

PHILOSOPHICAL LECTURE

The Mental Simulation Debate

MARTIN DAVIES

Corpus Christi College, Oxford, OX1 4JF

Introduction

IN A WELL-KNOWN SECTION from Chapter 3 of *Individuals*, Strawson (1959, pp. 99–100) says:

it is a necessary condition of one's ascribing states of consciousness, experiences, to oneself, in the way one does, that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself.

... the ascribing phrases are used in just the same sense when the subject is another as when the subject is oneself.

[But] how could the sense be the same when the method of verification was so different in the two cases ...

Strawson's example is the predicate 'is in pain'; but what he says goes equally for 'believes that penguins waddle'. It is a necessary condition of self-ascription:

I believe that penguins waddle

that I can also ascribe beliefs to others:

She believes that penguins waddle.

The predicate is used in the same sense in both cases — it is a single concept that has been mastered — despite the fact that the grounds for attribution are liable to be very different. In the first-person case, Strawson says (1959, p. 100), we would scarcely speak of there being a 'method of verification'; but in the third-person case, it is natural to

say, the grounds for attribution are provided by the observable situation and behaviour of the other.

This difference between the first-person and third-person cases, as regards grounds of attribution (or 'method of verification') might show up in a philosophically fundamental account of our ascriptions of beliefs — in an account of what constitutes mastery of the concept of belief — as a difference between two clauses in that account. But the account as a whole should honour the idea that it is just the same thing that is being predicated in the two cases. In particular, the account should guarantee, somehow, that where first-person and third-person attributions converge upon a single subject of predication the two attributions are true under just the same conditions. For it would be impossible to maintain that the predicate 'believes that penguins waddle' retains the same sense in both first-person and third-person attributions, if we could not rule out the possibility that

I believe that penguins waddle
might be true, say, in my mouth, while

He believes that penguins waddle
could be false when said of me.

This requirement — which Peacocke (1992, p. 171) calls the requirement of *referential coherence* — is surely a constraint upon any philosophically fundamental account of belief attribution (of our deployment of the concept of belief).

In an oft-quoted passage in *Word and Object*, Quine recommends the idea that third-person attributions of belief involve a kind of dramatic re-enactment (1960, p. 92):

We project ourselves into what, from his remarks and other indications, we imagine the speaker's state of mind to have been, and then we say what, in our language, is natural and relevant for us in the state thus feigned.

This is the idea — with its stress upon the ability to imagine the state of mind of the other — that has been developed into the claim that a fundamental account of belief attributions should be cast in terms of mental simulation: simulating in imagination the state of mind of the other.

Quine's idea and Strawson's constraint date from 1960 and 1959 respectively. As we shall see, the question whether it is possible to

meet the constraint while adopting the idea is still — over thirty years later — in need of an answer.

The idea of mental simulation — understanding another by imagining his or her state of mind — is not a new idea. Around the middle of this century — in the work of Collingwood (1946), for example — it had its home in disputes in the philosophy of social sciences and of history. These were disputes about whether the methods that were successful in explaining and predicting the physical world of material objects would also be appropriate to explanations and predictions of the human world.

The current form of the debate can be dated, not entirely arbitrarily, to 1986. For in that year, two key papers appeared: 'Replication and functionalism' by Jane Heal, and 'Folk psychology as simulation' by Robert Gordon.

The present paper has four parts. First, I shall sketch the two opposed positions in the current form of the mental simulation debate: the *theory theory* and the *simulation alternative*. Second, I shall give a brief review of some of the empirical literature — though it has to be said at the outset that, in my view, the empirical strand of the debate is at a rather inconclusive stage. In the third section, I shall focus upon the particular development of the simulation theory that has been offered by Alvin Goldman (1989; 1992). I shall use this exposition both as a way of motivating a particular way of delineating the two opposed views, and as a way of introducing the idea that simulation cannot be employed to give a philosophically fundamental account of our use of mental concepts. In the final section, I shall raise the question whether Robert Gordon's (1986; 1992a; 1992b; 1992; in press) development of the simulation view presents a more promising prospect.

1. The debate in outline

The friends of mental simulation are united by what they oppose; namely, the so-called 'theory theory' about commonsense (or 'folk') psychology. Let us begin there.

1.1 The theory theory

Suppose we ask: How are normal adult human beings able to negotiate a particular psychological task domain? How, for example, is a normal

adult able to judge whether a sentence of his or her natural language is grammatical? One strategy for answering this kind of question is to postulate that human beings are in possession of a theory — in the linguistic case, a grammar — and that they deploy this theory (grammar) in order to solve the problems that the task domain presents.

It is not reckoned to be any block on this strategy that the person who is supposed to be deploying the theory is unaware of the theory, nor that — in the vast majority of cases — the deployer of the theory will never be conversant with the conceptual resources that are required to state the theory explicitly. The normal adult's relationship to the theory is said to be that of 'tacit' or 'implicit' knowledge. This 'tacit theory' strategy is certainly one of the dominant paradigms in contemporary cognitive science.

Now, one impressive fact about human beings is that, with little or no formal training, they develop the capacity to deploy psychological concepts such as *belief* and *desire* in predictions and explanations of the actions and mental states of other members of the species. These predictions and explanations are said to *rationalize* the subject's actions or mental states; they present the subject's beliefs and desires as providing him or her with reasons for acting and thinking in certain ways. What is the nature of the basis of this ability?

In line with that dominant paradigm in cognitive science, many philosophers and psychologists argue that this everyday ability reflects the fact that normal adult human beings possess a primitive or 'folk' psychological theory. This theory, just like its less primitive cousins in more developed domains, postulates theoretical entities — in this case, mental states — and contains laws which relate the mental states to one another and to external stimuli (on the input side) and actions (on the output side). When I predict what someone will do, or explain why they have done something, I do so by deploying this theory. Most of us are, of course, quite unaware that this is what we are doing; but, as with grammatical theory, that fact is reckoned to be unimportant. Our relationship to the psychological theory is allowed to be 'tacit' or 'implicit' knowledge.

This view about our actual practice of predicting and explaining mental states — a very general empirical claim about the basis of that ability — is one strand of what has come to be known as the *theory theory* about the propositional attitudes of commonsense psychology.

But the theory theory has other strands too — clearly empirical psychological strands, and also more narrowly philosophical strands.

One of these strands is the claim that the best philosophical elucidation of the nature of the propositional attitudes is provided by a psychological theory's specification of the interconnected roles of the various attitudes (their causal or functional roles according to that theory). A further philosophical strand is the claim that an individual's mastery of the concepts of the propositional attitudes is precisely constituted by his or her tacit knowledge of such a theory embedding those notions. It is important to distinguish these two strands: one concerns mental states while the other concerns our concepts of mental states.

There are other empirical psychological strands, too — alongside the initial claim about our normal adult predictive and explanatory practices. Most importantly, these concern the nature of the psychological development that leads to the attained state of tacit mastery of a folk psychological theory. The theory theorist says that — by learning or by maturation — the various principles of an increasingly sophisticated psychological theory become available for deployment in prediction and explanation.

There may be yet further strands, but it is worth keeping in mind at least these four:

Psychological

- a strand that concerns the attained state of normal adults;
- a strand that concerns the course of development;

Philosophical

- a strand that concerns the elucidation of the nature of mental states;
- a strand that concerns the mastery of mental concepts.

1.2 The simulation alternative

Robert Gordon — in a series of papers beginning with the one that we have already mentioned: 'Folk psychology as simulation' (1986) — and Alvin Goldman — in a series beginning with 'Interpretation psychologized' (1989) — have mounted a challenge to the idea that folk psychology is best seen as a theory. These philosophers deny that our understanding of one another primarily proceeds by deployment of a theory, whether folk or otherwise. Rather, they claim that

human beings are able to predict and explain each others' actions by using the resources of their own minds to simulate the psychological aetiology of the actions of others. So, instead of being theorizers, we are *simulators*. We are mental simulators, not in the sense that we merely simulate mentation, but in the sense that we understand others by using our own mentation in a process of simulation.

Gordon begins by noting the prodigious ability we have to predict our own actions. He notes that before I act I often engage in practical reasoning, and suggests that to predict what I might do in a hypothetical situation I can 'simulate the appropriate practical reasoning' by engaging 'in a kind of pretend play'. Gordon fills this out with the following example (1986, p. 161):

I imagine, for instance, a lone modification of the actual world: the sound of footsteps from the basement. Then I ask, in effect, 'what shall I do now?' And I answer with a declaration of immediate intention, 'I shall now...' This too is only feigned. But it is not feigned on a *tabula rasa*, as if at random: rather, the declaration of immediate intention appears to be formed in the way a *decision* is formed, *constrained* by the (pretended) 'fact' that there is a sound of footsteps from the basement, the (*unpretended*) fact that such a sound would now be unlikely if there weren't an intruder in the basement, the (*unpretended*) awfulness of there being an intruder in the basement, and so forth.

Gordon then extends this strategy to the third person case, where he says (p. 162):

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, this is not the same as deciding *what I myself would do*. One tries to make *adjustments for relevant differences*.

So, according to Gordon, my ability successfully to predict the actions of another relies upon the ability to engage in simulation. And this is not merely to simulate that *I* am in the other's situation, but to simulate *being* in that situation, having those psychological traits, and so on; in short, to simulate *being the other*.

It is important to note that the outcome of the pretence or simulation is not supposed to be determined by any theory that I hold; if it were then the gap between the theory theory and the simulation alternative would be narrow indeed. Rather — on one plausible elaboration of the basic idea of simulation — I feed the pretend inputs into my own decision-making processes and let those processes run 'off-line', so that no action is actually produced. I then announce the decision

that is produced as output by those 'off-line' processes. Goldman spells out the idea of a process-driven (rather than a theory-driven) simulation quite clearly (1989, p. 173):

But must all mental simulations be theory-driven in order to succeed? I think not. A simulation of some target systems might be accurate even if the agent lacks such a theory. This can happen if (1) the *process* that drives the simulation is the same as (or relevantly similar to) the process that drives the system, and (2) the initial states of the simulating agent are the same as, or relevantly similar to, those of the target system. Thus, if one person simulates a sequence of mental states of another, they will wind up in the same (or isomorphic) final states as long as (A) they began in the same (or isomorphic) initial states, and (B) both sequences were driven by the same cognitive process or routine. It is not necessary that the simulating agent have a theory of what the routine is, or how it works.

This idea of a system of decision processes that can be operated 'off-line' is integral to Goldman's version of the simulation view. Gordon, in contrast, regards it as an 'ancillary hypothesis', though a 'very plausible' one (1992b, p. 87), and chooses to focus upon the less scientific-sounding idea of imaginative identification. These differences within the simulation camp are not unimportant; but at this introductory stage perhaps we can fairly say that, so far as the mechanisms of prediction and explanation go, the basic proposals of Gordon and Goldman are fairly similar.

It is crucial, however, even at this introductory stage, to impose upon the simulation view a differentiation of various strands, corresponding to the several distinct strands in the theory theory. For a friend of simulation might or might not move from a claim about normal adult processes of prediction and explanation to more purely philosophical claims about the elucidation of the nature of mental states, or about the conditions for the possession of mental concepts.

All manner of hybrid positions are at least *prima facie* possible. For example, suppose someone holds that mental simulation provides a useful — generally reliable — heuristic for the generation of hypotheses about how someone will behave, or about why someone has acted in a particular way. It is quite consistent to combine that view with the claim that mastery of the mental concepts used in those hypotheses is a matter of tacit knowledge of a theory, which in turn provides the resources for normative evaluation of the generated hypotheses.

Gordon's position is that simulation provides much more than an

effective heuristic. Indeed, he enters the bold suggestion that the simulation view will yield an alternative to the theory theorist's understanding of mental states themselves; or rather, as he puts it (1986, p. 166), will provide 'a way of interpreting ordinary discourse about beliefs'.

In contrast, Goldman is not optimistic about the prospects of the simulation view yielding a constitutive account either of mental states or of the possession conditions for mental concepts. However, Goldman does not see this as calling for the re-introduction of strands of the theory theory, for what the theory theory offers is essentially a third-person account of mental states. The simulation view already addresses the case of third-person attribution of mental states, but only by taking for granted the first-person case. So, Goldman concludes (1989, p. 183), 'If the simulation theory is right . . . it looks as if the main elements of the grasp of mental concepts must be located in the first-person sphere'.

I shall return to Goldman in Section 3, and to Gordon in Section 4. In the meantime, however, we should take account of the empirical strands of the mental simulation debate — particularly, the strand that concerns the course of development. There too, as we shall see, there has been a kind of devolution upon the first-person case.

2. The developmental evidence

One of the four strands that we discerned within the theory theory concerns the developmental course leading up to the normal adult's attained state of possession of a folk psychological theory. In a version that most closely parallels the case of linguistic theory, this strand would say that tacit knowledge of the theory is innate, that it is embodied in a special-purpose module of the mind, and that development is predominantly a matter of maturation rather than of learning. Of course, other versions are certainly conceivable: someone might deny the nativism; someone else might reject the idea of a task-specific system.

It is strictly speaking consistent to combine a theory theorist's account of the attained state with a mainly simulation theoretic account of the course of development. For example, the principles of the theory might be extracted, by some kind of overseeing system, from the operation of a developed system for off-line simulation (cf. Goldman, 1989, p. 176). But, despite that possibility, it is natural to start out with

a simplifying background assumption to the effect that there is more continuity between the stages of development and the attained state. Given such a working hypothesis, a simulation theorist or a theory theorist can use empirical evidence in favour of the developmental strand of his or her overall view in order to reinforce the strand that concerns the attained state.

There is a very considerable body of research in developmental psychology, addressing the question whether the child's development of the capacity to deploy psychological concepts such as belief and desire to predict and explain human actions reflects the gradual development of a theory or the development of some different kind of ability. Does the child develop a theory of the mind or, rather, develop the ability to imagine what it is like to be in another person's shoes: to pretend or to simulate? Both Gordon and Goldman call upon this research in support of the simulation view. Indeed, they both cite two classic studies involving false beliefs (Wimmer and Perner, 1983; Baron-Cohen, Leslie and Frith, 1985).

2.1 The false belief task

Gordon gives a simple description of the task used by Wimmer and Perner (Gordon, 1986, p. 168):

[T]he puppet-child Maxi puts his chocolate in the box and goes out to play. While he is out, his mother transfers the chocolate to the cupboard. Where will Maxi look for the chocolate when he comes back?

and of the key result:

In the box, says the five-year-old, pointing to the miniature box on the puppet stage: a good prediction of the sort we ordinarily take for granted. . . . But the child of three to four years has a different response: verbally or by pointing the child indicates the cupboard.

The simulation theorist's account of the difference between the performance of the younger and the older children is, of course, that the younger children as yet lack, while the older children have attained, the ability to engage in a kind of pretend play that involves simulating being someone whose cognitive position is substantially different from one's own. This account is then fortified by the finding of Baron-Cohen, Leslie and Frith that autistic children — who are typically poor at pretend play — make the same error as the younger children in the Wimmer and Perner study.

Gordon and Goldman are joined in their defence of the simulation view by Paul Harris (1989; 1991a; 1992), who offers a range of support for its developmental strand. Harris also notes (1992, pp. 121–4) that the false belief task — used in the experiments reported by Wimmer and Perner (1983) and much subsequent work — was originally introduced to make it impossible for a subject using a very simple simulation strategy to achieve predictive success. The simple simulation strategy — which Gordon (1992a, p. 13) calls ‘total projection’ — is the one that does not make ‘adjustments for relevant differences’ — particularly, for relevant cognitive differences. If a task — involving the prediction of behaviour, for example — can be performed by using that simple strategy then, of course, a creature’s success on that task does not constitute evidence of the creature’s possession of a psychological theory — a ‘theory of mind’. Researchers seeking evidence of a ‘theory of mind’ therefore began using the false belief task; for success on that task certainly requires *something* more sophisticated than total projection.

Now, whether the extra that is needed is deployment of a theory, or just more sophisticated simulation, is controversial. So, success on the false belief task is not clinching evidence on behalf of the theory theory. But, Harris’s point is that, given the original motivation for the false belief paradigm, we should not really expect it to be the starting point for results that are especially congenial to the simulation theory and problematic for the theory theory.

It seems fair to say that both the advocates of the theory theory (Perner, 1991; Perner and Howes, 1992; Gopnik and Wellman, 1992) and the friends of mental simulation (Harris, 1989; 1991a; 1992) can provide acceptably principled accounts of the developmental data from the third-person false belief task. According to the theory theory, children undergo ‘a change from one mentalistic psychological theory to another somewhere between 2½ and around 4’ (Gopnik and Wellman, 1992, p. 149). According to the simulation account, ‘the changes . . . stem from changes in the child’s imaginative flexibility, rather than from a change in the child’s so-called theory of mind’ (Harris, 1992, p. 131).

In order to make further progress with the empirical, developmental strand of the debate, we need to look elsewhere; and the obvious place to look is the first person. This case holds out the prospect of distinguishing empirically between the two theories, since here they seem bound to yield different predictions. For the theory theory, the

first-person case is just one case among others; but the mental simulation account of third-person attribution of mental states works by taking first-person attributions for granted. Thus it seems that the theory theory predicts similar errors in both first-person and third-person attributions, while the simulation account predicts distinctive errors introduced by the use of simulation to cantilever out to the third person from the first person. (In fact, to be a little more accurate, we should say that the empirical psychological strands of the two theories seem bound to yield different predictions about actual attributions. The more purely philosophical strands — concerning the nature of mental states and the mastery of mental concepts — do not obviously yield any distinctive predictions about the actual practice of attribution.)

Harris himself suggests (1992, pp. 139–41) that what would be embarrassing for the theory theory would be results demonstrating a marked difference in accuracy as between first-person and third-person reports. And he notes that there are results — even in the work of those advocating the theory theory — that seem to reveal the crucial asymmetry between the first-person and third-person cases. Even very young children are very accurate, for example, at reporting what they are currently thinking, pretending, seeing, or wanting.

2.2 Focus on the first person

Alison Gopnik and Henry Wellman are advocates of the theory theory, and recommend that the course of development should be understood as a series of changes in the child's theory of mind. Some of the advantages that they claim for the theory theory relate to the absence of an asymmetry between first-person and third-person attributions which — they say, apparently in agreement with Harris — would be predicted by the simulation theory.

The assumption shared by Gopnik and Wellman, on the one hand, and Harris, on the other, is that the simulation theory is introspectionist about first-person attributions. In particular, the simulation theory predicts that reading off one's own mental states should be relatively easy, while attributing to a third person mental states different from one's own should be relatively difficult. In contrast, the theory theory predicts that the difficulty of the attribution task should correlate with a difference between kinds of mental state, which cross-cuts the first-person vs. third-person distinction. This is the difference in theoretical complexity

between rudimentary notions of desire and perception, and the fully representational notion of belief.

Gopnik and Wellman review evidence that suggests that children who fail on the third-person version of the false belief task also fail on a first-person version of the task. They offer this as evidence that is problematic for the simulation theory, since it reveals no asymmetry between first-person and third-person attributions.

However, Harris (1992, p. 132) comments that this first kind of evidence does not really threaten the simulation theory. For Gopnik and Wellman's first-person version of the false belief task involves a past-tense attribution. And — whatever may be the appeal of introspection as a method for present-tense first-person attributions — it is quite consistent for a simulation theorist to group past-tense first-person attributions with the third-person case, since the simulation would involve the subject having to set aside his or her own *present* beliefs. Once again, it is hard to find anything decisive, one way or the other, in the developmental data.

This is not the end of the empirical story. There are other findings that have been offered in support of the theory theory, and as problematic for the simulation view. For example, Gopnik and Wellman also describe evidence that suggests that children who fail when asked to make first-person attributions of past beliefs are nevertheless able to make correct first-person attributions of other past mental states, such as desires and perceptions. The challenge to the simulation theory is to explain this difference between mental states. Why is a child who is unable to set aside his or her present beliefs nevertheless able to set aside his or her present desires?

The simulation view also comes under attack from developmental data presented by Josef Perner and Deborah Howes. In their experiment, children are presented with a story much like that in the earlier experiments of Wimmer and Perner. But, instead of asking, 'Where will [the puppet-child] look for the chocolate when he comes back?', the experimenters now ask, 'Where does [the puppet-child] think the chocolates are?' The children in the experiment are old enough to answer this question correctly; but many of them give the wrong answer to a further question:

What if we go over to the park and ask John [the puppet-child]: 'John, do you know where the chocolates are?' What will he say?

That is, many of the children answer, 'No', to this question. Perner and Howes argue that this finding is difficult for the simulation view. If the first question is answered correctly as a result of the children simulating John, then the second ('self-reflective') question should be answered correctly, too, 'since it is part and parcel of a belief to be convinced that one knows where the object is' (1992, p. 76), or again, since 'the believer's reflective conviction that he is right is simply part and parcel of a conscious belief' (1992, p. 81).

These findings are suggestive, to be sure; but it is difficult to be convinced that the friend of simulation is totally devoid of a response to them. For example, the friend of simulation might raise a query about the precise questions that Perner and Howes asked the children in their experiment. The question that the children tend to answer incorrectly:

What if we go over to the park and ask John [the puppet-child]: 'John, do you know where the chocolates are?' What will he say?

asks the experimental subject to imagine him- or herself going over to the park and talking to John. By explicitly placing the subject in the imagined scene, the question may make it more difficult for the subject to engage in a simulation of John. This difficulty is not present in the question that the same children can answer correctly:

Where does John think the chocolates are?

So, a better comparison would be with the question:

What if we go over to the park and ask John: 'John, where are the chocolates?' What will he say?

This question was not asked in the experiment reported by Perner and Howes (1992).

There are other complications, too. We need to consider the inevitable role of theory where questions about knowledge are concerned. Even the most minimal grasp upon the concept of knowledge requires this much theory:

If it is not the case that p , then x (here, John) does not know that p .

A full evaluation would need to consider the possibility that this little

piece of theory may introduce some confusion into an attempt to use simulation to answer questions involving knowledge.

Furthermore, we need to take into account that not every advocate of the simulation view shares the assumption that seems to condition this aspect of the debate; the assumption, namely, that the simulation theory must be introspectionist (or 'Cartesian'; see Perner and Howes, 1992, p. 75) about (present-tense) first-person attribution. A full evaluation would need to assess whether the predictions attributed to the simulation view really do depend upon the assumption of introspectionism. (I return briefly to the issue of introspectionism in Section 4.1 below.)

3. Goldman and the interpretation strategy

Goldman's initial development of the simulation alternative to the theory theory takes place within a broader dialectical context: evaluating what he calls the 'interpretation strategy'. This is the strategy of studying our actual practice of interpretation — of attributing propositional attitudes and predicting and explaining actions in terms of attitudes — in order to extract the conditions of mentality. The idea behind the interpretation strategy is that a philosophically fundamental account — either of the nature of mental states, or of the conditions for possessing concepts of mental states — can be wrung from a description of our actual practice of attributing mental states.

The question with which the proponent of the strategy begins is then (Goldman, 1989, p. 162):

[H]ow does the (naive) interpreter arrive at his/her judgements about the mental attitudes of others?

Goldman considers three possible answers to this question:

- 1 theories of 'radical interpretation' based upon principles of charity and rationality — in the style of Davidson or Dennett;
- 2 the 'folk theory' approach to interpretation; and
- 3 the simulation alternative.

His overall argument is that the first two possible answers are inadequate; a simulation account of interpretation is to be preferred. But, whereas either the 'radical interpretation' or the 'folk theory' account of the actual practice of attribution could be reconfigured into a putative account of mental states themselves, the simulation

alternative cannot — Goldman argues — be incorporated into a philosophically fundamental account of mentality. Thus, the interpretation strategy runs into the sand.

This broader dialectical context is not my main concern in this paper; so I shall leave aside Goldman's concerns about principles of charity and rationality, and ask about his grounds for ruling out the 'folk theory' account of our interpretive practice. He launches three objections against this account. First (1989, pp. 166–7):

[A]ttempts by philosophers to articulate the putative laws or 'platitudes' that comprise our folk-theory have been notably weak.

Second (p. 167):

[W]hy, one wonders, should it be so difficult to articulate laws if we appeal to them all the time in our interpretive practice?

Third, there is a worry about children's acquisition of knowledge of the putative folk theory (p. 167):

Are such children [at the age of four, five, or six] sophisticated enough to represent such principles? And how, exactly, would they acquire them?

It is, of course, very striking that exactly analogous objections could be raised against the idea that our judgements of grammaticality are subserved by our having knowledge of a theory. Articulating the principles of universal grammar and the rules of particular languages is a far from easy task: doubtless some early attempts were 'weak' and characterized by vagueness and inaccuracy. We can ask why it is so hard to spell out these principles if we are using them all the time as we perceive and produce linguistic items. And we can also ask whether young children are sufficiently sophisticated to represent such principles, and how children could come to acquire such knowledge. But very few people would regard those objections as doing serious damage to the credentials of the 'theory theory' of grammaticality judgements, since in that domain the theory theory is a *tacit theory* theory.

The natural response, then, to Goldman's reasons for ruling out the 'folk theory' account of our interpretive practice is to allow that the 'folk theory' may be a tacitly known theory. Indeed, this is the way that I introduced the theory theory side of the debate (in line with what I described as a dominant paradigm in cognitive science). To be sure, this way of responding to Goldman's three objections carries with it an obligation: to give an account of tacit knowledge of theories

which is not trivial. In particular, tacit knowledge must be a notion that discriminates more finely than 'extensional equivalence'. But work in the philosophy of linguistics gives us some grounds for optimism here (Evans, 1981; Davies, 1986; 1987; 1989; Peacocke 1986; 1989). So, let us suppose that we can meet the obligation to give an account of tacit knowledge of theories, so that the tacit theory theory can remain as a candidate account of our actual practice of interpretation.

Now consider the simulation alternative as that is described by Goldman (1989, p. 168):

[T]hey ascribe mental states to others by pretending or imagining themselves to be in the other's shoes, constructing or generating the (further) state that they would then be in, and ascribing that state to the other.

I would like to suggest that there is more than one way to construe this proposal. In fact, I want to distinguish two construals corresponding to two rather different imaginative processes, and then to point out that one construal of the simulation alternative places it under more threat of collapse into the theory theory than does the other. This, of course, is intended to provide some motivation for pursuing that other construal of the simulation alternative.

3.1 The threat of collapse

The first construal of the simulation alternative has it that the simulator (pretending or imagining him- or herself to be in the other's shoes) imaginatively entertains hypotheses concerning mental states:

I believe that p
I desire that q

and then proceeds to a conclusion about a further mental state, or about an action, or about an intention to act:

I believe that r ; or
I V ; or
I intend to V

(which mental state, action, or intention to act is then ascribed to the other).

If we recall Goldman's distinction between 'theory-driven' and 'process-driven' simulation, we can see that this use of the imagination can count as a process-driven simulation provided that a certain con-

dition is met. This condition is that the processing in the simulator (leading from the entertainment in imagination of the hypothesis to the arrival in imagination at the conclusion) is the same as, or relevantly similar to, or isomorphic to — let us say, follows the contours of — the process being simulated (that is, the process in the other that leads from mental states to further mental states, actions, or intentions to act).

This construal of the simulation alternative places it under some threat of collapse into the tacit theory theory. The crucial point to notice is that the processing in the simulator could *also* follow the contours of the derivational structure of a proof of a conclusion about, say, an intention to act from premises about, say, the agent's beliefs and desires — a proof cast in a psychological theory. This matching of structure between simulation process and deductive derivation could be quite general. Wherever, for example, two proofs draw upon a common axiom in the theory, the two pieces of simulation might draw upon a correspondingly common cognitive mechanism. The threat of a collapse ensues because it is just this idea of matching of structure between a causal process and a derivational process that is used in some accounts of what it is for a cognitive processing system to embody tacit knowledge of a particular theory of the task domain. Roughly speaking, a component processing mechanism embodies tacit knowledge of a particular rule or axiom if it plays a role in mediating causally between representational states that is structurally analogous to the role that the rule or axiom itself plays in mediating derivationally between premises and conclusions (see again Evans, 1981; Davies, 1986; 1987; 1989; Peacocke 1986; 1989).

In brief, the upshot is that on the first construal of the simulation alternative there is some threat of collapse of the debate, since a cognitive system for process-driven simulation might yet embody tacit knowledge of a psychological theory.

When I say that there is some threat of collapse of the distinction between the theory theory and the simulation alternative on this first construal, I do not mean to suggest that there is absolutely nothing that could be said to reinforce the distinction. For example, it might be that the theory theory could distinguish itself by insisting that the psychological information that is drawn upon in the attribution, prediction and explanation process should have a particular nomological status. This is, in fact, what Goldman himself suggests (1992, p. 110):

The crucial characteristic of the theory-theory is its postulation of the possession and employment of *nomological* information, information about causal generalizations.

Another possibility is that the theory theory might take on a commitment, not about the content of the psychological principles that are deployed, but about their representational format. The theory theory might distinguish itself by going beyond the account of tacit knowledge that I sketched, and insisting that the psychological information should be encoded in some specially favoured — perhaps sentential — format.

These tactics could serve to reduce the threat of collapse, though they would not eliminate it altogether. But, more importantly, they would disturb the shape of the debate. Many of the arguments for either the theory theory or the simulation alternative proceed by arguing against the other contender. We could already have reservations about this style of argument, since there appear to be intermediate positions worthy of exploration. But, if the theory theory takes on the highly specific commitments that we have just countenanced then, as Stephen Stich and Shaun Nichols point out (1992, p. 47; emphasis added), 'the falsity of the theory-theory (narrowly construed) is *no comfort at all* to the . . . simulation theorist'.

Stich and Nichols themselves take the heroic step (as friends of the theory theory) of proclaiming the kind of collapse that we have just been contemplating to be a victory for the simulation alternative. Their idea is that, if attribution of mental states to others proceeds by deployment of a psychological theory which is also deployed in one's own decision taking processes (and in self attribution of mental states), then the simulation alternative is correct. The theory theory is thus committed to other attribution of mental states drawing upon a theory that is not also implicated in one's own decision taking; or, as they put it (1992, p. 47, n. 7):

So, as we construe the controversy, it pits those who advocate any version of the off-line simulation account against those who think that prediction, explanation and interpretation are subserved by a tacit theory *stored somewhere other than in the Practical Reasoning System*.

On this construal of the controversy, it might turn out that the theory theorist has to defend the claim that we make use of two 'copies' of the very same theory: one copy for use in our own decision taking and the other for use in interpreting others.

As an alternative to all of these moves, I want to suggest a more

general response to the threat of collapse — a response based upon appeal to an imaginative process that is rather different from the entertainment in imagination of hypotheses concerning mental states. The imaginative process that figures in the first construal of the simulation alternative is that of entertaining hypotheses of the form:

I believe that *p*
I desire that *q*.

Since these hypotheses exhibit the general form:

x believes that *p*
x desires that *q*

the states of entertaining them are appropriate inputs to a mechanism that embodies tacit knowledge of a psychological theory. The alternative that I have in mind is a process in which the simulator imaginatively adopts those mental states themselves. The simulator imagines believing that *p* and desiring that *q*. This process seems to deserve the title 'imaginative identification'.

Given this imaginative process, we have a second construal of the simulation alternative. I recommend that we set up the mental simulation debate by allowing the theory theory to make use of the notion of tacit theory, and to construe 'theory' itself in an inclusive way, to include any body of information. A psychological theory is then a body of information about psychological states.

If a processing system is to embody tacit knowledge of such a theory — a psychological theory — then it must mediate 'inference like' transitions amongst representational states whose contents themselves concern mental states. Such contents might be of the forms:

I believe that *p*
I believe that *q*.

If, though, the proponent of the simulation alternative makes use of the idea of imaginative identification then the states of the simulator do not have contents such as these. Rather, they are 'pretend belief' and 'pretend desire' states, whose contents are simply that *p* or that *q*. Thus, the threat of collapse is avoided, since processing mechanisms that mediate transitions amongst states with such contents are not going to be embodiments of tacit knowledge of the principles of a psychological theory.

I said that I would use the exposition of Goldman's development

of the simulation theory to motivate a particular way of delineating the theory theory and the simulation alternative. This is what I have just been doing — without suggesting that this is how Goldman himself would see the debate as best being conducted. Now, I return to the main thread of Goldman's own argument, in order to reach the second idea that I want to motivate.

3.2 Mental simulation and what is philosophically fundamental

Goldman's overall project was, you recall, an investigation of the interpretation strategy. At the stage of his argument that we have now reached, two putative accounts of our actual practice of attributing mental states — the 'radical interpretation' account and the 'folk theory' account — have been rejected (though we did not accept Goldman's reasons for ruling out the 'folk theory' account, and did not even consider his reasons for finding the 'radical interpretation' account wanting). Thus, by Goldman's lights, the simulation account of our interpretive practice is the last remaining candidate. But now, the simulation alternative itself faces an objection (Goldman, 1989, p. 176):

[T]he simulation approach ostensibly makes this emphatic attitude the standard mode of interpretation. Is that not difficult to accept?

To this objection, Goldman offers two lines of reply. First, simulation need not be 'introspectively vivid'. If the theory theorist is allowed to appeal to tacit theory, then surely the friend of mental simulation may appeal to introspectively bland simulation. Second, simulation need not be our only method of making mental attributions. '[I]n many cases', Goldman says (1989, p. 176), 'the interpreter relies solely (at the time of interpretation) on inductively acquired information.'

The importance of this second line of reply is that it points up the fact that, where there are different grounds available for making a judgement, one set of grounds may offer a philosophically fundamental account of our practice of making judgements of that kind, while another set of grounds does not. Thus, for example, it might be that inductively based attributions are conceived as ultimately answerable to judgements based upon mental simulation. Equally, it might be that attributions made by way of mental simulation — however natural and heuristically important they may be — are conceived as ultimately answerable to some other set of grounds for judgement.

So, the question to be asked is whether mental simulation can enter a philosophically fundamental account of propositional attitude attribution. Can simulation figure in a philosophical account of our most basic mastery of mental concepts? Goldman's answer to this question is that it cannot; and consequently he reaches his final assessment of the interpretation strategy, namely, that it cannot be carried through. The conditions of mentality cannot be wrung from a description of our practice of attributing mental states.

The 'radical interpretation' account of our interpretive practice can be reconfigured into a constitutive account of mental states. We can say that what it is for a subject to believe that *p* is for the attribution of that belief to figure in an overall interpretation that makes the best sense possible of that subject's total life and conduct (Wiggins, 1980; p. 199). Similarly, the 'folk theory' account of mental state attribution can be reconfigured into a functionalist account of what mental states are. But these two accounts have been found wanting (says Goldman) as descriptions of our actual practice. The simulation alternative fares better by the standards of realistic description, but it cannot be transposed into an account of the truth conditions of mental attributions.

The reason given for this negative assessment of the prospects for a constitutive account of mental states based upon the idea of mental simulation is that simulation is too fallible (1989, p. 182):

Since simulation is such a fallible procedure, there is little hope of treating 'M is ascribed (or ascribable) to S on the basis of simulation' as constitutive of 'S is in M'.

A similar reason could be offered for casting doubt upon the prospects for an account of the conditions for mastery of mental concepts based upon the idea of simulation. For suppose we said that what is fundamentally required for a thinker to possess the concept of belief is — so far as third-person attributions are concerned — that the thinker should judge another to believe that *p* just in case the thinker arrives at the 'pretend belief' that *p* as a result of simulating the other (putting him- or herself in the other's shoes). Then the fallibility of the simulation process would leave us with the unintuitive result that a thinker's judgements made in perfect accordance with the fundamental conditions for possession of the concept of belief could yet turn out to be false.

Here is the promised motivation for the idea that mental simulation cannot be employed to give a fundamental account of our use of

mental concepts. But although the idea has now been motivated, it is not clear that the motivation is absolutely watertight.

Let us consider, first, whether the fallibility of mental simulation provides a conclusive reason for not using simulation in a constitutive account of the nature of mental states. In a constitutive account wrung from the 'radical interpretation' account of our actual practice, we do not say that what it is for a subject to believe that p is for that attribution to be made by an eminently fallible interpreter trying to make the best sense possible of the subject's total life and conduct. Likewise, when we reconfigure the 'folk theory' account as a functionalist proposal about the nature of mental states, we do not say that a belief state is a state that plays exactly the role specified by our presumably flawed current folk theory. So, it is reasonable to wonder whether we might not be able to employ an idealized notion of simulation in a constitutive account of the nature of mental states. Such an account might be open to other kinds of objections — circularity problems certainly threaten — and it might be excess to requirements in the context of some larger project. But it would not be ruled out simply by the fact that our actual ability to simulate others is a limited and flawed ability.

Whether this idea of idealized simulation can, in the end, be legitimately appealed to is a question that I shall not pursue here (though it arises again, briefly, at the end of Section 4.2). Any further consideration might begin from the case of 'radical interpretation'. A constitutive account of the nature of mental states that is based upon the idea of 'radical interpretation' does not allow that an attribution might be wrong even though it really does contribute to making the best sense possible of the subject's total life and conduct. But that constitutive account allows that in an actual empirical attribution based upon 'radical interpretation' fallibility inevitably enters, since what is available to the interpreter are mere snippets of the subject's total life and conduct. Thus, on the 'radical interpretation' approach, the constitutive account explains the fallibility of the empirical attribution process. The question that faces the simulation approach is whether it can provide a constitutive account that similarly explains the fallibility of the empirical process of attribution by simulation. More specifically, the question is whether the notion of idealized simulation can figure in such an account.

Leaving those questions aside for the time being, we might consider, second, whether an idealized notion of simulation could figure in an

account of mastery of mental concepts — whether such a notion could play a part in the specification of possession conditions for the concept of belief, for example. In an account of the conditions for mastery of mental concepts, circularity problems are likely to be less acute than they would be in an account of the nature of mental states, since there is no objection on grounds of circularity to the use of a concept in an account of its own possession conditions (Peacocke, 1992, p. 9). Thus, for example, there would be no objection on the grounds of circularity to an account of mastery of the concept *square* that said that in order to master the concept *square* a thinker must judge to fall under that concept those objects that are in fact *square*. (Of course, there would be other objections to that putative account.)

3.3 Review

In this section, I have used an exposition of Goldman's development of the simulation theory in order to do two things.

First (3.1), I have used the threat of collapse of the distinction between the theory theory and the simulation alternative to motivate a particular way of delineating the two opposed views in this debate. On the one hand, I recommend a fairly inclusive version of the theory theory. On the other hand, I suggest that we regard the simulation process as the adoption in imagination of (pretend) beliefs and desires rather than the imaginative entertainment of hypotheses about beliefs and desires.

Second (3.2), I have used the undoubted fallibility of the simulation process to motivate the idea that mental simulation cannot figure in a philosophically fundamental account of the nature of mental states, or the conditions for mastery of mental concepts. But, having motivated that idea, I have — in the last few paragraphs of Section 3.2 — entered the suggestion that the possibility of a simulation theoretic account of the mastery of mental concepts has not quite been ruled out.

4. Gordon and our discourse about beliefs

There are at least two important ways in which Gordon's development of the simulation view differs from Goldman's. One difference is that Goldman sees himself as giving a 'processing' version of the simulation view, in terms of the 'off-line' operation of a system of decision pro-

cesses. Gordon, in contrast, takes the un-scientized notion of imaginative identification as the fundamental primitive notion.

A second difference arises from the fact that Goldman — as we have just seen — is not optimistic about the prospects of the simulation view yielding an account of our mastery of mental concepts. There are really two components to this pessimism about mental simulation's contribution to constitutive accounts. One component — as we noted at the end of Section 1 — is that the simulation view addresses the case of third-person attribution of mental states by taking the first-person case for granted. It is this component that moves Goldman to say (1989, p. 183):

If the simulation theory is right . . . it looks as if the main elements of the grasp of mental concepts must be located in the first-person sphere.

The simulation view needs to be augmented, and Goldman (1993) seeks the needed further elucidation of the nature of mental states and of our grasp of the concepts of those states in the idea that mental states like belief have intrinsic, introspectible qualities — qualia. The other component is the fallibility of mental simulation, which presents a problem for the use of simulation even to cantilever out from the first-person case to the third-person case.

4.1 First person and third person

Gordon differs over both the first-person and the third-person case. In the first-person case, Gordon agrees that the simulation view needs to be augmented; but he disagrees about the nature of the augmentation that is required. In particular, Gordon rejects a certain kind of introspectionism.

Suppose that you are asked whether you believe that the planet Neptune has rings. According to the introspectionism that Goldman favours, the canonical way to answer this question is to look inwards and discern the intrinsic and introspectible qualities of inner states. Gordon (in press), in contrast, denies that there is any introspective identification of a state as a belief that the planet Neptune has rings. Rather, you simply ask yourself whether Neptune has rings. If you answer, 'Yes, Neptune does have rings', then you report yourself as believing that Neptune has rings.

In this, Gordon explicitly follows Evans, who said (1982, pp. 225–6):

We can encapsulate this procedure for answering questions about what one believes in the following simple rule: whenever you are in a position to assert that *p*, you are *ipso facto* in a position to assert 'I believe that *p*'.

But, Evans went on to say (p. 226):

But it seems pretty clear that mastery of this procedure cannot constitute a full understanding of the content of the judgement 'I believe that *p*'. Understanding of the content of the judgement must involve possession of the psychological concept expressed by ' ξ believes that *p*', which the subject must conceive as capable of being instantiated otherwise than by himself.

And, in a similar spirit, Gordon (in press) distinguishes between *comprehending* and *uncomprehending* ascriptions. Someone who mechanically prefaces her assertions with 'I believe that' is making only uncomprehending ascriptions. So, the account of the first-person case must be coupled with an account of third-person attributions of belief, in such a way as to make clear that it is the very same concept that is applied in the two cases.

In the third-person case, Gordon maintains that the simulation view offers a radical alternative account of our mental talk — 'a way of interpreting ordinary discourse about beliefs' (1986, p. 166). Thus (Gordon, in press):

To ascribe to O a belief that *p* is to assert that *p* within the context of a simulation of O.

This is certainly a use of simulation to cantilever out to the third-person case; but it is explicitly cast as an account of our 'discourse about beliefs' — it tells us how to play the language game, but it does not immediately yield any statement of the truth conditions of 'O believes that *p*'. And because he is not directly concerned with truth conditions, Gordon is untroubled by the fallibility of simulation.

As Jane Heal (in press) notes, there is a potential problem for Gordon's account at just this point. For, to the extent that he does not award truth conditions to ascriptions of beliefs, but instead casts his account in terms of a certain kind of speech act — assertion within the context of simulation — he faces an obligation to explain how ascriptions of belief can occur in embedded contexts, for example, in the antecedents of conditionals ('If O believes that *p*, then . . .'). Gordon (in press) makes some initial moves in response to this objection, by considering conditionals of the form:

If O believes that *p*, then she will do *X*.

Thus (Gordon, in press):

Where I am supposing another person, Mary, to believe that *p*, I embed the supposition that *p* within a simulation of Mary and, using my *adjusted* motivational and other resources, 'decide' whether or not to do *X*.

But, as he acknowledges, what he says about that kind of conditional does not generalize; the problem of embedded belief ascriptions is left open.

However, I do not want to press this objection, since it does not seem obvious that the problem of embedded contexts is insoluble. Gordon can perhaps draw some encouragement from the case of metaphor. If someone says, 'Metaphorically, *p*', then he does not straightforwardly assert that *p*. But nor (on, say Davidson's [1984] account of metaphor) is there some other proposition, *q*, which the speaker advances with assertoric force. It is not that the operator 'Metaphorically' converts the proposition that *p* into another proposition with which the speaker does the familiar assertoric thing. Rather, the only proposition involved is the proposition that *p*; but with this proposition, the speaker does a less familiar thing. The analogy with belief ascriptions is clear enough. If someone says, 'O believes that *p*', then he does not straightforwardly assert that *p*. But nor (on Gordon's account of belief ascriptions) is there some other proposition, *q*, which the speaker advances with assertoric force – since 'O believes that *p*' is not truth conditioned. The only proposition involved is the proposition that *p*; but with this proposition, the speaker does a less familiar thing: he asserts it within the context of a simulation of O. The reason why Gordon might draw some comfort from this analogy is, of course, that there seems to be no doubt that 'Metaphorically, *p*' can occur embedded, for example, in the antecedent of a conditional. Whatever account is given if this occurrence of 'Metaphorically, *p*' can, perhaps, be transposed to the case of 'O believes that *p*'.

Thus, Gordon's development of the simulation alternative as an account of our discourse about beliefs may survive the objection from embedded contexts. But, as it stands, it does not deliver an answer to the question with which we began.

4.2 The concept of belief

Strawson's constraint upon first-person and third-person attributions of mental states begins from the idea that 'is in pain' and 'believes that

penguins waddle' are predicates — expressions that go together with names of individuals (here, of persons) to form sentences that are evaluable as true or false. It is the idea that these expressions are predicates that motivates the constraint. If someone claims that to say 'I am in pain' is tantamount to saying 'Ouch', and extends his claim to the third person by adding that to say 'O is in pain' is to say 'Ouch' within the context of a simulation of O, then he has achieved some conformity with a requirement of univocality. But he can hardly be said to have measured up to Strawson's constraint.

Likewise, if someone claims that to say 'I believe that penguins waddle' is tantamount to saying 'Penguins waddle' and extends this claim to the third person by adding that to say 'O believes that penguins waddle' is to say 'Penguins waddle' within the context of a simulation of O, then Strawson's constraint has been by-passed rather than met. For this account of our 'discourse about beliefs' does not treat 'believes that penguins waddle' as a genuine predicate.

The theory that I have just sketched (the 'Ouch' theory of belief attribution) is not Gordon's. In fact, Gordon is very sensitive to the need to give an adequate account of *comprehending* first-person belief ascriptions, which 'require conceptually prising one's own present beliefs apart from the facts, so that, like the beliefs of another, they may be false or at variance with the facts' (in press). But the 'Ouch' theory serves well enough to highlight the fact that what Gordon gives us is not cast in the form of an account of 'possession of the psychological concept expressed by "ξ believes that *p*"'.

We can see easily enough what a simulation based account of possession of that psychological concept would be like. Roughly, the first-person clause would say that a thinker judges 'I believe that *p*' when she has the (conscious) belief that *p* (and so is ready to assert that *p*). The third-person clause would then cantilever out using the notion of simulation. Such a clause can be modelled upon a third-person clause offered by Peacocke (1992, p. 163) which does not make use of the notion of simulation. In outline, what Peacocke offers is:

If a thinker judges 'O believes that *p*', then the thinker incurs a commitment; namely, a commitment to O being in an internal state with a certain 'functional role'. The thinker does not, however, have to know what the functional role associated with believing that *p* is. The commitment is simply that O should be in an internal state with the same 'functional role' as the internal

state that the thinker would herself be in, were she herself to believe that *p*.

This clause presumes upon the first-person clause in that the thinker's commitment embeds the thought that the thinker herself believes that *p*. And the cantilever depends upon the idea of a relation between the first-person and the third-person; namely, the relation of having internal states with the same 'functional role'. Following the contours of this clause, we can offer an alternative, with mental simulation as the first-person/third-person relation upon which the cantilever depends:

If a thinker judges 'O believes that *p*', then the thinker incurs a commitment; namely, a commitment to believe (judge, assert) that *p* (within the scope of a simulation) if she were to simulate O.

The fallibility of ordinary everyday attempts at simulation poses a problem for this account of the possession conditions for the concept 'ξ believes that *p*' — particularly, for the appropriate connections between the incurred commitments being met and the original judgement being true. A thinker's judgement 'O believes that *p*' may be true, even though within the scope of her best — but flawed — effort at simulating O, the thinker would not, in fact, believe, judge, or assert that *p*. (Cf. Peacocke, 1992, pp. 169–70.)

The next move to make would be to introduce an idealized notion of simulation, as suggested at the end of Section 3.2. That move would — more or less by stipulation — avoid the problem of fallibility. But, it is a further question whether it could contribute to a philosophically fundamental, and distinctive, account of mastery of the concept of belief, while also providing an explanation of the way that fallibility enters the empirical process of attribution. Answering that question might involve us in following out analogies, for example, between ideal simulation and ideal verification, and between simulation and spatial relocation. It might also cause us to reflect more upon the fallible and the fundamental. Clearly, these are not matters to embark upon now. But until the question is answered, we still do not know whether it is possible to meet Strawson's constraint while adopting Quine's idea.

As I remarked at the outset, the constraint and the idea date from 1959 and 1960, respectively. Here — as in so many other areas of philosophy — progress is slow. But what makes the progress slow is

also what makes the mental simulation debate so fascinating. For it brings us up against much that is genuinely difficult in our conception of our own mentality.

Note. Early versions of this material were presented in talks at Harvard University, McGill University, Washington University St. Louis, and the University of Alberta, during the autumn of 1992. I learned much from those occasions, and especially from discussions with Robert Gordon. Comments by Christopher Peacocke on the penultimate draft were very helpful in the preparation of this final version. Special thanks to Tony Stone for his advice and assistance.

Bibliography

- Allison, H. E. 1983: *Kant's Transcendental Idealism*. New Haven: Yale University Press.
- Baron-Cohen, S., Leslie, A. M. and Frith, U. 1985: Does the autistic child have a 'theory of mind'? *Cognition*, 21: 37-46.
- Bennett, J. 1966: *Kant's Analytic*. Cambridge: Cambridge University Press.
- Bisiach, E., Berti, A. and Vallar, G. 1985: Analogical and logical disorders underlying unilateral neglect of space. In M. Posner and O. Marin (eds), *Attention and Performance*, vol. 11. Hillsdale, New Jersey: Erlbaum.
- Bisiach, E., Geminiani, G., Berti, A. and Rusconi, M. L. 1990: Perceptual and premotor factors of unilateral neglect. *Neurology*, 40: 1278-1281.
- Bisiach, E. and Luzzatti, C. 1978: Unilateral neglect of representational space. *Cortex* 14: 129-133.
- Bisiach, E. and Vallar, G. 1988: Hemineglect in humans. In P. Boller and J. Grafman (eds), *Handbook of Neuropsychology*, vol. 1. Amsterdam: Elsevier.
- Boden, M. 1990: *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press.
- Brewer, B. 1992: Unilateral neglect and the objectivity of spatial representation. *Mind and Language*, 7: 222-239.
- Cassam, Q. 1987: Transcendental arguments, transcendental synthesis, and transcendental idealism. *Philosophical Quarterly*, 37: 355-378.
- Cassam, Q. 1989: Kant and reductionism. *Review of Metaphysics*, 43: 72-106.
- Cassam, Q. forthcoming: Transcendental self-consciousness. In P. K. Sen and R. Verma (eds), *The Philosophy of P. F. Strawson*.
- Chisholm, R. 1981: *The First Person*. Brighton: Harvester Press.
- Collingwood, R. G. 1946: *The Idea of History*. Oxford: Oxford University Press.
- Coltheart, M. 1980: Deep dyslexia: a right-hemisphere hypothesis. In M. Coltheart, K. Patterson and J. C. Marshall (eds), *Deep Dyslexia*, pp. 326-380. London: Routledge and Kegan Paul.
- Davidson, D. 1984: What metaphors mean. In *Inquiries into Truth and Interpretation*, 245-264. Oxford: Oxford University Press.
- Davies, M. 1986: Tacit knowledge, and the structure of thought and language. In C. Travis (ed.), *Meaning and Interpretation*, 127-158. Oxford: Blackwell.
- Davies, M. 1987: Tacit knowledge and semantic theory: Can a five per cent difference matter? *Mind*, 96: 441-462.

- Davies, M. 1989: Tacit knowledge and subdoxastic states. In A. George (ed.), *Reflections on Chomsky*, 131–152. Oxford: Blackwell.
- Dennett, D. 1984: Cognitive wheels: the frame problem of AI. In C. Hookway (ed.), *Minds, Machines and Evolution*, 129–151. Cambridge: Cambridge University Press.
- Dennett, D. C. 1991: *Consciousness Explained*. Boston: Little, Brown.
- Diamond, S. 1972: *The Double Brain*. London: Churchill Livingstone.
- Evans, G. 1973: The causal theory of names. *Proceedings of the Aristotelian Society*, supp. vol. 47: 187–208.
- Evans, G. 1981: Semantic theory and tacit knowledge. In S. Holtzman and C. Leich (eds), *Wittgenstein: To Follow a Rule*, 118–137. London: Routledge and Kegan Paul. (Reprinted 1985 in *Collected Papers*, 322–342. Oxford: Oxford University Press.)
- Evans, G. 1982: *The Varieties of Reference*, ed. J. McDowell. Oxford: Oxford University Press.
- Fogelin, R. 1985: *Hume's Skepticism in the Treatise of Human Nature*. London: Routledge & Kegan Paul.
- Gallistel, C. R. 1980: *The Organization of Action: A New Synthesis*. Hillsdale, New Jersey: Erlbaum.
- Gallistel, C. R. 1990: *The Organization of Learning*. Cambridge, Mass.: MIT Press.
- Gazzaniga, M. 1988: In A. J. Marcel and E. Bisiach (eds), *Consciousness in Contemporary Science*, 226ff. Oxford: Oxford University Press.
- Gibson, J. J. 1979: *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Goldman, A. I. 1989: Interpretation psychologized. *Mind and Language*, 4: 161–185.
- Goldman, A. I. 1992: In defense of the simulation theory. *Mind and Language*, 7: 104–119.
- Goldman, A. I. 1993: The psychology of folk psychology. *Behavioral and Brain Sciences*, 16: 15–28.
- Gopnik, A. and Wellman, H. 1992: Why the child's theory of mind really is a theory. *Mind and Language*, 7: 145–171.
- Gordon, R. M. 1986: Folk psychology as simulation. *Mind and Language*, 1: 158–171.
- Gordon, R. M. 1992a: The simulation theory: objections and misconceptions. *Mind and Language*, 7: 11–34.
- Gordon, R. M. 1992b: Reply to Stich and Nichols. *Mind and Language* 7: 87–97.
- Gordon, R. M. 1992c: Reply to Perner and Howes. *Mind and Language*, 7: 98–103.
- Gordon, R. M. in press: Simulation without introspection or inference from me to you. In M. Davies and T. Stone (eds), *Mental Simulation: Philosophical and Psychological Essays*. Oxford: Blackwell.
- Harris, P. L. 1989: *Children and Emotion: The Development of Psychological Understanding*. Oxford: Blackwell.
- Harris, P. L. 1991a: The work of the imagination. In A. Whiten (ed.), *Natural Theories of Mind: The Evolution, Development and Simulation of Everyday Mindreading*, 283–304. Oxford: Blackwell.
- Harris, P. L. 1991b: Letter to Josef Perner, 30 May 1991.

- Harris, P. L. 1992: From simulation to folk psychology: the case for development. *Mind and Language*, 7: 120–144.
- Heal, J. 1986: Replication and functionalism. In J. Butterfield (ed.), *Language, Mind and Logic*, 135–150. Cambridge: Cambridge University Press.
- Heal, J. in press: How to think about thinking. In M. Davies and T. Stone (eds), *Mental Simulation: Philosophical and Psychological Essays*. Oxford: Blackwell.
- Hurley, S. L. in preparation: *The Reappearing Self*.
- Jeeves, M. A. 1965: Agenesis of the corpus callosum — physio-pathological and clinical aspects. *Proceedings of the Australian Association of Neurologists*, 3: 41–48.
- Johnson-Laird, P. N. 1983: *Mental Models*. Cambridge: Cambridge University Press.
- Kant, I. 1933: *The Critique of Pure Reason*. Tr. Kemp Smith, N. London: Macmillan.
- Lockwood, M. 1989: *Mind, Brain and the Quantum: The Compound 'I'*. Oxford: Blackwell.
- Marcel, A. J. 1993: Slippage in the unity of consciousness. In Ciba Foundation Symposium No. 174, *Experimental and Theoretical Studies of Consciousness*. Chichester: John Wiley.
- Marks, C. E. 1981: *Commissurotomy, Consciousness and the Unity of Mind*. Cambridge, Mass.: MIT Press.
- Milner, A. D. and Jeeves, M. A. 1979: A review of behavioural studies of agenesis of the corpus callosum. In I. S. Russell, M. W. Van Hof and G. Berlucchi (eds), *Structure and Function of Cerebral Commissures* 428–483. London: Macmillan.
- Nagel, T. 1979: Brain bisection and the unity of consciousness. reprinted in T. Nagel, *Mortal Questions*. Cambridge: Cambridge University Press. (First published in 1971 in *Synthese*, 20.)
- O'Keefe, J. 1985: Is consciousness the gateway to the hippocampal cognitive map? A speculative essay on the neural basis of mind. In D. A. Oakley (ed.), *Brain and Mind*, 59–98. London: Methuen.
- O'Keefe, J. 1990: A computational theory of the hippocampal cognitive map. In J. Storm-Mathisen, J. Zimmer and O. P. Ottersen (eds), *Progress in Brain Research*, 83: 301–312. Amsterdam: Elsevier.
- O'Keefe, J. 1991: The hippocampal cognitive map and navigational strategies. In J. Paillard (ed.), *Brain and Space*, 273–295. Oxford: Oxford University Press.
- O'Keefe, J. 1993: Kant and the sea-horse. In N. Eilan, B. Brewer and R. McCarthy (eds), *Spatial Representation: Problems in Philosophy and Psychology*, 43–64. Oxford: Blackwell.
- O'Keefe, J. and Nadel, L. 1978: *The Hippocampus as a Cognitive Map*. Oxford: Oxford University Press.
- Parfit, D. 1984: *Reasons and Persons*. Oxford: Oxford University Press.
- Peacocke, C. 1986: Explanation in computational psychology: language, perception and level 1.5. *Mind and Language*, 1: 101–123.
- Peacocke, C. 1989: When is a grammar psychologically real? In A. George (ed.), *Reflections on Chomsky*, 111–130. Oxford: Blackwell.
- Peacocke, C. 1992: *A Study of Concepts*. Cambridge, Mass.: MIT Press.
- Peacocke, C. 1993: Externalist explanation. *Proceedings of the Aristotelian Society*, 93: 203–230.

- Perner, J. 1991: *Understanding the Representational Mind*. Cambridge, Mass.: MIT Press.
- Perner, J. and Howes, D. 1992: 'He thinks he knows': and more developmental evidence against the simulation (role taking) theory. *Mind and Language*, 7: 72–86.
- Piaget, J. and Inhelder, B. 1951/1975: *The Origin of the Idea of Chance in Children*. New York: Norton.
- Powell, C. T. 1990: *Kant's Theory of Self-Consciousness*. Oxford: Oxford University Press.
- Quine, W. V. O. 1960: *Word and Object*. Cambridge, Mass.: MIT Press.
- Rorty, R. 1970: Strawson's objectivity argument. *The Review of Metaphysics*, 24: 207–244.
- Schwyzler, H. 1990: *The Unity of Understanding*. Oxford: Oxford University Press.
- Sergent, J. 1990: Furtive incursions into bicameral minds. *Brain*, 113: 537–568.
- Seymour, S., Reuter-Lorenz, P. and Gazzaniga, M. 1994: The disconnection syndrome: basic findings reaffirmed. Abstracted in *The Society of Neuroscience*, 1993.
- Shebilske, W. L. 1984: Context effects and efferent factors in perception and cognition. In W. Prinz and A. F. Sanders (eds), *Cognition and Motor Processes*. Berlin: Springer-Verlag.
- Shoemaker, S. 1984: Causality and properties. In S. Shoemaker, *Identity, Cause and Mind*. Cambridge: Cambridge University Press.
- Sperry, R. W. 1990: Forebrain commissurotomy and conscious awareness. In C. Trevarthen (ed.), *Brain Circuits and Functions of the Mind*. Cambridge: Cambridge University Press.
- Stich, S. and Nichols, S. 1992: Folk psychology: simulation or tacit theory? *Mind and Language*, 7: 35–71.
- Stich, S. and Nichols, S. in press: Second thoughts on simulation. In M. Davies and T. Stone (eds), *Mental Simulation: Philosophical and Psychological Essays*. Oxford: Blackwell.
- Strawson, P. F. 1959: *Individuals*. London: Methuen.
- Strawson, P. F. 1966: *The Bounds of Sense*. London: Methuen.
- Tegnèr, R. and Levander, M. 1991: Through a looking glass. *Brain*, 114: 1943–1951.
- Trevarthen, C. 1974: Analysis of cerebral activities that generate and regulate consciousness in commissurotomy patients. In S. Dimond and J. G. Beaumont (eds), *Hemisphere Function in the Human Brain*. London: Elek Science.
- Trevarthen, C. 1984: Biodynamic structures. In W. Prinz and A. F. Sanders (eds), *Cognition and Motor Processes*. Berlin: Springer-Verlag.
- Walker, R. 1978: *Kant*. London: Routledge.
- Wiggins, D. 1980: What would be a substantial theory of truth? In Z. van Straaten (ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson*, 189–221. Oxford: Oxford University Press.
- Wilkie, D. M. and Palfrey, R. 1987: A computer simulation model of rats' place navigation in the Morris water maze. *Behavioural Research Methods, Instruments and Computers*, 19: 400–403.
- Williams, B. 1978: *Descartes: The Project of Pure Inquiry*. Harmondsworth: Penguin.
- Wilson, M. D. 1987: *Descartes*. London: Routledge & Kegan Paul.

- Wimmer, H. and Perner, J. 1983: Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13: 103-128.
- Wimmer, H., Hogrefe, G.-J. and Perner, J. 1988: Children's understanding of informational access as a source of knowledge. *Child Development*, 59: 386-396.