basingstoke: Palgrave Macmillan.

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

Saxe, R., Carey, S., and Kanwisher, N. (2004). Understanding other minds: Linking developmental

psychology and functional neuroimaging. *Annual Review of Psychology* 55: 87-124.

Siegel, S. (2010). *The Contents of Visual Experience*. New York: Oxford University Press.

Silins, S. (2013). The significance of high-level content. *Philosophical Studies* 162: 13-33.

Smith, J. (2015). The phenomenology of face-to-face mindreading. *Philosophy and Phenomenological Research* 90: 274-293.

Spaulding, S. (2015). On direct social perception. *Consciousness and Cognition*. http://dx.doi.org/10.1016/j.concog.2015.01.003

Tooby, J. and Cosmides, L. (1995). Foreword. In S. Baron-Cohen, *Mindblindness*. Cambridge, MA: The MIT Press.

Travis, C. (2004) The silence of the senses. *Mind* 113: 57-94.

Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge, MA: The MIT Press.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Oxford University Press.

van Fraassen, B. C. (2001). Constructive empiricism now. *Philosophical Studies* 106: 151-170.

Wellman, H. M. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.

Zahavi, D. (2011). Empathy and direct social perception: a phenomenological proposal. *Review of Philosophy and Psychology* 2: 541-558.