CHAPTER 27

Knowledge of other minds

Chapter contents

So far the chapters in this part of the book have focused on the nature of the mind and mental states of central importance to psychiatry. The questions that they have addressed have included:

◆ What are mental states and how are they related to physical states including brain states?

◆ How are mental states *about* anything?

◆ In what sense is the aboutness or intentionality of the mind a *natural* property?

◆ What is the connection between thought and language?

◆ Can mental states cause actions?

Empirical versus metaphysical

These are all questions about the nature of mental phenomena, about their ontology. While they are not straightforwardly *empirical* questions we have looked at the bearing that empirical findings, such as the results of brain imaging or theories of mental function from cognitive psychiatry, have on them. However, they have all been questions of ontology, of the nature of what is. By contrast the subject matter of this chapter is the *epistemology* of mind of how we know about our and others' mental states.

Ontology versus epistemology

The focus of this chapter is how we can have *knowledge* of mental states rather than what mental states are. In fact, in the case of the philosophy of mind, the distinction between epistemology and ontology is not sharp. Getting clear on how we find out what we and other people think promises to shed light on the very nature of thought. However, the agenda will be set by following epistemological issues first.

The importance of the issue to mental health care

The question of how we know about the mental states of others is obviously important to an understanding of mental health care. Think, for example, of a clinical encounter between doctor and patient, or social worker and client. One of the key aims of the doctor or social worker is to find out about the state of mind of the patient or client. Now the Present State Examination is one attempt to increase the consistency (or inter-rater reliability) of such encounters. However, it does not raise the deep issue of how it is at all possible to know about someone else's mental states.

There is another source of interest in this question that results from clinical work. One of the central symptoms of autism is an impoverishment of a subject's ability to understand others. So if people who suffer autism have an impairment of an ability that non-sufferers enjoy, this promises to shed light on that ability. Just what is it that autistic people lack? What is the nature of their disability in this area?

Knowledge, access, and the role of justification

Given that we are concerned with epistemology: with how we know about other people's mental states there is a further preliminary point worth making. In general in philosophical discussion, epistemology is mainly concerned with *justification*. It addresses the *pedigree* of our beliefs (about the external world, the future, the past) by asking what justification can be given for them on the long-standing traditional assumption that justification (or something similar) is an element of knowledge. (The traditional thumbnail analysis is that knowledge is *justified true belief*.) This is reflected in a traditional question in the philosophy of mind: How can we *know* the contents of other minds, or know even whether other minds exist?

In this chapter we will be equally concerned with a question that appears to be prior to that of justification: How is it possible to form beliefs about the contents of other people's minds at all? How is any such access possible? In fact the question of access and the question of justification are closely related. As we do take ourselves to be justified in our beliefs about other minds, no account of access that leaves such judgements as having the status of mere guesswork will do. The very idea of talking about access to other minds implies some degree of reliability. Thus explaining how it is possible to form beliefs about mental states must dovetail with an explanation of why the method we generally deploy is justified.

Three main approaches

What gives the question of access to other minds its bite is the persuasive thought that we cannot literally see (into) the minds of others. Only I can directly access my mental states. Other people must *infer* them as I must infer both the existence and the contents of their mental states. On what basis and how can I draw inferences? To anticipate the rest of this chapter, there are currently two main rival responses to this question.

Theory-theory

One account replies that access to other minds is mediated by a *theory*. This takes as its input observable behaviour and, as its output, claims about unobservable mental states. Third-person access is thus akin to our access to the unobservable microstructure of the world provided by the theoretical physical sciences. In recent jargon this is called the 'theory-theory' approach.

Simulation theory

Its main rival is the 'simulation-theory' approach. This is based on a contrasting thought. Instead of deploying a *theoretical* knowledge of the workings of minds, we access other minds by *using* our own mental faculties. We put ourselves in the shoes of other people on the basis of an assessment of their situation, actions, and speech. Thus we simulate the mental states of others. Rather than *knowing* and applying general principles that govern the workings of minds, we simply *use* our own minds of whose working we might be ignorant.

These two sketches maximize the differences between theory-theory and simulation theory. As we will see later in the chapter, they may not be as different as all that. Very briefly: no theory theorist believes that the theory in question is a matter of *explicit* knowledge. The theory is, at most, *tacitly* known and codifies our abilities to ascribe mental states on the basis of behavioural evidence. It resembles the use of linguistic theory, which is

supposed to codify our ability to use and understand grammatical sentences. Such linguistic theories—the stuff of Chomskian linguistics, for example—are not thought of as explicitly known by speakers but merely tacitly known. Likewise in this case the theory of others minds is tacit. But if so the ability it codifies may be an ability to simulate. At the same time, simulation theory requires that the results of the 'thought experiments' on which it depends are taken to be guides to how others think. It is open to question whether this can escape a theoretical inference that you are thinking the same thoughts as those I have just simulated.

What may serve as the clearest difference between these two approaches is the way rationality plays a constitutive role in the mind. But whether that role is best cashed out in either such theory is questionable.

Direct access?

There is also a lesser known third possibility. This is to deny that there is the fundamental epistemological barrier, that both of the preceding views presuppose, to the direct access to other's mental states. Spelling out this third option will require some care because there is something unarguable about the claim that we cannot directly observe other people's mental states. Nevertheless, this need not be seen as requiring the kind of explanatory account at which both the theory-theory and simulation theory aim.

But we will begin not with any of these contemporary views but with a piece of intellectual history.

The plan of the chapter

◆ *Session 1* examines the origin of the Problem of Other Minds in Descartes' account of the mind and the attempt to solve it using the Argument from Analogy.

◆ *Session 2* introduces the 'theory theory' approach.

◆ *Session 3* introduces the rival 'simulation theory' approach which emphasises the role of rationality.

◆ *Session 4* looks at the current state of the debate including the role of evidence from autism.

◆ *Session 5* examines a more direct approach to undermine the Problem of Other Minds.

Session 1 **An historical starting point: Descartes and the argument from analogy**

The origin of the Problem of Other Minds

In the Anglo-American tradition, the philosophy of mind, has until very recently, turned away from epistemological issues and concentrated instead on metaphysical and ontological questions about the nature of mental states. It has focused on the sorts of questions discussed in previous chapters. However, during the first half of the twentieth century, a central 'live' issue in the philosophy of mind was the Problem of Other Minds. How is it possible to account for our knowledge of the mental states of other people or even that they have them?

Like most philosophically sceptical questions, this question requires some initial stage setting for it to seem pressing. After all, it is obvious that in everyday life we routinely make judgements about other people's beliefs, moods, desires, and suchlike. Making arrangements to meet at a specific place and time requires just this sort of understanding. In law courts, evidence can be offered about the state of a person's mind with just as much claim to validity or objectivity as evidence about the state of their bank balance. Many human customs and practices would be impossible if it we were not able reliably to 'read' other people's minds. However, once one has been initiated into a characteristically philosophical way of thinking about the mind, this ability can come to seem mysterious or illusory.

Historically, the express route to the Problem of Other Minds was one or other form of Cartesian dualism. Descartes' arguments for a dualism of thinking non-extended stuff and extended non-thinking matter were touched on in Chapter 22. However, without thinking of the details of his argument—an argument whose validity is open to question—it is worth recalling generally its presuppositions.

EXERCISE 1   (30 minutes)

Read the short extract from the start of:

> Descartes, R. ([1641] 1996). *Meditations on First Philosophy*. Cambridge: Cambridge University Press. (Extract: pp. 12–13.)

Link with Reading 27.1

◆ What is striking about the approach that Descartes adopts to doing philosophy?

◆ How might this affect his account of the mind?

The solitary starting point

Descartes' project is one that, he suggests, should be undertaken once in his life. He should examine his beliefs in order to establish some foundation for them. Thus he begins: '*I am here quite alone*, and at last will devote myself sincerely and without reservation to the general demolition of my opinions.' (p. 12, italics added)

What is striking about this starting point is its solitariness. In general we take the acquisition of knowledge to be a social endeavour. Scientists generally work in teams in large laboratories. The case conference is widely used in health care. Even individual scientific research is vetted and reproduced by others and the mark of its success turns on its wider publication in journals or at conferences. Descartes' project, by contrast, is one he undertakes alone. This is not to say that others might not also follow the same thought processes. However, they will also do it alone. As we will see, and as often happens in philosophy, his final account of the mind owes much to this apparently innocent choice of starting point.

Sources of scepticism

As should be familiar, Descartes then attempts to set aside any beliefs that can be the subject of doubt in order, he hopes, to arrive at a firm indubitable foundation on which to base his subsequent beliefs. Rather than attempting to run through his beliefs piecemeal and assess each individually, he thinks of three general sources of doubt about empirical beliefs. First, beliefs grounded in his senses are subject to perceptual illusions. Secondly, as at any moment he cannot reliably test that he is not dreaming, empirical beliefs based on the testimony of his senses or his memory may be false. Thirdly, he may be subject to delusion by an evil genie. 'I shall think that the sky, the air, the earth, colours, shapes, sounds and all external things are merely the delusions of dreams which he has devised to ensnare my judgement. I shall consider myself as not having hands, or eyes, or blood or senses…' (p. 15).

Thus all beliefs that turn on the reality of the external world are equally subject to doubt and are (temporally) rejected.

(One possible line of objection to Descartes' project is to object that there is no natural kind that comprises our knowledge of the external world. What Descartes presents as merely a short-cut is an essential presupposition. He seems to assume two things. First, that there is a logical gap between inner experience and the state of the outer world. It is this which enables him to say that we could not tell from inner experience alone that we were not deceived by an evil genie (or nowadays that we were not a brain in a vat). However, he also assumes that to counter scepticism, we must form justified beliefs about the outer world *from* our inner experience. He assumes, in other words, a form of foundationalism, which is far from mandatory. For this line of criticism see Williams (1996) *Unnatural Doubts*.)

The cogito

A sceptical loss of knowledge of the outer world is the dismal conclusion of the first meditation. In the second, Descartes continues in this sceptical vein but is stopped in his tracks by the following thought:

> But I have convinced myself that there is absolutely nothing in the world, no sky, no earth, no minds, no bodies. Does it follow that I too do not exist? No: if I convinced myself of something then I certainly existed. But there is a deceiver of supreme power and cunning who is deliberately and constantly deceiving me. In that case I too undoubtedly exist, if he is deceiving me; and let him deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something. So after considering everything very thoroughly, I must conclude that this proposition, *I am, I exist*, is necessarily true whenever it is put forward by me or conceived in my mind.   (p. 17)

Thus the one thing that resists his sceptical attack is that he himself exists. It is another matter in what respect he exists: what he is. A similar procedure is followed to analyse this: 'I will then subtract anything capable of being weakened, even minimally, by the arrangements now introduced, so that what is left at the end may be exactly and only what is certain.'

As a result of this method, Descartes rejects definitions of himself as a rational animal or as embodied with a face, hands, and arms. All these substantial descriptions include elements whose existence cannot meet the strict epistemological pedigree imposed within the meditations. This leaves the minimal claim that he, Descartes, is a thing that thinks, that doubts, understands, affirms, and suchlike. It is worth briefly digressing here for a few paragraphs.

Digression on the validity of the argument

As has long been realized, if Descartes intends to ground a substantial negative conclusion that he Descartes is nothing more than a thinking thing, the argument suggested in the second meditation, is the wrong way to go about it. Using the resources available in this meditation, the best line of argument appears to be as follows:

1. One of the features or properties of myself is that its existence cannot be doubted by myself.

2. One of the features or properties of my body is that its existence can be doubted by myself.

3. If two apparently different things are really numerically identical (i.e. the very same thing) they must have all and only the same properties.

4. Therefore I am not my body.

This resembles a sound argument based on Leibnitz's law (the third premiss). I and my body differ in one property or feature and thus cannot be the very same thing. Suppose you were falsely accused of stealing a tie from a shop on an occasion when one but only one tie had been stolen and resembled the one carelessly tied around your own neck. One possible line of defence—if this were appropriate—would be to show that your tie and the one stolen differed in some feature. Perhaps the other (must have) had a 'St Michael' label and yours says 'Pierre Cardin'. If so, because they differ in this respect, they cannot really be one and the same and thus yours is not the stolen tie in question. (Perhaps you stole it from a different shop!)

In Descartes' case, however, this form of argument will not work. Consider (to use an oft-quoted example) that the Prime Minister were to awake suffering from amnesia. He might wonder who he was. Suppose he were to employ Descartes' method of doubt. Whimsically, or on the basis of some evidence found on the hospital ward, he might consider the possibility that he were, in fact, the Prime Minister. Now even if it were true that he were the Prime Minister, this fact would still be open to sceptical doubt. By contrast the fact of his own existence would not be open to him to doubt. By applying the same form of argument he could conclude that he were not the Prime Minister after all. As in this case true premisses would clearly lead to a false conclusion, the form of the argument in general cannot be valid.

The problem in Descartes' argument as formalized above is that the properties that plug into Leibnitz's law (whether in Descartes' case or the Prime Ministerial example) occur within what is called an 'intensional context' of a propositional attitude—in this

case the Descartes *doubting that* such and such. In well behaved *extensional* contexts, substituting terms that refer to the same thing keeps the sentence true if it was true or false if it was false. Consider: 'The prime minister of the UK in 2005 was a man.' This is true. Thus so is: 'Tony Blair was a man.' These two sentences only differ in the substitution of 'Tony Blair' for 'The prime minister of the UK' which refer (in 2005) to the same man. But consider: 'Smith believes that the prime minister of the UK in 2005 was a man' and 'Smith believes that Tony Blair was a man'. It is perfectly possible that one is true and the other false.

Substituting co-referring terms within the content of mental states does not generally preserve the truth value of such reports. I may be a fan of the Roman orator Cicero and thus: I believe that *Cicero was a fine orator*. But it is not necessarily the case that: I believe that *Tully was a fine orator*, because I may not believe that Cicero was Tully (even though in fact 'Cicero' and 'Tully' are different names for the same person). Leibnitz's law does not apply to differences within intensional contexts.

End of digression

From the point of view of a sympathetic reading of Descartes, it is not clear that the fallacious argument set out above is what underpins his negative conclusion. In Meditation 6, he brings out the extent to which the assumption that God would not deceive him in his (Descartes') appraisal of what is and what is not possible. But what matters to this chapter is not so much Descartes' dualist conclusion, but the way his starting point shapes both his and others' assumptions about our epistemological predicament.

The origin of the Problem of Other Minds

Descartes builds a picture of the world outwards from a perspective even more alone than a philosopher sitting alone before the fire in his dressing gown. He is alone with his thoughts only, the existence of his body being a matter of inference that will (later) be underpinned by God. His account of the world starts from his own first-person perspective and builds outwards. A key presupposition is that this first perspective is not threatened or changed by scepticism about the outside world. In this assumption he differs from the externalists about content discussed in Chapter 25. They hold that the very ability to hold at least some thoughts requires that one stands in some actual relation to the 'outside' world. Descartes subscribes to an internalist assumption that although his beliefs will be false if there is no external world, they will remain the very same beliefs. Their existence is world-independent.

Now one problem, forcefully presented recently by the philosopher John McDowell, with this view of one's standing in the world is that it makes the ability of thoughts to concern worldly objects mysterious (see Reading guide). If mental states are free-standing elements in my mental wardrobe, why should they be any more *about* anything than my trousers? On Descartes' account, mental states are immaterial, but they retain a thing-like quality. If so how can they, as free-standing things be about anything?

However, even setting this fundamental problem aside leaves a long-standing further difficulty. How can one have knowledge of, or even access in some weaker sense to, other people's minds? If mental states are elements in a private theatre with only one spectator, how can others gain access to them? This is the Problem of Other Minds. The classic solution is the argument from analogy to which we will now turn.

An attempt to solve the Problem of Other Minds

EXERCISE 2 (10 minutes)

Think about how one might know about other people's mental states if Descartes' account of the mind were true. How would access to your own mental states help? Try to take the problem seriously and consider what resources might be available to you. But also think how you do in fact experience other people's minds.

The argument from analogy

There are a number of plausible answers that might be given to the question: How can I know other people's mental states?

Perhaps the obvious answer is by *analogy* with my own. Of course another response is to assume in the face of Descartes' account of the mind as a private subjective realm that we do *not* know other people's minds. In other words, one might adopt a limited form of scepticism directed just at others' mental states while allowing knowledge of their outer behaviour. However, that is a radical and ultimately impractical line to take. (Try it!) Sir Bertrand Russell, for example, says:

> We are not content to think that we know only the space-time structure of our friends' minds, or their capacity for initiating causal chains that end in sensations of our own. A philosopher might pretend to think that he knew only this, but let him get cross with his wife and you will see that he does not regard her as a mere spatio-temporal edifice of which he knows the logical properties but not a glimmer of the intrinsic character. We are therefore justified in inferring that his skepticism is professional rather than sincere.
>
> Bertrand Russell 'Analogy' from *Human Knowledge: its scope and limits* ([1948] 1991, pp. 89–91)

Russell takes as the basic datum to be explained the fact that *we do know* the mental states, beliefs, moods, and desires of other people. We do not merely claim to know their physical properties, and their abilities to act (in the way that we know the dispositional properties of chemicals, for example). How is this possible, then? Russell sets out to articulate the principle on which such claims can be grounded given the basic Cartesian first-person perspective as a starting point.

His suggestion is that 'we must appeal to something that may be vaguely called "analogy". The behaviour of other people is in many ways analogous to our own, and we must suppose that it must

have analogous causes.' Russell suggests that because he knows from his own case that his thirst is the normal cause of such an utterance, when he hears the sentence 'I'm thirsty' when he himself is not thirsty he can conclude that it is probable that someone else is. He assumes this the more readily given other behavioural factors such as seeing a hot drooping body. 'It is evident that my confidence in the "inference" is increased by increased complexity in the datum and also by increased certainty of the causal law derived from subjective observation.' (p. 90)

Russell sees this form of inference as akin to the reasoning from effects to causes in the physical sciences. While no such inference can be certain—because the same effect may be caused in different ways—it can, he suggests, be probable. The general principle is summarized as follows: 'If, whenever we can observe whether A and B are present or absent, we find that every case of B has an A as a causal antecedent, then it is probable that most B's have A's as causal antecedents, even in cases where observation does not enable us to know whether A is present or not.' (p. 91)

Against analogy

Although there is something initially attractive about the pairing of a Cartesian picture of mind with the argument from analogy, there are powerful arguments against this as a coherent position. Many are developments of remarks made by Wittgenstein about *criteria* to which we will turn shortly. But Wittgenstein himself also deployed two other considerations.

Wittgenstein's criticisms

A simple argument is that analogical reasoning provides poor justification. Unlike the establishment of correlations in physics, for example, there is in principle only *one* kind of correlation on which to base one's analogy: the connection between one's own mental states and behaviour. This goes no way towards undercutting the worry that only I have mental states in the first place: that there is a fundamental distinction between my experience and anyone else's.

Imagination

A different objection is raised in the following passage from the Wittgenstein's (1953) *Philosophical Investigations*:

> If one has to imagine someone else's pain on the model of one's own, this is none too easy a thing to do: for I have to imagine pain which I *do not feel* on the model of the pain which I *do feel*. That is, what I have to do is not simply make a transition in imagination from one place of pain to another. As, from pain in the hand to pain in the arm. For I am not to imagine that I feel pain in some region of his body (which would also be possible). (§302)

The problem that Wittgenstein raises here is the difficulty of thinking of someone else's pain given only one's own experience of one's own pain. Given only this first-person starting point, what are the materials available to think of a pain that someone else experiences? It is not enough simply to imagine one's own pain in another body because this would be a case of feeling pain on their leg, for instance. So how could one break out of a first-person grasp of pain as pain-as-I-feel it to imagine pain independently of my experience?

Norman Malcolm and third-person concepts

Norman Malcolm does not rely on the difficulty of *imagining* others' pain given the first-person starting point but on a related point based instead on one's *understanding* or conception of others' pain. (In passing he mentions but does not develop a more purely epistemological point that the argument from analogy is riskily based on inference from one case: one's own mental states.)

Malcolm's argument could be summarized like this. The argument from analogy helps itself to something to which it has no right. It is presented simply as an epistemological solution to an epistemological problem: How can I make judgements about Jones' mental states, his pains, for example? Russell (1991) specifies a principle, which, he argues, provides the necessary justification for such judgements. On the basis of my own experience of the causes of my pain-behaviour (my pains usually), I can infer from Jones' manifest pain-behaviour that 'Jones is in pain' is probably true. But, Malcolm argues, this presupposes that I already know what 'Jones is in pain' means. How?

One possible response is anticipated by Malcolm. The objector suggests that given that I understand what 'pain' means from my own case, I can infer what it means in the sentence 'he has pain' because it means just the same. But while the last claim is true, it is no help in this case until I understand what *sameness* here involves. As we have already seen, it does not involve my feeling pain in his body, for example.

Criteria . . .

Malcolm goes on to argue that 'Smith is in pain' can only be given meaning if the rules for its use are specified via third-person *criteria* for its correct application. But once such criteria are in place there is no need for an argument from analogy because the criteria will provide direct rules for judging when the state of affairs it describes obtains.

The use of the words 'criterion' and 'criteria' come from Wittgenstein. There has been much discussion of what these are supposed to mean and what philosophical work they can do. We will return to that debate later in this chapter (in Session 4). To begin with, though, criteria are rules for the employment of words, rules which thus determine their meaning. Malcolm's point is that only given such criteria does the phrase 'Smith is in pain' have a meaning but as these rules will specify the circumstances under which it can be applied they will undercut the need for analogical reasoning from one's own case. Instead of judging that Jones is in pain because he is behaving the way I behave when I am in pain, I make this judgement because I know it is licensed directly by Jones' behaviour. That such behaviour is sufficient, in normal circumstances, to warrant the judgement is something I know simply by knowing the meaning of the phrase 'Jones is in pain'.

...and symptoms

This last point is an important feature of criteria that Wittgenstein contrasts with what he calls symptoms. The connection between phenomena and their symptoms is a contingent matter, something to be investigated empirically. By contrast the connection between criteria and what they are criteria of is an a priori matter. Wittgenstein comments that criteria and symptoms can be swapped. What has previously been regarded as a contingent concomitant can come to be regarded instead as definitional. (A neurological indicator of a psychiatric disease may come to play an essential or definitional role, replacing behavioural signs.)

Are criteria defeasible?

Without pre-empting the later discussion of criteria, it is worth here noting one important feature of the connection between mental states and behaviour. Apparently qualitatively identical behaviour can sometimes be the expression of a mental state—such as pain—and can sometimes be a sham. It may, for example, be part of a play. So any account on which behaviour plays a criterial role will have to be one that also allows the relation between apparently identical behaviour and underlying mental states to come apart. For this reason, the first commentators described criteria as analytic but *defeasible* indicators.

Although it has epistemological consequences, what drives Malcolm's argument are considerations about meaning. The key idea is that Descartes' starting point—a necessarily isolated individual identifying his or her own sensations prior to identifying other people's—cannot yield a conception of other people's mental states. Thus, as an account of how we can know about other minds, it cannot be right because we do have a conception of other minds.

A slightly formal argument against a Cartesian account of other minds

In the UK in the 1960s arguments of this form were influential and gave rise to a form of logical behaviourism: roughly the view that there are logical analytic connections between descriptions of behaviour and descriptions of mental states. For a while it seemed that solutions to the Problem of Other Minds amounted to a choice between the scepticism that results from starting from a first-person perspective and then attempting to work outwards, or the behaviourism that resulted from starting from a third-person perspective. One can think of this choice as following from something like the following argument:

i   No one can perceive anyone else's sensations; (assumption).

ii   All that can be perceived is behaviour (perhaps tautologically defined by what can be perceived); (assumption).

iii Knowledge of particular matters of fact is either directly perceived or inferred from direct perception; (assumption).

iv   Others' sensations are not directly perceived; (from i).

v   Either they are not known or are inferred from behaviour; (from iv, iii, ii).

vi   Inference is either deductive or inductive; (assumption).

vii Deductive inference involves logical truths; (assumption).

viii Inductive inference involves observed correlations; (assumption).

ix   There are no observed correlations of others' sensations and behaviour; (from ii).

x   Either others' sensations are not known or there are logical truths relating sensations and behaviour, i.e. either scepticism or logical behaviourism (from ix, vi, v, vii).

But as we will see, there is another option: theory-theory.

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. How does the central question of this chapter relate to the rest of Part V?

2. How does the Problem of Other Minds arise? How is it connected with a Cartesian picture of the mind?

3. How is the argument from analogy supposed to solve the problem? Is it successful?

Session 2   Introduction to the 'theory-theory' approach

Theory-theory as a response to logical behaviourism

This session introduces a very influential way of thinking about how we have knowledge of other minds now generally called the

theory-theory approach. A clear way of understanding it is by contrasting it with the 1960s criteria-based accounts influenced by Wittgenstein discussed in the previous session. This is not to say that all theory theorists either then or now were explicitly reacting to logical behaviourism, but it is true of the first reading.

In brief, the theory-theory assumes that we access others' minds by employing a tacit theory, which postulates unobservable mental states in much the way that physical theory postulates unobserved particles to explain the behaviour of observable bodies.

In a paper discussed below (and the reading linked with Exercise 4) with 'Operationalism and ordinary language: a critique of Wittgenstein', Charles Chihara and Jerry Fodor attempt to undercut the dilemma that the Wittgensteinians attempted to impose: either scepticism about other minds or logical behaviourism. They press for a third option. Access to mental states is via an empirical theory. Their argument against the Wittgensteinian account is justified in part by an analogy between knowledge of others' mental states and knowledge of atomic particles.

The background: Chihara and Fodor on Wittgenstein and criteria

Chihara and Fodor (1991) summarize the Problem of Other Minds as traditionally turning on the assumption that there are no conceptual or logical connections between behavioural and mental predicates. Combined with the assumption that one can only have direct first-hand knowledge of one's own states, there will turn out to be insufficient justification for ascribing mental states to others. Hence a form of scepticism. The reason for this is that any justification would have to rely on an observed inductive correlation between mental states and behaviour and only *one* such correlation can be observed.

They go on to report that, in opposition to this, philosophers influenced by Wittgenstein denied the premise and thus committed themselves to a form of logical behaviourism. Chihara and Fodor provide some general remarks about why this move fitted well into Wittgenstein's broader, operationalist approach to meaning. This, they suggest quite plausibly, is reinforced by and reinforces Wittgenstein's metaphilosophical claim that philosophical problems can be dissolved by attending to the way language is actually used, including cases where apparently similar grammatical forms mask very different uses. (Philosophical problems can result from assuming that there must be an underlying similarity.) Concentration on the use of words—such as 'pain' in 'Smith is in pain'—and on the way words are taught leads them to the role that criteria are supposed to play in a Wittgensteinian account of mind.

Wittgenstein's basic idea, they suggest, is that for mental concepts to be teachable, there must be criteria that can be invoked for their use and which are more basic than any inductive criteria. The reason for this dependence is that the correlations of, for example, symptoms with underlying conditions requires that one already understands the concepts of the kind of symptom and kind of underlying (mental state) condition in question. But one

can only understand the latter in mental cases if one already understands criteria for it. They summarize the role of criteria thus: 'X is a criterion of Y in situations of type S if the very meaning or definition of "Y" justify the claim that one can recognise, see, detect, or determine the applicability of "Y" on the basis of X in *normal* situations of type S.' (p. 141)

The stress on normality is the result of the assumption that criteria are (on this reading of criteria) defeasible. That is, under the 'wrong' circumstances the criteria can be satisfied without the condition for which they are criteria being the case. Thus a ball going into the net may be the criterion of a goal but only in a game of football, when the ball is in play, when no player is off side and so forth. (We will return to a different interpretation of criteria later in the chapter.)

EXERCISE 4   (30 minutes)

Read the short extract from

Chihara, C.S. and Fodor, J.A. (1991). Operationalism and ordinary language: a critique of Wittgenstein. Reprinted in *The Nature of Mind* (ed. D. Rosenthal). Oxford: Oxford University Press, pp. 137–150. (Extract: pp. 145–146)

Link with Reading 27.3

This paper attempts both to explain a broadly Wittgensteinian, logical behaviourist account of knowledge of other minds and then to argue against it.

◆ What is the nature of the alternative that the authors set out in this passage?

◆ How plausible is the analogy on which it is based?

The theoretical alternative

In effect, Chihara and Fodor offer the following argument. The Problem of Other Minds originates from the assumption that the only justification for ascribing mental states to other people turns on observed inductive correlations and there are insufficient such correlations to achieve justification. The only alternative, according to logical behaviourism, is to say that the justification for the ascription of mental states to others turns on *criteria* for the application of mental concepts. It is a priori that in *normal* circumstances such and such behaviour is sufficient for ascribing so and so in the way of mental states. This flows from our understanding of the mental and behavioural concepts in question.

But, according to Chihara and Fodor, there is a third possible answer to the question of how we gain access to other people's mental states. This is neither by observed correlations nor by logical connections but instead through theoretical inference. The best explanation of the outward behaviour is the existence of inner states described by the theory of folk psychology. Thus third-person access is justified by the possession of an appropriate form of theory.

In fact this sort of approach has already been hinted at in two previous readings in previous chapters. By returning to those, a little more flesh can be put on the theory-theory.

EXERCISE 5   (60 minutes)

Re-read the extracts from Fodor and from Dennett from Chapters 24 and 25, i.e. extracts from:

Fodor, J.A. (1991). Propositional attitudes. In *The Nature of Mind* (ed. D. Rosenthal). Oxford: Oxford University Press, pp. 325–338

Dennett, D. (1987). True believers: the intentional strategy and why it works. In *The Intentional Stance*. Cambridge. MA: MIT Press, pp. 13–35

Now read the short extract from

Fodor, J.A. *Psychosemantics*. Cambridge, MA: MIT pp. 1–2

Link with Reading 27.4

◆ How might a theory of mind solve the Problem of Other Minds?

◆ What sort of theory is it?

Fodor on theory

In the paper 'Propositional attitudes' and further in his book *Psychosemantics*, Fodor presents a reductionist philosophical explanation of mental content. As we saw in Chapter 24, Fodor defends a particular philosophical theory of how the intentional properties of mental states (and thus derivatively, and in turn, of language) result from underlying causal connection between internal mental representations and worldly states of affairs. That is not the concern of this chapter. What is of concern is the way Fodor begins: with an account of the connection between mental states and the *theory* which he claims is implicit in folk psychology. Here he begins by arguing that our knowledge of the mental states of other people takes the form of an implicit theory that he calls 'folk psychology'.

(The argumentative strategy that Fodor continues in *Psychosemantics* as a whole is this. On the assumption that our knowledge of other people's minds can be codified in a theory, some explanation of why this theory works, of what features of the world it is true of, has to be given. Fodor argues that the explanation of this is that folk psychology is a theory that governs the internal mental representations described in Fodor's Representationalist Theory of Mind. This is how he attempts to ground a causal theory of the mind. We are here concerned with the first stage of that overall argument: the claim that our knowledge of other minds is codified in a theory of folk psychology.)

Folk psychology and folk explanation

Fodor argues that we can explain and predict the behaviour of others by postulating an underlying ontology of mental states.

(Fodor thinks that possession of a mental state can itself be explained by the possession of an internal mental representation. But that is a further stage.) The behaviour of these can be codified in the generalizations of folk psychology. One such generalization is that *if x is y's rival then x prefers y's discomfiture, all else being equal*. These fit together into a deductive structure to explain particular actions, broadly resembling the Deductive-Nomological model of explanation (see Chapter 14). Given knowledge of the particular circumstances of a person, we can apply generalizations to deduce their subsequent thoughts and actions. How is it possible to begin this process and have knowledge of particular matters? The answer is the same as how one knows of particular but unobservable matters in physical science: by a holistic fitting of the observed facts into a broader theory.

Fodor argues that in practice (and possibly in principle as well) there is no alternative to the deployment of this theory if one wants to explain and predict behaviour. How else could one predict that one would be met from an airport in 3 weeks' time, having made such an arrangement over the telephone? So folk psychology is *practically* compulsory. But as long as it is a more or less true theory of how we think and behave, it will underpin inferences from behaviour to their underlying mental causes. This is a matter not of observed correlations, nor of logical criteria, but of the best explanation of observable effects. It ties a variety of behavioural effects together in a plausible overall explanatory theory.

The analogy with a theory of meaning

It is worth noting that theories of this sort have been offered for other competencies. We have already briefly encountered one in Chapter 25. Davidson's formal theory of meaning is an attempt to codify the practical competence that speakers of a language have in the production and understanding of novel sentences. It has widely been assumed in recent analytic philosophy that such a theory goes some way to explain an ability by showing that it can be built up from component abilities: tacit knowledge of axioms of the theory, perhaps. What has not been explained is the precise explanatory burden such a theory is supposed to carry. Fodor is a rare exception in this regard. He is quite explicit that he thinks that corresponding to each of the elements described in a theory of grammar there will be mental elements, real mental entities. But other philosophers, including Davidson, have not wished to purchase clarity at this great cost. Given, however, that the theory-theory has been invoked explicitly in this chapter to explain third-person epistemology, one of its burdens will be to explain precisely how it is so explanatory. Before going on to examine a rival to the theory to ease just that assessment, we will briefly digress to revise Dennett's views on this issue.

Dennett on theory

Dennett agrees with Fodor in his emphasis on the importance of folk psychology as a predictive and explanatory strategy. To that extent it looks as though he shares a basic theory-theory

approach. There are two important differences from Fodor and possibly theory-theory in general, however, in what he goes on to say. The first was touched on in Chapter 25. Dennett does not think that folk psychology as a theory stands in need of the further explanation that Fodor goes on to give. Dennett does not subscribe to Fodor's 'industrial strength realism' of *causally* active internal states. Folk psychological explanations are distinguished from those physical science explanations that cite the behaviour of unobservable *entities*. It thus does not require the further postulation of those entities to explain its success. Instead it resembles physical science explanations, which describe calculation bound entities or *abstracta*.

Explanation and rationality

This leads to a second difference from Fodor's account. Dennett stresses the claim that folk psychological explanations resemble explanation by comparison with an ideal rather than by subsumption under a generalization. Think of this as the difference between explanations by appeal to how people *ought ideally* to behave and those that chart the most *probable* ways in which people behave based perhaps on past statistical observations. Dennett's emphasis on the ideal results from the central role that *rationality* plays in the Intentional Stance. The idealizations involved are those of how people ought rationally to behave. In fact this is very pertinent to the subject matter of this chapter. It provides a reason for thinking that Dennett does not subscribe to a standard form of the theory-theory. But before that point can be explained, it will be useful to look first at its major competitor. (We will find a similar emphasis on the role of rationality in one of the founding papers of the rival to theory-theory.)

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. How does theory-theory attempt to solve the Problem of Other Minds?

2. What is the connection between it and a functionalist account of mind?

Session 3   Simulation theory

Summary so far

The previous session presented the origins of theory-theory as a reaction to an argument that logical behaviourism was the only opposition to scepticism about other minds (conjoined with the claim that such scepticism was obviously false). The reaction stressed the broadly theoretical nature of folk psychology: those everyday generalizations and rationalizations that can be used to explain and predict behaviour. On the assumption that folk psychology can be classed as a theory, then the physics of the microscopic can provide an analogy of how we can know other people's mental states. We have a working theory of the underlying unobservable (mental) causes of observable (behavioural) effects. This underpins the inference from behaviour to mental states.

Theory-theory was also an attractive account of the epistemology of third-person access to mental states because it fitted into a broader programme of explaining complex abilities by breaking them up into component abilities in turn codified in a theory. If, for whatever reason, the existence of complex open-ended abilities requires the truth of a theory that articulates their structure, then why not assume that the use of that theory underpins more directly the relevant epistemology?

Simulation theory opposes these views at least with respect to reading other people's minds. Rather than relying on knowing a theory of mind, simulation theorists argue that we simply use the 'mental mechanisms' that theory theorists' theories were supposed to codify. The reading contains an extract from one of two seminal papers which, independently, arrived at broadly similar positions although for different reasons. The second is discussed below.

EXERCISE 6   (15 minutes)

Read the extract from the beginning of:

Heal, J. (1995). Replication and functionalism. In *Folk Psychology* (ed. M. Davies and T. Stone). Oxford: Blackwell, pp. 45–59. (Extract: pp. 45–47)

Link with Reading 27.5

◆ What is the role that rationality plays in Heal's argument against theory-theory?

◆ What is the alternative that Heal proposes?

◆ Does she avoid merely repeating the argument from analogy?

Heal's account of the motivation of theory-theory

Heal begins by summarizing what she sees to be the main motivation behind theory-theory. She does this by sketching out a broadly functionalist theory of mind. (It is worth asking whether all forms of theory-theory share the features that she identifies in functionalism. Think about the role of rationality in functionalism. This important issue will return shortly.) Functionalism provides a way of explaining the open-ended nature of our ability to ascribe mental states to other people. This ability is open-ended both because there is a vast number of different *kinds* of psychological state (beliefs, hopes, fears, desires, expectations, etc.) that can interact, but also because the *contents* of such states are not limited. I can believe that red is my favourite colour. And I can also believe that red is my favourite colour *and that* Paris is

the capital of France. There is no limit on the number of different beliefs that can be achieved through conjunction alone. So how is it possible to be able to ascribe states of this complexity and open-ended multiplicity to other people?

A theory such as functionalism goes some way towards breaking this ability down into component parts standing in functional relations. A theory such as Fodor's Representationalist Theory of Mind, which deploys a language of thought, goes even further. It tries to explain how the content of complex beliefs is built up from component 'words' in a mental language. So here we have a connection between the subject matter of this chapter—the epistemology of mind—and that of Chapter 24. A theory of mental content explains how it is possible to entertain an unlimited open-ended number of different mental states. But if we already have tacit *knowledge* of such a theory, that would also explain our ability to handle complex *ascriptions* of mental states to others. Heal suggests that this is the motivation for what is now called a theory-theory approach (a label she does not herself use). But the motivation for ascribing tacit knowledge of such a massively complex theory (albeit one that reduces the complexity of the states it codifies) is undercut if there is a simpler explanation of our ability to 'read minds'.

Heal's alternative

The alternative that Heal suggests is 'replication' (or 'simulation' as the strategy has become known). Given that we have imaginative abilities to help us cope with a complex and unclear future by thinking through what it would be sensible to think and do in possible situations, we can use this to ascribe thoughts to others *without any further elaborate theorising about them*. (p. 47)

> I place myself in what I take to be his initial state by imagining the world as it would appear from his point of view and I then deliberate, reason and reflect to see what decision emerges . . . To get results from the method I require only that I have the ability to get myself into the same state as the person I wish to know about and that he and I are in fact relevantly similar.   (p. 47)

What marks this strategy out is that one simply *uses* one's own 'mental machinery' in deliberating to find out what the other person thinks. One need not tacitly know a *theory* of how one's mind works.

Three objections

To defend this strategy, Heal has to show that it really is distinct from a theory-theory approach, that it does not presuppose theoretical knowledge. She suggests three potential criticisms and responses to them:

1. Replication, as Heal describes it, requires that one is able to put one's self in the same initial state as the other person, prior to running the simulation. But surely this will have to require an inference from behaviour to mental states and this simply smuggles a theory-theory back in.

   ◆ *Response*: Heal's response to this is, first, that there may be a more direct way of reading another's mental state than

theoretical inference. Secondly, the simulator can gauge the initial state by looking at how the world impacts on the other in order to simulate the initial state. As her first answer is not consistent with pure simulation theory but an example of the third general approach to the Problem of Other Minds we will postpone discussion of it until later.

2. Simulation will require putting oneself in a state of *make-believe belief*. Knowing what follows from some make-believe belief will surely itself be a piece of theoretical knowledge.

   ◆ *Response*: this is not so. We already use our imaginative abilities use our in practical reasoning about our own future experiences, thoughts, and actions. We do not employ theories to think what we should think under hypothetical circumstances, but simply reason directly. (This is called 'off line' reasoning by others.)

3. Working out what follows from an initial state will itself be a matter of deploying a theory of mind of applying principles of reasoning. Thinking is thus just first-person access to a theory of mind.

   ◆ *Response*: such principles are not causal generalizations but normative and rational principles about what ought to follow from what or what would make a particular belief true or desire satisfied. In part, this response stems from the way Heal characterizes her opponent as a functionalist rather than more broadly as a theory theorist. It turns, in other words, on assuming an opponent like Fodor rather than Dennett. This is brought out in section iii of her paper.

The role of rationality

In addition to the basic argument that replication is simpler and more plausible than ascribing tacit or implicit knowledge of a theory of mind that has yet to be articulated, Heal also deploys a positive argument in favour of replication over theory-theory. This turns on the role and nature of rationality in psychological explanation.

The basic idea is this: 'in giving a psychological explanation we render the thought or behaviour of the other intelligible, we exhibit them as having some point, some reasons to be cited in their defence. Another way of putting this truism is to say that we see them as exercises of cognitive competence or rationality.' (p. 52)

This claim resembles the distinction mentioned earlier in the chapter between explanation by subsumption and explanation by appeal to an ideal.

However, the presence of rationality makes a second difference. Heal suggests that the demands of rationality cannot be codified into any kind of theory that would serve the theory-theory. To be rational does not guarantee that any specific belief must be true. One can always ask the question: 'is this belief really true?' without that casting doubt on one's cognitive competence. Following Quine, we cannot pick out any belief as immune to future revision. But nor could rationality or competence be identified with making the correct application of particular rules of inference.

I can fail to follow simple and reliable inference rules and adopt some most unreliable ones, and recognise later that this was what I was doing, quite compatibly with continued trust in my then and present cognitive competence. The only constraint is that I should be able to make intelligible to myself why I failed to notice so-and-so or seemed to assume such-and-such.   (p. 53)

Either erroneous beliefs or erroneous inference rules can be made intelligible providing that one can give the right stage-setting. But what the right stage-setting is cannot itself be specified in advance or codified. Of course partial models of good thinking are encoded in, say, first order deductive logic, but failure to follow such models does not by itself imply a failure of overall cognitive competence.

If these thoughts are right and the demands of rationality cannot be codified, then the theory that any plausible theory theorist requires cannot be framed. This clearly counts against the theory-theory; however, it does not count against replication or simulation. In this alternative strategy, one assumption built into the replication is that others are like me in being rational. This is made no more problematic for the absence of a final account of what such rationality comprises.

Robert Gordon's account of simulation theory

Jane Heal was not the only philosopher to arrive at the idea of replication. Robert Gordon arrived at similar conclusions—although he coined the now more popular label 'simulation theory'—in R.M. Gordon (1995a) 'Folk psychology as simulation' (in Davies and Stone, ed. *Folk Psychology*).

Gordon's paper begins more straightforwardly and makes no explicit appeal to the connection between psychological explanation and rationality. (There is, however, an implicit connection through the role of practical reasoning.) However, he makes some interesting further claims towards the end.

While Heal appeals to the imaginative abilities we have to consider our own possible reactions to possible (distant) future events, Gordon starts with an appeal to our ability to predict our immediate future actions on the basis of our intentions to act. We can predict how we are about to act without the use of any theory, or general principles, which govern human action, except for the principle that we shall act as we intend, in general. As intentions are often the product of practical reasoning—reasons that justify or rationalize actions—simulated practical reasoning with hypothetical premises can serve as a predictive device of our own more distant future actions. The paper then suggests that this same ability can be put to use to predict the actions of others, 'As in the case of hypothetical self-prediction, the methodology essentially involves deciding *what to do*; but, extended to people of 'minds' different from one's own.' (p. 63)

So we can apply the very same non-theory-based ability to predict our own actions through the use of practical reasoning to predict the actions of others providing we can simulate their practical reasoning by 'putting ourselves in their shoes'. One projects oneself into another's situation *but without any attempt*

to project oneself into, as we say, the other's 'mind' (p. 63). Likewise it may not be possible for me ever to be in the situations of others whom I attempt to simulate—perhaps because the situations lie in the past—and so I do not have to decide what I myself would do or believe that I am in those circumstances. Instead I project myself into the other person's circumstances. Thus simulation does not involve implicit analogy with my own case. These final qualifications are important and we will return to them shortly. They are supposed to head off an objection that might have been niggling you so far in this session.

Simulation and concepts

So far we have focused on both Gordon's and Heal's epistemological arguments (although the latter's paper covers broader areas also). We have not touched on the question of what account of the *concepts* of various mental states can be given from the perspective of simulation theory. As well as accounting for our ability to ascribe mental states to others, can it also account for what it is that we conceive when we conceive of mental states (from a third-person point of view)?

This question can be brought into focus by drawing two different contrasts. Theory-theory provides not only an account of how we can know of others, mental states (answer: by implicit knowledge of such a theory of mind), but it also accounts for what beliefs are. They are just those (internal, causal) states described by the theory that have the inputs and outputs and general behaviour articulated in the theory. As simulation theory subscribes to no such theory, what account can it give?

Malcolm's Wittgensteinian attack on the argument from analogy discussed in Session 1 was based on the claim that the argument presupposed an account of the *concept* of a mental state characterized in third-person terms—such that Smith could be in pain—while precluding the materials to construct just such a concept. Malcolm concluded that unless one already had a more or less direct way of determining the third-person applicability of mental descriptions, arguing by analogy from one's own first-person case could not get started. Once one had those, there was no need for the argument from analogy. Can simulation theory escape a related charge and provide the resources for explaining the general concept of mental states such that they allow both first- and third-person ascription? (If the answer is no then it may be that the best account of the mind will involve elements of both simulation theory and theory-theory.) Gordon's paper provides one possible line of defence.

Gordon suggests that the concept of belief comes from the interaction of simulators. In order to distinguish a bona fide belief in *persona propria* from a statement of belief within the context of a simulation or pretence, some linguistic device is needed. One way of doing this would be to announce that one was about to run a simulation:

1 Let's do a Smith simulation. Ready? *Dewey won the election.*

The same task might be accomplished by saying:

2 *Smith believes that* Dewey won the election.

My suggestion is that (2) be read as saying the same thing as (1), though less explicitly... To attribute a belief to another person is to make an assertion, to state something as a fact, *within the context of practical simulation* (Gordon, 1995a, p. 68).

Empirical evidence

This is just the beginning of a suggestion about what concept of belief arises out of simulation rather than a worked out account. However, it leaves the following doubt unresolved. Is Gordon's account really sufficient to explain the concept of belief that we have, the concept of a state that can represent the world both truly and falsely and can serve as part of a reason for action? Gordon thinks that it is and suggests, furthermore, that there is some empirical evidence for this view.

1. The fact that children only learn to pass the false belief test at the age of 4 or 5, but are nevertheless able prior to that to make predictions based on true beliefs (or at least *shared* false beliefs) fits simulation theory better than theory-theory. According to the latter, to understand the very concept of beliefs is to understand a theoretical context. Thus Gordon argues that the *change* in children's ability can be explained as a development of a child's ability to simulate obstacles to other people believing what the child knows. By contrast a theory-theory has no account of this change. If a child can explain behaviour at all by ascribing mental states, the difference between true and false cases should be unimportant.

2. The fact that there is a correlation between a lack of understanding of others and a lack of imaginative play among autists fits a simulation account of third-person access. A lack of the latter would preclude the former.

We will return to whether empirical evidence supports simulation theory rather than theory-theory shortly. That debate is far from closed yet. However, Gordon's suggestion that empirical evidence also supports a simulation account of mental concepts appears less convincing. The problem is not so much that simulation theory could not underpin mental concepts but that no gesture has been given as to how it might. The false belief test, for example, suggests that children younger than 4 have not mastered the idea of beliefs as states whose contents can diverge from how the world is. But it is not clear that possessing the concept of beliefs or other mental states can be equated with an ability to simulate, or to form beliefs about the world in a pretend or off line mode. Perhaps simulation theory provides the materials for learning how to ascribe mental states to other people, but a full understanding of what such states are requires that children form a theoretical understanding of beliefs.

Is simulation theory just a version of the argument from analogy?

Putting matters like that, however, raises a further objection to simulation theory. Can it avoid the various objections raised against inferring from the first person to the third as an answer to the Problem of Other Minds? In fact Gordon (1995b) faces just this sort of objection in a different paper ('Simulation without introspection or inference from me to you' in Davies and Stone, ed., *Mental Simulation*).

There, Gordon claims that other philosophers who have developed simulation theory have subscribed to three principles all of which he rejects:

1. an analogical inference from oneself to others,

2. premissed on introspectively based ascriptions of mental states to oneself, and

3. requiring prior possession of the concepts of the mental states ascribed.

The first two views fit naturally together. As we saw earlier, the argument from analogy was deployed in conjunction with a broadly Cartesian model of first-person access. If one finds out about one's own mental states by a kind of inner perception of private mental items, it seems that the only hope of discovering the states of others is a form of analogical reasoning.

Gordon suggests that many other simulationists think along these lines. One first discovers what one would think in someone else's circumstances and one then reasons that they will think the same way. Gordon argues that this version of simulation theory will face some of the same objections that the argument from analogy faced. What justifies the inference from one's own case to others? How can one come to think of the mental states of others on the basis of one's own? He argues that either these objections will be impossible to overcome or if they can be overcome then the resultant theory will be a form of theory-theory. The inference from my own case to others, for example, will be a theoretical inference.

Gordon's proposed alternative works like this. Instead of imagining oneself in another person's position, seeing what state one is in and inferring that the other will be in the same type of state—an inference that requires one has a prior grasp on the concept of a type of mental state—one imagines that one is the other person in their position. One imagines, *in the first person*, the other person missing a flight or whatever. One does not transfer one's mental states to another, but transforms oneself into them, in imagination.

Similarly, Heal (1995) says of a related criticism that it 'misdescribes the direction of the gaze of the replicator. He is not looking at the subject to be understood but at the world around that subject. It is what the world makes the replicator think which is the basis for the beliefs which he attributes to the subject.' (p. 138) But she does also concede 'one simple assumption is needed: that they are like me in being thinkers' (p. 137).

So the question of whether Gordon's is really a coherent and plausible account of mental gymnastics is open to question. Can it escape the charge that there is still an analogical assumption present? Clearly, for example, there would be no question of confusing one's thoughts, arrived at by imagining the world surrounding the subject in question, with one's own. Their

sorrow in the face of bereavement remains their sorrow even if one can imaginatively grasp its scope and object. Thus one line of objection to simulation theory is that it is not itself finally coherent. But in fact, that is not the most obvious difficulty which is instead to distinguish it from its main rival. It is to this question we will now turn.

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What is the main simulationist criticism of theory-theory?

2. How does simulation theory avoid this criticism itself?

3. Is simulation theory a version of the argument from analogy?

Session 4 **The current state of the debate: evidence from autism**

Summary

The previous two sessions have introduced both the theory-theory and simulation theory in the context of solving the Problem of Other Minds. The theory-theory likens our ability to 'read other minds' to a structure or body of knowledge we have of the mental workings of other people. Such a structure of knowledge underpins our ability to explain and predict their behaviour. It was introduced as an alternative to logical behaviourism: the view that there are logical connections between mental descriptions and behavioural descriptions, to which we will return in the final session.

Simulation theory is itself a reaction to theory-theory. It holds that our ability to read other people's minds turns on our non-theoretical ability to *use* our own minds in hypothetical practical reasoning. We saw that there was some risk that simulation theory is vulnerable to the argument from analogy, that it is really of a piece with a Cartesian theory of mind.

Evidence from autism

The discussion has, so far, been framed in largely philosophical terms. The arguments involved have concentrated on whether the various ideas are coherent, whether they would lead either to formal contradictions or informal absurdities. But the debate about theories of mind has also made much use of empirical findings. Some of the original research was done in the field of primatology. More recently, autism has been deployed as a rich source of empirical tests of philosophical theory. In this session we will examine some work on autism. One of the more apparent lessons

from this will be just how difficult it is to draw clear-cut conclusions once one steps away from pure philosophy and looks instead at the interplay between philosophy and an, as yet, unsettled empirical matter.

This session will consider a recent short summary of the debate and look towards the further applied work that has also been carried out. We will begin by examining how clear-cut the distinction between the two theories is.

EXERCISE 7   (30 minutes)

Read the extract from:

Stone, T. and Davies, M. (1996). The mental simulation debate: a progress report. In *Theories of Theories of Mind* (ed. P. Carruthers and P. Smith). Cambridge: Cambridge University Press, pp. 119–137. (Extract: pp. 131–134)

Link with Reading 27.6

◆ Is there really a clear distinction between theory-theory and simulation theory?

Stone and Davis set out a general overview of the opposition between theory-theory and simulation theory and then question—in the extract above—the state of play. What emerges is that, by contrast with the discussion in the previous session, there may be less of a clear distinction between the two explanations of our access to other minds than was at first thought.

Three forms of theory

What is particularly interesting is the discussion of just what sort of theory the theory-theory presupposes. The paper suggests that there are three dominant strategies varying from the most specific and tightly defined (and therefore least likely to be true) to the most relaxed view.

1. The first is to model the theory of folk psychology on a scientific theory: a tightly deductive theory that introduces entities in what are called Ramsey sentences. This view fits particularly with identifying psychological explanation construed as a form of regularizing explanation with a Deductive-Nomological model of scientific explanation. But as folk psychology differs from a scientific theory in various ways, this is also a (needlessly) implausible analogy.

2. More plausibly, one can compare folk psychology with the theory that is supposed to underpin our linguistic abilities and of which, according to much cognitive science, we have tacit knowledge. On the other hand the paper suggests, however, folk psychological explanation implies that we do have explicit knowledge of at least some of the principles in the theory. This marks a contrast with the linguistic case. In fact things are not so clear here as the paper here suggests, as Heal rightly emphasized, the principles contained in a theory-theory would have to include an unlimited number of detailed principles about

the behaviour of mental states according to their different contents. These would surely have to remain tacit and distinct from the principles used in everyday explanation.

3. Least contentiously, theory-theory can simply describe folk psychology as a body of knowledge akin to that of cookery, which explains our ability even if it cannot be deductively structured with the rigour demanded in the first case. Given this construal, to mark off a clear alternative simulation theory must deny that we negotiate the social world in virtue of some general principles about how people behave, and thus must assert that this differs from how we negotiate the physical world.

This summary of theory-theory encourages a twofold characterization of simulation theory (roughly how it has been described above). It involves an imaginative transference or transformation of perspective that requires the use of one's mental faculties rather than a theory of them. And it gives rise to forms of explanation that make sense of, or render intelligible, behaviour without subsuming it under statistical generalizations.

Three other dangers

There remain three dangers of the positions still blurring together which the paper picks out:

1. Much depends on the account of tacit knowledge of a theory if the ability to simulate others is not itself to count as such tacit knowledge.

2. In order to cut down on the complexity of their proposed theories, theory theories sometimes invoke an element of simulation within it. This is to enable an interpreter of another person's behaviour to understand the consequences of ascribing a belief with such and such content to them. Instead of having a massively complex theory of what each such content entails, an interpreter may simply simulate a belief of that content and see what it entails.

3. Most simulation theorists agree that some inductively based generalizations play some part in simulation.

Finally, Stone and Davies here are rare within the current debate on theory-theory and simulation in realizing that these positions do not exhaust all the possible contemporary positions. A third option is to abandon a 'regularising' conception of psychological explanation without thinking that this necessitates an imaginative identification with other people. One way of doing this is to subscribe to a form of *direct* access to which we will turn in the next and final session.

How is empirical evidence brought to bear?

The interdisciplinary debate about both the nature of autism, its explanation, and the light it sheds on the rest of the population is still wide open. We will shortly mention some representative recent papers. But the key issue to be clear on is *how* empirical evidence might shed light on such a debate and what the role of the philosophical theories is.

Take the case of autism and, especially, Wing's triad. This is a group of three symptoms that are centrally important in the diagnosis of autism. They comprise impairments of social competence; impairments of communication skills; and lack of pretend play. It is an empirical matter that these three distinct symptoms often occur together (in autism). This empirical finding might then be thought to shed light on how we have knowledge of other minds in something like the following way. Because it turns out that Wing's triad includes lack of pretend play, this suggests that the key deficit may be a lack of ability to simulate and it is this lack that explains the poor social competence. Thus, Wing's triad is evidence for simulation theory. (In fact a rival argument can be given that it supports theory-theory.)

The point is this. If it is legitimate to analyse a complex ability (such as the ability to read minds) as consisting in (and perhaps built from) simpler component abilities, then deficit studies should present evidence for or against particular analyses. The battle between simulation and theory theories is over which analysis is correct. Does both mind reading and pretend play turn on simulation or the possessing of theoretical 'metarepresentations' of others? Of course an alternative view is that there is something illicit or unnecessary about the attempt to analyse mind reading as dependent on lower but still mental skills. We will return to that option in the last session.

The complexity of the empirical findings

Two philosophers who have considered the role of empirical evidence and who take opposing views on it are Gregory Curie and Peter Carruthers. (See Currie 'Simulation-theory, theory-theory and the evidence from autism' and Carruthers 'Autism as mind-blindness' both in Carruthers and Smith, ed., *Theories of Theories of Mind*, 1996.)

Both Currie and Carruthers present some of the central or classic symptoms of autism and consider the relative merits of simulation theory and theory-theory in the light of them. As Currie comments at the very end of his paper: 'a fully satisfactory assessment of the simulation-versus-theory debate, even as confined to the present state of the evidence, would have to take into account the whole vast and bewildering range of autistic symptoms. Here I have focused on just a few of them.' (p. 256)

Thus, the extent to which he is able to marshall the evidence to present a case for one or other side may depend on careful selection only of 'Wing's triad' and some executive function deficits. The same consideration also applies to Carruthers.

This is the chief interest, in this context, of Boucher's paper also in Carruthers & Smith, (eds) 1996. It points out in some detail the clinical difficulties in pinning down just what is to be explained. Can autism really be tied to specific shared deficits, or are there a range of overlapping deficits but with no specific symptom or symptoms necessary and sufficient for such a diagnosis? The kind of explanation that both Currie and Carruthers aim at requires for its plausibility just such a clear clinical picture.

Wing's triad

Currie and Carruthers agree that both theories explain equally plausibly two of the deficits of Wing's triad: impairment in social relationships and in verbal and non-verbal communication. These features are named after Lorna Wing, a foremost autism researcher, who established the triad in the 1980s. However, both agree that there is at least a prima facie case that simulation theory can better explain the third: the absence of pretend play. It is reasonable to argue that pretend play just *is* the simulation of other roles: whether pirates or bears. Thus Carruthers, who attempts to defend a theory-theory, has to provide an account of this deficit using the explanatory prior idea that autists suffer from a deficit in their theory of mind. His idea is that autists do not engage in spontaneous pretence because they are deprived of the source of enjoyment in such play. This is, he claims, the manipulation of one's own mental states, which, he argues, requires consciousness of the pretence. Thus only those children able to form a second order belief (a belief about their beliefs) will be able to enjoy pretend play. The plausibility of this claim turns on the idea that this is what makes pretence pleasurable.

Problem solving

A further area of dispute is what better explains the findings that autists do worse at practical problem solving. Currie argues that while this is not at first an obvious consequence of deficits at simulation, a plausible connection can be found. Solving problems such as the Tower of Hanoi is generally a matter of simulated trial and error. One thinks through what the consequences would be of making a certain move, and what one ought to do as a result so as to eliminate bad strategies. Thus if autism is an inability to simulate, autists will also be bad at such practical problem solving.

Carruthers, however, suggests that an equally plausible explanation can be given that relates such cases to disruption of a theory of mind. Practical problem solving requires that one can both access and reflect on one's past thoughts and processes of reasoning. However, both of these abilities are mediated by a theory of mind module. Note that in saying this, Carruthers takes it that first-person access to one's past thoughts is a matter of theory, a view that Currie denies (and which is not part of the relaxed view of theory-theory discussed above). But whether or not this is so, it is at least consistent with theory-theory to say that the assessment of reasoning processes will require a working theory of mind.

So is there a difference between theory and simulation?

What the contemporary debate shows is that things are not as clear-cut as they may have at first seemed. In fact, this general debate is beginning to lose momentum as the participants, especially those in primate studies, begin to realize that the opposition between theory on the one hand and the ability to simulate on the other is not sharp. In the final session we will turn to a different tack that owes something to the Wittgensteinian account of the 1960s in its refusal to *theorize* at all. For now,

however, the following distinction in emphasis may be a helpful way of thinking about theory-theory and simulation theory. Simulation theory is better able to account for the role of rationality in mind.

Take Heal's argument for simulation theory. A key idea was that we can make sense of other people by thinking about what it would be rational to think in their position. We do this not by using a *theory* of rationality but just by using our rationality, our faculty of rationality, perhaps.

To use the language of a reading from Chapter 15 from McDowell, the structure of the 'space of reasons' is a rational structure. It concerns the support that beliefs give one another. Significant divergences from such a rational structure would not merely be unfortunate, they would be *unintelligible* as a structure of beliefs at all. But this idea that there are limits on what we could even understand finds no echo in the resources of natural laws, which would have to say merely that one sort of internal state tended to be followed by another. A sideways view of the rational structure could not explain how it marked the limits of intelligibility.

Of course this rational structure is a kind of ideal up to which we do not always live. Reason explanations have a different kind of logic to natural scientific nomological explanations. We *compare* with an ideal rather than *subsuming* under a universal law. We can make sense of virtually any mistaken beliefs they might have as long as we understand how they were thought to follow from their other beliefs. This elasticity in the structure of beliefs resists capture in a *theory* of other minds. It again brings out a difference between the structure of rationality that constitutes our thinking and a structure of laws. As an ideal we must also be alive to the thought that our conception of what follows from what, what gives reason for what, stands in need of correction and improvement. Again, this structure looks very different to the structure of a theory.

These considerations were deployed in Chapters 23 and 24 against the idea that mental states can be cashed out in functional terms. In other words, they were considered in the context of the *ontology* of mind. Do they also apply in the case of the *epistemology* of mind? Not entirely, but there is some sort of suggestive connection.

The reason they do not simply carry over is the difficulty of knowing what role a theory has and what constraints are put on the nature of that theory. No one thinks we explicitly use a theory to explain the everyday behaviour of others. It is at best tacit. And if a rational stance such as Dennett's Intentional Stance—which explicitly employs principles of rationality—counts as a theory, it is not clear that theory-theory cannot build rationality in. At the same time, however, it is against the spirit of theory-theory to acknowledge the central role of rationality given that its demands cannot be codified in a theory.

Having said that, both theory-theory and simulation theory assume that a philosophical account is needed in response to the Problem of Other Minds. In the final session we will turn to a different approach.

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. How can the debate between simulation theory and theory theory be assessed?

2. What role does evidence for autism play?

3. Are the two positions really distinct?

Session 5 Rationality and direct access to mental states

This chapter began with the observation that the knowledge of other minds is central to mental health care. On the one hand, there are specific syndromes or conditions that involve a partial breakdown of our everyday ability to make judgements about others' minds. On the other hand, that general ability also underlies clinical practice. Only if one can form a picture of a client's state of mind is clinical intervention possible. We have examined the two dominant contemporary philosophical views and begun to see how empirical evidence from autism might impact on an assessment of their plausibility.

Is there a need for a substantial theory at this (mental) level?

The accounts of third-person access considered in the last two sessions have both assumed that there was a substantial problem to be addressed whose solution would require a philosophical theory. Both theory-theory and simulation theory are *substantial* theories in that they attempt to explain an ability that can seem mysterious: we are able to 'read other people's minds' despite the fact that minds are generally hidden. It is, in part, because of the assumption that mental states are not directly available to people other than their subjects that drives the need for an explanation of how we are able generally to know about mental states.

(Again, it is important to remember that this chapter has not investigated the question of how *first-person* 'access' to mental states is possible. But not that as some of the readings have suggested, some theory theorists argue that even this is a matter of theoretical *inference*.)

One of the advantages of such a construal both of the problem and of the general form of the solution is that it suggests that they may be an explanation of breakdowns of the ability to mind-read pitched at a mental level. If the full ability to understand and explain others' behaviour can be analysed as comprising more basic abilities, then there is the prospect of explaining failures of the full ability in terms of failures of these lower-level abilities, which still deserve mental descriptions. Contrast this case with a case where the breakdown has to be explained as a breakdown at an underlying physical level. (One very crude analogy is to think of the contrasts between software bugs and hardware faults in computing.)

However, there is a different kind of response that might be given to the original problem that will be briefly flagged in this final session. This is to construe the Problem of Other Minds as itself stemming from a dubious picture of the epistemological situation. Once this is rethought, there may be no problem remaining to be solved. (There may still be much neurological work to be carried out on how it is possible to hear meaning in others' speech or mental states in their behaviour. However, these explanations will be deployed using lower-level terms.)

In Heal's original paper, this possibility was suggested in her response to an anticipated criticism: that simulation required some other way of determining the initial conditions from which one runs one's simulated development. She suggested that it might be possible to read this initial state *without* the use of a theory that translated behavioural responses—thinly described as mere movements—into underlying mental states. She also cited McDowell in support of this view. The focus of discussion in this, final, session is therefore a paper by McDowell who puts forward just such a view in the context of re-evaluating Wittgenstein's use of the idea of criteria.

The standard view of criteria for other minds again

In 'Criteria, defeasibility, and knowledge' (1982) McDowell introduces a line of thinking about the Problem of Other Minds, of the relation between behaviour and mental states, and the role of Wittgenstein's notion of criteria here that generalizes to have important repercussions in epistemology more generally. Getting clear on the particular case of our epistemic standing with respect to other minds will clarify what happens when we open our eyes to the world in other cases, or so the hope goes. This connection makes McDowell's paper very important but it also helps to explain why it is more difficult than the others in this chapter.

McDowell begins by attempting to cast doubt on the coherence of the more widely held interpretation of criteria as a priori, conventional, but defeasible indicators or conditions of some (underlying) fact. The influential Wittgenstein exegete P.M.S. Hacker, writing in the *Oxford Companion to Philosophy*, defines a criterion thus:

A standard by which to judge something; a feature of a thing by which it can be judged to be thus and so. In the writings of the later Wittgenstein it is used as a quasi-technical term. Typically, something counts as a criterion for another thing if it is necessarily good evidence for it. Unlike inductive evidence, criterial support is determined by convention and is partly constitutive of the meaning of the expression for whose application it is a criterion. Unlike entailment, criterial support is characteristically defeasible. Wittgenstein argued that behavioural expressions of the 'inner', e.g. groaning or crying out in pain, are neither inductive evidence for the mental (Cartesianism), nor do they

entail the instantiation of the relevant mental term (behaviourism), but are defeasible criteria for its application.

Key features of this definition are that the criteria of, for example, an 'inner' state such as pain are fixed by convention and are partly constitutive of what we mean by pain. Thus groaning and crying out are not mere symptoms but rather part of what we understand by pain. They are connected by definition not induction.

EXERCISE 8   (10 minutes)

On the standard view of criteria, the behavioural criteria for mental states provide some support, as a matter of meaning, for the ascription of mental states; however, they are also defeasible. Sometimes the criteria are satisfied when the subject is not in the relevant mental state. Before going on think whether this allows one ever to have knowledge of others' mental states? Think about the requirements one might make for someone in general to have knowledge (e.g. the analysis of knowledge as justified true belief). Can the criterial approach fit our normal beliefs about whether we have knowledge?

McDowell's objection to the standard view of criteria for other minds

McDowell's main objection is this. If a criterion can be satisfied while the underlying fact for which it is a criterion does not obtain, then knowing that the criterion is satisfied cannot legitimate (or legitimate 'criterially') the claim that one knows that the fact obtains. The problem is that if there is the slack between criteria and facts which the standard reading supposes then criteria cannot ever be *sufficient* for knowledge, whether or not there are conventions which supposedly declare this to be so.

McDowell attempts to bring out just what is wrong with this idea in the following thought experiment:

Consider a pair of cases, in both of which someone competent in the use of some claim experiences the satisfaction of (undefeated) 'criteria' for it, but in only one of which the claim is true. According to the suggestion we are considering, the subject in the latter case knows that things are as the claim would represent them as being; the subject in the former case does not . . . However, the story is that the scope of experience is the same in each case: the fact itself is outside the reach of experience. And experience is the only mode of cognition—the only mode of acquisition of epistemic standing—that is operative; appeal to theory is excluded . . . How can a difference in respect of something conceived as cognitively inaccessible to both subjects . . . make it the case that one of them knows how things are in that inaccessible region while the other does not—rather than leaving them both, strictly speaking, ignorant on the matter?   (pp. 459–460)

(Appeal to *theory* is excluded because criteria are supposed to be conventional and a priori rather than a posteriori and contingent.)

The problem is the wedge between appearance and reality

The problem that McDowell picks out is that the standard view of criteria puts a wedge between experience and the knowledge that such experience is supposed to warrant or justify. Thus even in favourable circumstances where the fact that the criterion is a criterion for obtains, it seems that experiencing the criterion cannot make the right sort of difference to constitute knowledge as opposed to accidentally true belief.

If experiencing the satisfaction of 'criteria' does legitimize ('criterially') a claim to know that things are thus and so, it cannot also be legitimate to admit that the position is one in which, for all one knows, things may be otherwise. But the difficulty is to see how the fact that 'criteria' are defeasible can be prevented from compelling that admission; in which case we can conclude, by contraposition, that experiencing the satisfaction of 'criteria' cannot legitimize a claim of knowledge. How can appeal to 'convention' somehow drive a wedge between accepting that everything that one has is compatible with things not being so, on the one hand, and admitting that one does not know that things are so, on the other?   McDowell (1982, p. 458; 1998, pp. 372–373)

(This objection resembles McDowell's claim that a conception of mental states as free-standing internal states makes their having a bearing on the world—possessing content—deeply mysterious. That argument was discussed in Chapter 25.)

The origin of the normal view of criteria

The main exegetical argument for the standard view of criteria is that Wittgenstein suggests that whether something is a criterion for something depends on the context or the particular circumstances. The standard view is that the background context can undermine a criterion (for some fact) that is satisfied. If so, then the criterion is defeated. However, McDowell suggests an alternative view in which the context determines whether some condition really is a criterion for a fact. What depends on context is not whether a satisfied criterion is defeated or not, but whether a condition really is a criterion in the first place. If it is, then it cannot be defeated by circumstances or context. Of course, what is at first taken for a criterion may turn out not to be.

McDowell supports this interpretative possibility by considering a passage in which Wittgenstein discusses criteria in a non-mental context.

The fluctuation in grammar between criteria and symptoms makes it look as if there were nothing at all but symptoms. We say, for example: 'Experience teaches that there is rain when the barometer falls, but it also teaches that there is rain when we have certain sensations of wet and cold, or such-and-such visual impressions.' In defence of this one says that these sense-impressions can deceive us. But here one fails to reflect that the fact that the false appearance is precisely one of rain is founded on a definition.   Wittgenstein (1953, §354)

Wittgenstein here rejects the temptation to say that both the fall of a barometer and also sensations of wet and cold (or visual impressions) are mere *symptoms* of rain. Instead, and by contrast with the barometer fall, the connection between the sensations (or the visual impressions) and rain is definitional or criterial. They are used in an explanation of what 'rain' means. This thought can, however, be interpreted in two ways.

> Commentators often take this to imply that when our senses deceive us, criteria for rain are satisfied, although no rain is falling. But what the passage says is surely just this: for things, say, to look a certain way to us is, as a matter of 'definition' (or 'convention'…), for it to look to us as though it is raining; it would be a mistake to suppose that the 'sense-impressions' yield the judgement that it is raining merely symptomatically—that arriving at the judgement is mediated by an empirical theory. That is quite compatible with this thought… when our 'sense-impressions' deceive us, the fact is not that criteria for rain are satisfied but that they *appear* to be satisfied.
> McDowell (1982, p. 466; 1998, 381)

Similarly in the case of criteria for mental states, pretence can make it *seem* that the criteria for pain, for example, are satisfied when, in fact, they are not.

Appearances and the Problem of Other Minds

This idea is clarified in McDowell's response to a further argument for the standard view, which draws explicitly on assumptions about psychological descriptions central to this chapter. Quoting a case where Wittgenstein suggests that the criterion for someone else experiencing a mental image of redness is what they say and do, McDowell offers the following diagnosis of the usual interpretation. We ascribe mental states to other people on the basis of behavioural conditions that can be detected *independently* of (and antecedent to) their mental states. However, this does not follow from what Wittgenstein here says. The claim that one can tell what state someone is in by what they say and do does not imply that the latter is a 'condition that one might ascertain to be satisfied by someone independently of knowing that he has [e.g.] a red image' (p. 465). Nor is it the case that a specification of such a criterion gives a general recipe for generating other examples of criteria: namely anything which meets that specification. To repeat, criteria depend on *particular* circumstances.

McDowell suggests that the traditional view is of a piece with the Argument from Illusion, which is often used to ground a form of general scepticism. Instead of appearances in general, we here have bodily behaviour and instead of underlying reality in general, we have mental states, but the basic idea is the same. The latter is supposed to be epistemologically inaccessible in a way in which the former is not and the challenge is to try show how judgements about the former could ever warrant judgements about the latter. Rather than being a device deployed within this framework, McDowell suggests that criteria are a part of Wittgenstein's rejection of the sceptic's framework. We do not satisfy ourselves of the criteria for another's mental states as a route to finding out about their states. 'This flouts an idea that we are prone to find natural, that a basis for a judgement must be

something on which we have a firmer cognitive purchase than we do on the judgement itself; but although the idea can seem natural, it is an illusion to suppose that it is compulsory.' (p. 471)

Appearances and the Argument from Illusion

The third section of McDowell's paper draws some broader connection between this view of criteria and epistemology more generally. The Argument from Illusion is often taken to suggest the following picture of experience. As appearances can sometimes be deceptive, sometimes when it looks that such and such it is not in fact the case, although things seem just as they do when such and such is the case. Thus what is taken in in experience is the same in both cases: a mere appearance that is not the same as, and stops short of, the fact. He labels this the 'highest common factor' conception. It places a veil of ideas (in this case: appearance) between ourselves and the world. In its place, he suggests that the fact that experience is fallible only justifies the claim that experience is either of merely an appearance or it really is of the full fact: the disjunctive conception.

The main argument against the highest common factor conception is that it 'undermines the very idea of an appearance having as its content that things are thus and so in the world "beyond" appearances' (p. 474). Once one reifies appearances, then there being appearances of something becomes mysterious. Furthermore, our ability to have knowledge of the world is undercut because the facts we aim to know always remain 'blankly external' to our experiences.

What makes the highest common factor conception attractive in the first place is the following thought. Whether one is justified in a judgement cannot depend on something that could vary independently of how things are subjectively. This leads to the idea that one's 'epistemic entitlement ought to be something one could display for oneself… from within;… non-question-begging[ly] from a neutrally available starting point' (p. 475). McDowell argues that the first requirements can be met by the disjunctive conception while rejecting the second: 'When someone has a fact made manifest to him, the obtaining of the fact contributes to his epistemic standing on the question. But the obtaining of the fact is precisely not blankly external to his subjectivity, as it would be if the truth about that were exhausted by the highest common factor.' (p. 476)

This is because external facts can be taken in in veridical experience, so there is no need to attempt to fill out an account of justification, of reasons for belief, which make use of merely internal matters.

Finally, McDowell goes on to suggest that the Problem of Other Minds depends on reading the highest common factor conception of experience on to the gap between behaviour and minds. It turns both on thinking of persons as a product of the accessible behaviour of material bodies and inaccessible minds. Human bodies replace human beings (p. 469). This is in turn premissed on the idea that human behaviour cannot be considered in itself expressive, any more than the behaviour of planets. However, once one abandons the highest common factor conception, other minds become accessible

in their direct expression in behaviour. The behaviour does not merely license an *inference* to underlying mental states.

Summing up

We began the section with the origins of the Problem of Other Minds in Cartesian philosophy. But, although the Cartesian picture of the mind as a private inner realm observed only by a single person makes the Problem of Other Minds obvious, some such problem can seem mere common sense. We all want to say at some times in our lives that we are not understood by others, that they haven't grasped what things are like for us. Equally, we sometimes find it difficult to gauge other people's states of mind. It can be tempting to conclude informally, whether or not we know anything about Cartesian philosophy, that minds are private, hidden behind behaviour.

Such a thought, however, exaggerates the everyday Problem of Other Minds, allowing it to inflate into the philosophical Problem of Other Minds. It ignores much of our everyday experience of taking account of other people's mental states immediately, transparently. It ignores what McDowell calls the *expressive* character of behaviour. Even people who say when asked that minds are private, do not actually take that approach in practice. No one takes the Cartesian perspective seriously in living with others.

Whatever exactly the right account is of the source of our knowledge here, the one thing nearly all philosophers now agree on is that one should not begin an account of the mind from an individual first-person perspective. There is some difference in whether one should think of a third-person perspective, as in their different ways both Norman Malcolm and the theory theorists do, or as a shared first-person perspective as the simulation theorists and McDowell do. But the Cartesian perspective of a solitary, solipsistic ego is a hopeless non-starter.

But how does that leave the issue of the empirical findings from autism? Here much depends on what the role of a philosophical theory is. What is its explanatory force or role? If a theory-theory, what is the nature of our knowledge of it? If simulation theory, do we have unconsciously to simulate? Once this is settled there is then the further question of how deficit studies impact on normal cases. What is the sense in which complex abilities can be broken down into components? Is this a kind of logical analysis or does it actually reveal inner functional modules? Again, empirical work has a bearing on our understanding here, but no less than our getting clear on the concepts employed.

Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What other options are there for solving the Problem of Other Minds?
2. What role can behaviour play? How is its description different from the description presupposed by theory-theory?

Reading guide

- A good place to start is the discussion of third-person access in Burwood, Gilbert, and Lennon (1999) *Philosophy of Mind*.
- A good historical approach is provided in Avramides (2000) *Other Minds*.

Theory-theory versus simulation theory

Theory-theory and simulation theory have recently been discussed as rival theories in a number of collections of papers:

- Davies and Stone's (ed.) (1995a) *Folk Psychology: a guide to the theory of mind debate*, and their (1995b) *Mental Simulation: evaluations and applications*.
- Carruthers and Smith (ed.) (1996) *Theories of Theories of Mind*.
- Hobson's (1991) *Against the 'Theory of Mind'*, and his (2002) *The Cradle of Thought*, provides important critiques.
- Stich, S. and Nichols, S. (2003) Folk psychology.

Direct access

- McDowell's account of direct experience of other people's minds is developed from a discussion of Wittgenstein. The most accessible source of this is in Wittgenstein (1972) *The Blue Book and Brown Books*.
- A traditional view of what Wittgensteinian meant by behavioural criteria for other minds is developed by the American philosopher Norman Malcolm in book length form in: Malcolm (1971) *Problems of Mind: Descartes to Wittgenstein*.
- And by the Oxford Wittgensteinian Peter Hacker (1972) in *Insight and Illusion*.
- For a number of essays arguing for a different view but in the same general area as McDowell, try Wright (ed.) (1986) *Realism Meaning and Truth*. This contains essays by Wright on Other Minds, on the use of theories to explain knowledge of a language and also epistemology more generally. It is quite difficult.
- For a different perspective try Glendinning (1998) *On Being with Others: Heidegger, Derrida, Wittgenstein*.

Empirical work

For work on the theory of mind and autism among other things see:

- Baron-Cohen (1995) *Mindblindness: an essay on autism and theory of mind*, and (1993) (ed.) *Understanding Other Minds: perspectives from autism*.
- Eilan *et al.* (ed.) (2005) *Joint Attention—Communication and Other Minds: issues in philosophy and psychology (consciousness & self-consciousness S.)*.

◆ Gordon and Barker (1995) Autism and the theory of mind debate (in Graham and Stevens (ed.) *Philosophical Psychopathology*).

References

Avramides, A. (2000). *Other Minds*. London: Routledge.

Baron-Cohen, S. (ed.) (1993). *Understanding Other Minds: perspectives from autism*. Oxford: Oxford University Press.

Baron-Cohen, S. (1995). *Mindblindness: an essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Burwood, S., Gilbert, P., and Lennon, K. (1999). *Philosophy of Mind*. London: UCL.

Carruthers, P. (1996). Autism as mind-blindness. In *Theories of Theories of Mind* (ed. P. Carruthers and P. Smith). Cambridge: Cambridge University Press, 257–273.

Carruthers, P. and Smith, P.K. (ed.) (1996). *Theories of Theories of Mind.* Cambridge: Cambridge University Press.

Chihara, C.S. and Fodor, J.A. (1991). Operationalism and ordinary language: a critique of Wittgenstein. In *The Nature of Mind* (ed. D. Rosenthal). Oxford: Oxford University Press, 137–150.

Currie, G. (1996). Simulation-theory, theory-theory and the evidence from autism. In *Theories of Theories of Mind* (ed. P. Carruthers and P. Smith). Cambridge: Cambridge University Press, 242–256.

Davies, M. and Stone, T. (ed.) (1995a). *Folk Psychology: a guide to the theory of mind debate*. Oxford: Blackwell.

Davies, M. and Stone, T. (ed.) (1995b). *Mental Simulation: evaluations and applications*. Oxford: Blackwell.

Dennett, D. (1987). True believers: the intentional strategy and why it works. In *The Intentional Stance*. Cambridge. MA: MIT Press, pp. 13–35.

Descartes, R. (1996). *Meditations on First Philosophy*. Cambridge: Cambridge University Press.

Eilan, N., Hoerl, C., McCormack, T., and Roessler, J. (ed.) (2005). *Joint Attention—Communication and Other Minds: issues in philosophy and psychology* (*consciousness & self-consciousness S.*) Oxford: Oxford University Press.

Fodor, J.A. (1991). Propositional attitudes. In *The Nature of Mind* (ed. D. Rosenthal). Oxford: Oxford University Press, pp. 325–338.

Fodor, J.A. *Psychosemantics*. Cambridge, MA: MIT.

Gipps, R. (ed.) (2004). *Autism and Intersubjectivity* (Special issue). *Philosophy, Psychiatry, & Psychology*, 11(3).

Glendinning, S. (1998). *On Being With Others: Heidegger, Derrida, Wittgenstein*. London: Routledge.

Gordon, R.M. (1995a). Folk psychology as simulation. In *Folk Psychology* (ed. M. Davies and T. Stone). Oxford: Blackwell, pp. 60–73.

Gordon, R.M. (1995b). Simulation without introspection or inference from me to you. In *Mental Simulation* (ed. M. Davies and T. Stone). Oxford: Blackwell, pp 53–67

Gordon, R.M. and Barker, J.A. (1995). Autism and the theory of mind debate. In *Philosophical Psychopathology* (ed. G. Graham and G.L. Stephens). Harvard, MA: MIT Press, 163–181.

Hacker, P.M.S. (1972). *Insight and Illusion*. Oxford: Oxford University Press.

Hacker, P.M.S. (1995). Criteria. In *The Oxford Companion to Philosophy*, (ed. T. Honderich). Oxford: Oxford University Press.

Heal, J. (1995). Replication and functionalism. In *Folk Psychology* (ed. M. Davies and T. Stone). Oxford: Blackwell, pp. 45–59.

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

Hobson, R.P. (2002). *The Cradle of Thought*. London: MacMillan.

Malcolm, N. (1958). Knowledge of other minds. *Journal of Philosophy*, 55 pp. 969–978. (Reprinted in Rosenthal, D (ed.) (1991). *The Nature of Mind*. Oxford: Oxford University Press, pp. 92–97.)

Malcolm. N. (1971). *Problems of Mind: Descartes to Wittgenstein*. London: Allen & Unwin.

McDowell, J. (1982). Criteria, defeasibility, and knowledge. *Proceedings of the British Academy*, 68 pp. 455–479 reprinted in Mc Dowell, J. (1998). Meaning knowledge and reality. Cambridge, MA: Harvard University Press, pp. 369–394.

Russell, B. (1948). Analogy (1948). In *Human Knowledge: its scope and limits* London: Allen & Unwin. London: Allen & Unwin. (Reprinted in Rosenthal, D (ed.) (1991). *The Nature of Mind*. Oxford: Oxford University Press, pp. 89–91.)

Stich, S. and Nichols, S. (2003). Folk psychology in The Blackwell Guide to the Philosophy of Mind (ed S. Stich and T. Warfield). Oxford: Blackwell, pp. 235–255.

Stone, T. and Davies, M. (1996). The mental simulation debate: a progress report. In *Theories of Theories of Mind* (ed. P. Carruthers and P. Smith). Cambridge: Cambridge University Press, 119–137.

Williams, M. (1996). *Unnatural Doubts*. Princeton. Princeton University Press.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Basil Blackwell.

Wittgenstein, L. (1972). *The Blue Book and Brown Books*. Oxford: Blackwell.

Wright, C. (ed.) (1986). *Realism Meaning and Truth*. Oxford: Blackwell.