

Mental Simulation, Tacit Theory, and the Threat of Collapse*

MARTIN DAVIES AND TONY STONE

According to the theory theory of folk psychology, our engagement in the folk psychological practices of prediction, interpretation and explanation draws on a rich body of knowledge about psychological matters. According to the simulation theory, in apparent contrast, a fundamental role is played by our ability to identify with another person in imagination and to replicate or re-enact aspects of the other person's mental life. But amongst theory theorists, and amongst simulation theorists, there are significant differences of approach.

On the theory-theory side of the debate, some use the term 'theory' in a narrow sense to require that the body of knowledge about psychological matters should be organised around generalisations that have the status of psychological laws. Others use the term in a more inclusive sense to encompass any body of information or misinformation about psychology. Some theory theorists say that the theory is learned, either by processes analogous to scientific investigation or else by cultural transmission. Others say that the theory is, to a considerable extent, innate. On the simulation-theory side, some think of mental simulation as restricted to the re-enactment of rational transitions in thought. Others contrast the simulation approach with accounts of mental-state attribution that make use of an assumption of rationality. Some abstract away from questions about cognitive machinery and describe simulation in personal-level folk psychological terms. Others regard the simulation theory as a contribution to cognitive science.¹

1. Introduction: Cognitive science, tacit knowledge and the threat of collapse

Alvin Goldman, who is one of the three philosophers primarily responsible for developing and defending the simulation theory over the last fifteen years,² is a cognitive-scientific simulation theorist. His first paper on the topic, 'Interpretation psychologized' (1989), argues for the simulation theory in the context of the question, '[H]ow does the (naïve) interpreter arrive at his/her judgments about the mental attitudes of others?' (1989/1995, p. 74), and the accompanying remark that, 'Philosophers who have addressed this question

* A very early version of some of this material was used in a talk by MD at a workshop on simulation theory and tacit knowledge held in Adelaide in December 1997. (This was the first of a series of four workshops on simulation theory supported by a grant from the Australian Research Council.) Some of the material also figured in talks by MD at Cornell University, the University of Maryland, the University of Canterbury, Victoria University of Wellington, the University of Auckland and the University of Otago, and a penultimate version was used as a basis for discussion at a seminar at New York University. MD learned much from the audiences on all these occasions. We are especially grateful to Gregory Currie, Hartry Field, Alvin Goldman, Robert Gordon, Mark Greenberg, Jane Heal, Frank Jackson, Graham Macdonald, Christopher Peacocke, Ian Ravenscroft, Stephen Schiffer and Stephen Stich, for comments and conversations.

¹ Goldman favours the 'nomological construal of theories' (1989/1995, p. 89); Stich and Nichols, 1992, use the term 'theory' in the more inclusive way. On learning and innateness, see e.g. Gopnik and Wellman, 1992; Fodor, 1992. The differences amongst simulation theorists are discussed in Section 2 below.

² Robert Gordon and Jane Heal are the other two.

have not, in my view, been sufficiently psychological, or cognitivist, even those who are otherwise psychologically inclined' (ibid.). A later paper is even more explicit (1995, p. 186):

My framing of the problem deliberately construes it as an empirical question about the psychology of attributors: 'What actually goes on in the heads of attributors that accounts for their attributions?' . . . We should seek to identify the cognitive processes involved in this task.

On the opposite side of the debate from Goldman, Stephen Stich and Shaun Nichols also adopt a cognitive-scientific approach. Indeed, they defend the theory theory as an instance of 'the dominant explanatory strategy in cognitive science' (1992/1995, p. 124), namely, the strategy of 'positing an internally represented "knowledge structure" . . . which serves to guide the execution of the capacity to be explained' (ibid., p. 121). According to their version of the theory theory, our ability to engage in the folk psychological practices of prediction, interpretation and explanation rests on a 'largely tacit psychological theory' (ibid.). In contrast, according to Goldman's version of the simulation theory, our folk psychological practices rest on mental simulation; and this is driven, not by a body of theory, but by processes that are the same as those operating in the system that is being simulated (1989/1995, p. 85). If the same processes operate on initial states that are the same as those in the system being simulated, then the simulation exercise should end up producing a final state that is the same as the final state of the system being simulated. Even if the initial state of the simulation exercise is not exactly the same as the initial state of the system being simulated, provided that the two are relevantly similar we may expect that the final states will also be similar. So process-driven mental simulation seems to offer a methodology for predicting the mental states of other people; and the same methodology can also figure in the practices of attributing mental states on the basis of observed behaviour and of explaining behaviour in terms of mental states.

Goldman accepts that folk psychological practices may depend on the presence of a special-purpose component or module. But where the theory theory says that this module stores tacit knowledge about psychology, he urges that a module could embody 'the off-line simulation heuristic' (1992/1995, p. 194). He allows that the theory theory is an instance of one well-established explanatory strategy in cognitive science, but insists that 'choice of the simulation theory would not be a radical departure from familiar paradigms in cognitive science' (ibid., p. 192). Cognitive science postulates 'knowledge-poor' as well as 'knowledge-rich' procedures.

Goldman's version of the simulation theory is a cognitive-scientific theory. But not all defenders of the simulation theory regard it as a contribution to the empirical scientific study of the mind. In a challenging series of papers,³ Jane Heal argues against a way of looking at the theory versus simulation debate according to which (1994, p. 132):

it is seen as an empirical question about how our undoubted ability to predict others' future thoughts, feelings and actions on the basis of knowledge of their current psychological states is implemented at a sub-personal level.

³ See especially Heal, 1994, 1998a, 2000.

So she opposes the idea that the simulation theory is ‘an a posteriori hypothesis about the workings of sub-personal cognitive machinery’ (1998a, p. 477). One strand in Heal’s argument against this way of setting up the debate is that it presents a *threat of collapse*. She argues that if a mechanism is used to simulate the operation of other mechanisms of the same kind then it will turn out to embody tacit knowledge of a substantive theory about how mechanisms of that kind operate. So if the simulation theory, conceived in this mechanistic way, turned out to be true then the tacit-theory theory would also turn out to be true. The debate between theory and simulation would collapse, and it would not collapse in a way that was neutral as between the two contending parties. The collapse would leave only a debate between different variations on the theory-theory theme. Heal’s ‘threat of collapse’ argument is the main topic of this paper.

If Heal is right about the threat of collapse then this is potentially serious for Goldman’s version of the simulation theory. For Goldman, on the side of the simulation theory, and Stich and Nichols, on the side of the theory theory, view the debate in just the way that Heal opposes. So Heal’s argument, which stands unanswered in the literature on the theory versus simulation debate,⁴ would lead to the conclusion that Goldman’s version of the simulation theory collapses into a version of the tacit-theory theory.

It is true that there is a way for Goldman to accept the letter of Heal’s argument and yet to deny that it has such serious consequences. For the argument depends on a crucial background assumption (Heal, 1994, p. 134; emphasis added):

[I]f this is how we think mental simulation works to deliver answers to psychological questions about others, is there really any significant difference between it and the theory theory? I suggest that there is none, *at least on the definition of tacit knowledge we have assumed*.

So it is open to Goldman to avoid the threat of collapse by rejecting the account of tacit knowledge on which Heal’s argument relies.⁵ In order to judge whether this would be a good strategy for Goldman to adopt, we need to consider what kind of account of tacit knowledge would be most appropriate to the theory versus simulation debate as Goldman and Stich and Nichols conceive it. In particular, we need to ask whether tacit knowledge should be taken to require explicit representation. The notion of explicitness here connects with issues about format and storage in information-processing psychology rather than with verbalisability. We take it that explicit representation makes use of a language-like format (the language of thought, for example) and that information that is stored by being explicitly represented has to be accessed before it can be used.

On the account that Heal assumes for the purposes of her argument, explicit representation is *not* necessary for tacit knowledge. This is crucial, for the argument would not go through if we were to stipulate that tacit knowledge of a psychological theory requires stored sentence-like representations of a battery of psychological generalisations. Indeed,

⁴ In fact, some of the resources for a response are already present in Christopher Peacocke’s editorial Introduction to the volume in which Heal’s paper appears (Peacocke, 1994, pp. xx–xxvi).

⁵ The account is that of Davies, 1981a,b, 1987, 1989, 1995a; see also Evans, 1981; Peacocke, 1986, 1989. Heal cites Davies, 1987, and notes that it develops a suggestion by Evans.

Heal herself suggests that imposing the requirement of explicit representation would be a good way of responding to the threat of collapse.⁶ But in our view this would not be well motivated either by the structure of the debate or in its own right.

We shall leave it until later (Section 4.3) to argue that the stipulation would not be well motivated in its own right. What is more important for our introductory purposes here is that such a stipulation would distort the structure of the debate. Stich and Nichols (1992) intend to set up the debate in such a way that the theory theory and the simulation theory exhaust the possibilities and, for that reason, they adopt an inclusive notion of an internally represented theory. There are two aspects to this inclusiveness. The first concerns the term ‘theory’ itself. Stich and Nichols allow that any body of information or misinformation about psychological matters counts as a theory whether or not it is organised around psychological laws. The second aspect concerns the notion of internal representation. It is not supposed to be ruled out, for example, that a connectionist network might embody tacit knowledge of a theory even though networks do not make use of explicit representations in a language-like format. Goldman accepts this way of setting things up (1992/1995, p. 191): ‘I concur with Stich and Nichols in their assertion that the theory-theory and the simulation theory are the only two games in town’. Now, in fact, it might reasonably be doubted whether there is any notion of theory so inclusive that the theory theory could occupy all of the logical space not claimed by the simulation theory. But it would not even begin to be plausible that the theory theory and the simulation theory exhaust the options if the theory theory were to require explicit representation of psychological principles.

Our aim in this paper is to defend the legitimacy of Goldman’s way of conceiving of the theory versus simulation debate. So we respond to Heal’s ‘threat of collapse’ argument even while accepting the account of tacit knowledge that she assumes. In the next section, we provide a scheme for classifying different versions of the simulation theory and, in Section 3, we consider the role that elements of psychological theory play in the use of mental simulation. In Section 4, we review the account of tacit knowledge that figures in Heal’s argument. Then, in Section 5, we present the argument itself and offer our response.

2. Variations on the simulation theory

Our main concern in this paper is with a difference of approach between two simulation theorists, Goldman and Heal. In this section, we set this difference within a more general framework.

Simulation theorists (and their opponents) differ over the level of description at which the simulation theory is primarily pitched. Let us say that the *personal level* of description is the level at which persons are described as such, using everyday mental notions. At the personal level, we talk about people as experiencing, thinking subjects and agents. Their thinking, both theoretical and practical, is subject to norms and they arrive at judgements and decisions on the basis of reasons. Some philosophers would go so far as to maintain that personal-level explanations of judgements and decisions in terms of reasons are of a

⁶ Heal, 1994, p. 136: ‘Another (and I think more promising) move by which a genuine dispute could be re-introduced would be to insist that a “theory” should have a sentence-like mode of representation.’

distinctive kind, not to be assimilated to explanations that proceed by subsuming events under covering laws about how the world happens to work.⁷ The personal level of description can be distinguished from various *subpersonal levels*, including the biological level of description and, most importantly for the purposes of this paper, the level of information-processing mechanisms. The events and processes described at subpersonal levels may be regarded as underpinning, or perhaps as constituting, as perhaps even as being identical with, the events and processes described at the personal level.

Views of inter-level relationships range from autonomy for higher levels through to reductionism.⁸ Thus, some theorists regard the relationship between the personal level and the level of information-processing mechanisms as one of independence. According to this view, the correctness of personal-level descriptions – and especially of the descriptions that are of central importance for the philosophy of mind – is not answerable to empirical discoveries about how the information-processing machinery in our heads works. Others take a strongly reductionist view: investigation of personal-level phenomena must proceed by first recasting descriptions of those phenomena into the language of information-processing psychology.

Intermediate conceptions of the inter-level relationship are also possible. We ourselves would reject the idea that empirical discoveries in neuroscience or cognitive psychology are irrelevant to the personal-level descriptions that are important for the philosophy of mind. But we would allow that, for at least some personal-level phenomena such as experience, thought, or acting for a reason, accounts that are cast purely in the terms of information-processing psychology leave upward explanatory gaps. The most familiar example of this is, of course, the explanatory gap associated with phenomenal consciousness.

The distinction between personal and subpersonal levels of description provides one dimension of variation amongst simulation theorists. Some regard the simulation theory as primarily pitched at the personal level, others at a subpersonal level of description. There is also a second dimension of variation: simulation theorists differ over the epistemological status of the simulation theory. Some regard the theory as being straightforwardly empirical while others take it to have a more a priori character.

2.1 Subpersonal-level and empirical

As we have seen, Goldman and Stich and Nichols regard the simulation theory and the theory theory as competing accounts of the relationships between pieces of information-processing machinery. As they see the debate, the competing theories are primarily pitched at a subpersonal level of description.⁹ What is at issue is particularly clear if we consider the relationship between the information-processing machinery that subserves my own decision-making and the machinery that subserves prediction of another person's decisions. The machinery that subserves my decision-making takes representations corresponding to my

⁷ E.g. McDowell, 1985. Note that Heal also takes this view of everyday psychological explanations; see Heal, 1986/1995, p. 52.

⁸ Stone and Davies, 1999; Davies, 2000.

⁹ Goldman, 1989, 1992, 1995, 2000; Gallese and Goldman, 1998; Stich and Nichols, 1992, 1995, 1996, 1997; Nichols et al., 1996; Nichols and Stich, 1998.

beliefs and desires as inputs and produces a representation corresponding to a decision as output,¹⁰ and the production of this output representation typically leads on to a piece of behaviour.

The theory theory says that there are important elements in the decision-*predicting* machinery that are different from anything that is involved in my own decision-*making*.¹¹ These elements include, in particular, a stored body of information about the kinds of decisions that people typically make in various circumstances. The simulation theory, in contrast, stresses that decision-*predicting* involves, in large part, use of the same information-processing machinery that is implicated in my own decision-*making*.¹² In information-processing terms, the simulation theory says that my own decision-*making* machinery can be taken 'off line' from its usual inputs and outputs and can be fed input representations that correspond to another person's beliefs and desires. Considered within my own cognitive economy, these input representations amount to 'pretend beliefs' and 'pretend desires' rather than real beliefs and desires. Operating in this off-line mode, the system then produces output representations that amount to 'pretend decisions' rather than real decisions. These representations do not lead to real behaviour on my part. Rather, they provide the content of the decision that I expect the other person to reach.

As Goldman and Stich and Nichols see the debate, the simulation theory and the theory theory are straightforwardly empirical and the issue between them cannot be settled by a priori considerations. For someone who takes this view, considerable interest attaches to discoveries that components of the visual system are also active during visual imagery and that components of the system that is implicated in intentional action are also involved in motor imagery (Currie, 1995; Currie and Ravenscroft, 1997). For these discoveries make it more empirically plausible that imagined or pretended decision-making should make use of some of the same information-processing components as real decision-making.

It is no surprise, then, to find Goldman (2000; see also 1995) drawing on these results about visual and motor imagery in order to support the simulation theory. In these cases, voluntarily produced 'pretend' mental states or 'facsimile events' share many of the properties of their unpretended counterparts. This does not yet show that 'pretend' mental states figure in our everyday psychological understanding of other people. It does not show that mental imitation plays a role in our practices of prediction, interpretation and explanation. But Goldman assembles further plausibility considerations from work on very young infants' imitation of facial expressions and of actions:¹³

¹⁰ We use the term 'corresponding to' so as not to prejudge the relationship between propositional attitude states and states of information-processing machinery. But it is important that the representations, like my beliefs and desires, concern the world. They are not representations *of* my having beliefs and desires.

¹¹ For a clear statement, see Stich and Nichols, 1992/1995, p. 154, n. 7.

¹² The simulation theory also allows, of course, that predicting another person's decisions involves some resources that are not implicated in decision-making. These resources include, for example, machinery for keeping track of whose decision is being predicted.

¹³ Goldman, 2000, p. 000. See Meltzoff and Moore, 1983, 1995, on facial imitation and Meltzoff, 1995, on the imitation of actions. Eighteen-month-old infants who witness imperfectly executed actions enact (what we would naturally describe as) what was intended by the agent rather than what actually happened.

Meltzoff's work suggests that the capacity for intentional imitation of behavior is an innate and fundamental capacity of human nature. Currie and Ravenscroft show that people have capacities for intentional production of mental facsimiles as well, at least facsimiles of their own mental states. It is not too dramatic a leap, therefore, to conjecture that people might intentionally deploy mental facsimiles to interpret the mental states of others, as [the simulation theory] proposes.

With Vittorio Gallese, Goldman also mounts an argument in favour of the simulation theory from findings concerning mirror neurons in macaque monkeys.¹⁴ Mirror neurons respond both when the monkey is performing a particular action and when the monkey observes the same action performed by another. Suppose that internally generated activation in mirror neurons constitutes an action plan. Then we can think of externally generated activation in the same mirror neurons as also constituting a plan to perform the same action, though a plan that is not executed. It is inhibited or taken 'off line'. The externally generated activation of mirror neurons may, then, be a precursor of something that would figure in a simulation-theory account of the attribution of mental states to other people, namely the use of mental facsimiles in order to attribute goals to others.

2.2 Personal-level and a priori

As we saw in the last section, Heal is opposed to the idea that the simulation theory constitutes an answer to an empirical question about information-processing mechanisms. She argues (1998a, p. 478) that 'it is an a priori truth that simulation, in some sense, must be given a substantial role in our personal-level account of psychological understanding'. Equally, it is 'an a priori truth, and not an a posteriori one, that theory-theory (at least on one strong but natural understanding of "theory-theory") is unacceptable as an account of our personal-level abilities' (ibid.). The version of the simulation theory that Heal claims can be established a priori is the *co-cognition* theory (ibid., p. 484):¹⁵

It is an a priori truth that thinking about others' thoughts requires us, in usual and central cases, to think about the states of affairs which are the subject matter of those thoughts, i.e. to co-cognize with the person whose thoughts we seek to grasp.

The (strong) version of the theory theory to which this stands opposed says (ibid., p. 485):

[T]houghts about X form a separate subject matter which is independent of the subject matter X. . . . [G]rasp on the subject matter of vegetables is one thing and grasp on the subject matter of thoughts about vegetables is a quite different and independent thing.

These competing theories are clearly pitched at the personal level of description. So we can ask what, on Heal's account, is the relationship between the a priori, personal-level co-cognition theory and an empirical, subpersonal-level simulation theory of the kind that

¹⁴ Gallese and Goldman, 1998.

¹⁵ Heal, 1998a, p. 483: 'Co-cognition is just a fancy name for the everyday notion of thinking about the same subject matter.'

Goldman defends and Stich and Nichols oppose. The answer she gives (*ibid.*, pp. 492–5) is that the a priori theory does not entail the empirical one but might, in the context of other empirical claims, support it.¹⁶ But Heal does not say explicitly how her co-cognition theory would fare if empirical investigation were to reveal information-processing machinery with a structure as indicated by Stich and Nichols's version of the theory theory. In particular, what are we to make of the apparent possibility that there should turn out to be a system encoding a substantial body of psychological theory that is drawn on for decision-predicting but not for decision-making? Heal seems to say that by the time we come to the empirical investigation of information-processing machinery, 'The strong theory-theory is nowhere in the running as an option' (*ibid.*, p. 495). But this leaves it unclear whether the apparent empirical possibility that we have just described is supposed to be a genuine possibility but compatible with the co-cognition theory or not a genuine possibility because ruled out by the argument against the strong theory theory.¹⁷

It does not seem very plausible that Heal should intend her co-cognition theory to be compatible with Stich and Nichols's version of the theory theory. But nor is it easy to see how an empirical, subpersonal-level theory theory is ruled out by the a priori considerations in favour of co-cognition. Part of the difficulty here is that there is an extremely thin construal of the notion of thinking of the same subject matter on which the co-cognition theory seems undeniable, yet nothing interesting about the actual prediction of decisions seems to follow from the co-cognition theory so construed.

Anyone who is utterly unable to frame or entertain the thought that p is likewise unable to entertain the thought that O believes that p . Consider, for example, Yvonne who believes that the square on the hypotenuse is 25 units and Vincent who has no grasp at all of the concept of an hypotenuse. Because of his conceptual lack, Vincent cannot entertain the thought that the square on the hypotenuse is 25 units. Similarly, he is unable to entertain the thought that Yvonne believes that the square on the hypotenuse is 25 units, for the content of that second thought embeds the content of the first. Vincent is unable to think about states of affairs involving hypotenuses and as a result is unable to think about Yvonne thinking about states of affairs involving hypotenuses. If Vincent were able to think about Yvonne believing that the square on the hypotenuse is 25 units then he would, in that very thinking,

¹⁶ Heal also says that the subpersonal-level theory about taking decision-making machinery off line and feeding it pretend inputs could turn out to be false without endangering the co-cognition theory. The example that she gives to illustrate this latter point has theoretical and practical inference underpinned by information-processing machinery in which 'inference mechanisms do not take beliefs as input but rather take items which represent or encode propositional contents without attached attitudes' (Heal, 1998a, p. 494). In this example, it would not be correct to say that the machinery is on line and takes real beliefs as inputs when it is used in the service of decision-making but is off line and is fed pretend beliefs when it subserves decision-predicting.

Heal is surely right that there could be a common core in the machinery that underpins inferential transitions, whether these figure in the context of belief, supposition, pretence or imagination. At the level of information-processing machinery, it might be that no priority attaches to beliefs. She is also right, in our view, to maintain that if things were to turn out this way then that would still be consistent with the basic idea of the simulation theory.

¹⁷ Cf. Heal, 1994, p. 138, n. 4.

be thinking about the square on the hypotenuse being 25 units. Vincent and Yvonne would be engaged in co-cognition.

This result generalises to yield a thin thesis about thinking of the same subject matter. If Vincent is able to think that Yvonne believes that p then Vincent is able to think (though perhaps not to believe) that p . But nothing follows from this thin thesis alone as to how one person should actually set about predicting the thoughts of another. To see this, we only need to observe that the thin thesis that holds for ‘Yvonne believes that’ holds also for ‘Yvonne dreams that’, or ‘In her notes towards a surrealist novel, Yvonne says that’, or ‘Yvonne is upset by the memory that her mother once doubted that’. The thin thesis about thinking of the same subject matter seems to be entirely compatible with the claim that substantive bodies of empirical psychological theory are required for predictions about dreams, novel writing, or emotional upsets. So how can the thin thesis rule out a theory-theory account of actual decision-predicting?¹⁸

In order to understand Heal’s position on this issue, we need to take account of the fact that she is concerned only with a limited range of cases of folk psychological prediction. The key examples are predicting another person’s belief given information about other beliefs and interests (e.g. predicting ‘She believes that q ’ given the information that ‘She believes $p_1 - p_n$ and is interested in whether q ’), and similar cases of predicting the results of practical reasoning (ibid., 479–80).

If Vincent is capable of thinking that Yvonne believes that $p_1 - p_n$ and that she is interested in whether q then Vincent himself is capable of entertaining the thoughts $p_1 - p_n$ and q . Using his own ability to grasp these thoughts and to appreciate their inferential potential, he can suppose that $p_1 - p_n$ and consider whether q follows. Even if Vincent does not share Yvonne’s belief that $p_1 - p_n$ are all true he will, in considering this question, be aiming to track connections that are relevant to truth. Furthermore, he can assume that, as Yvonne tries to work out whether q , she too is aiming to track the truth.¹⁹ Suppose that Vincent reaches the conclusion that q does follow from $p_1 - p_n$. He will appreciate that departures from right reasoning are eminently possible. But, in the absence of any grounds for expecting mistakes on Yvonne’s part or his own, Vincent can reasonably predict that Yvonne will believe that q . So in this kind of case it is at least plausible, and on relatively a priori grounds, that there is a method that Vincent can use to make a prediction about Yvonne’s belief without having to rely on a stored body of knowledge about what else people who believe that $p_1 - p_n$ typically believe. This method is available to Vincent because he can think about the same subject matter as Yvonne.

Thus, the a priori considerations in favour of the co-cognition theory do support the claim that, at least for this circumscribed range of cases, an alternative to the theory-theory method of predicting decisions and beliefs is *available*. But those considerations do not, by

¹⁸ For a critical discussion of Heal’s claim that the co-cognition claim can be established a priori, see Nichols and Stich, 1998, esp. pp. 504–11.

¹⁹ He would not be entitled to make the corresponding assumption about Yvonne’s dreaming or her notes for a surrealist novel. Nor does the fact that q follows from p provide any guidance as to whether Yvonne will be upset by the memory that her mother once doubted that q given that she is upset by the memory that her mother once doubted that p .

themselves, establish that this alternative method is actually *used*. So there remains a question about Heal's view of the relationship between her a priori, personal-level cognition theory and empirical, subpersonal-level theories of the kind that Goldman and Stich and Nichols disagree about. We shall not pursue this question further here.

In an earlier paper, Heal is even more explicit about the circumscribed domain of mental simulation (1996, p. 56):

The kind of simulationism I would like to defend says that the only cases that a simulationist should confidently claim are those where (a) the starting point is an item or collection of items with content, (b) the outcome is a further item with content, and (c) the latter content is rationally or intelligibly linked to that of the earlier item(s).

This way of expressing the restriction highlights a further aspect of the difference between Heal's version of the simulation theory and Goldman's. For, when Goldman (1989) first presented his version, it was as an alternative both to the theory theory and to the 'rationality approach' to attributing beliefs to other people. It is true that Goldman characterises the rationality approach as involving the strong assumption that 'the agent in question . . . conforms to an ideal or normative model of proper inference and choice' (1989/1995, p. 75) while Heal's assumption is only of 'very minimal rationality' (1998a, p. 487). But the difference of approach remains. For Goldman, it is similarity rather than rationality that is important to the success of simulation.²⁰

2.3 Personal-level and empirical or subpersonal-level and a priori

Goldman's and Heal's versions of the simulation theory differ on both the personal versus subpersonal dimension and the empirical versus a priori dimension. But not all personal-level simulation theorists share Heal's view about the epistemological status of the theory.

Robert Gordon certainly pitches his version of the simulation theory at the personal level. Here, for example, is what he says about the term 'off-line simulation theory' (1992b/1995, p. 174):

This packs into the very name of the theory what I regard as an ancillary hypothesis: that when we simulate – that is, use our imagination to identify with others in order to explain or predict their behavior – many of the same cognitive systems that normally control our own behaviour continue to run as if they were controlling our behavior, only they run off-line. . . . This hypothesis is very plausible . . . But I think the imaginative identification theory remains attractive even without the off-line hypothesis.

This sounds very much like Heal, and Gordon also agrees with Heal on the importance of the point that in simulating another person I think of the same subject matter. I can often come to understand what another person has done, or predict what he will do, by directing my

²⁰ See Heal, 2000, for an extended defence of the rationality approach over the similarity approach.

attention onto the environment that I share with him and thinking, as he thinks, about that subject matter.²¹

So Gordon accepts the idea of co-cognition. But he does not link mental simulation with an assumption of rationality as Heal does (Gordon, 1992a/1995, p. 104):

[The simulation theory] predicts – correctly – that we will sometimes attribute irrationality to others. . . . [It] even predicts – again correctly, it would appear – just *when* we are most likely to do so: namely, when we are most prone to such irrationality ourselves.

For Gordon, as for Goldman, it is similarity that is important rather than rationality.

When I look at the environment that I share with someone whose behaviour I am trying to understand, I look for relevant features of that environment. But (ibid.), ‘we are not in general trying to find good reasons, whether they move us or not; rather we are seeking to be moved, whether rationally or not’. I may come to understand what another person has done by identifying a feature of the environment that moves me emotionally as it moved him. But I may also recognise that it is irrational for me to be moved in that way and I may attribute that same irrationality to the other person.

Both Gordon and Goldman point to the phenomenon of emotional contagion, along with motor mimicry, as counting in favour of the simulation theory.²² This is one indication that Gordon shares with Goldman an empirical approach even though Gordon does not cast his version of the simulation theory in cognitive-scientific terms. We can also see that Gordon and Heal are likely to differ over the epistemological status of the simulation theory by observing the use that Gordon makes of a striking a posteriori fact about emotion.

Our emotional responses to imagining being in a terrifyingly dangerous situation, for example, are similar to the responses that we would have if we were really in such a situation.²³ So, suppose that Vincent knows that Yvonne is, and believes herself to be, in an extremely dangerous situation and that he wants to predict her emotional response. Vincent imaginatively identifies with Yvonne. He imagines being in that dangerous situation and, as a result, he has a real emotional response of his own. This is the response that he predicts she will have. It is at least plausible, then, that there is a method that Vincent can use to make a prediction about Yvonne’s emotional response without having to rely on a stored body of knowledge about the typical emotions of people who believe themselves to be in extreme danger.

On Gordon’s account, the prediction of emotional responses is a favourable case for simulation theory. But notice that the effectiveness of the simulation method depends on the

²¹ For this ‘direction of gaze’ point, see Heal, 1986/1995, p. 48: ‘[The replicator] is not looking at the subject to be understood but at the world around that subject’; and cf. Gordon, 1992a/1995, p. 102. Note that Gordon also shares with Heal the view that folk psychological explanations are different from covering-law explanations (ibid., p. 117): ‘An explanation in terms of the agent’s reasons for acting should enable one to understand what features of the world made the action attractive to the agent. This requires that we see the world, and the agent’s situation in it, as if through the agent’s eyes. The empathetic method gives us all the explanatory understanding we want.’

²² See, e.g., Gordon, 1996, p. 13; Goldman, 1995, pp. 197–9.

²³ Walton, 1997. See also Goldman, 1995; Ravenscroft, 1998.

a posteriori fact about emotional responses. We can contrast this with the cases that are central for Heal. Victor's success in predicting that Yvonne will believe that q depends on the fact that his grasp of the thoughts $p_1 - p_n$ and q furnishes him with an appreciation of their inferential potential whether they occur in the context of belief, supposition, pretence or imagination. But this is an a priori, rather than an a posteriori, fact about grasp of thought contents.²⁴

Although Gordon is a personal-level simulation theorist, his view of the epistemological status of the simulation theory is like Goldman's rather than Heal's. Various kinds of empirical evidence could count against the personal-level simulation theory including findings about information-processing mechanisms. So Gordon returns a straightforward answer to the question how his version of the simulation theory would fare if empirical investigation revealed information-processing machinery with a structure as indicated by Stich and Nichols's version of the theory theory:²⁵

I would count it as a blow to my view if subpersonal systems involved in practical and theoretical reasoning were not also involved in predicting such reasoning (in hypothetical, counterfactual, or other-person cases). . . . If it's all done by mentalese entailments from mentalese generalizations, then I think I'd be ready to say that the simulation theory is wrong about *how* we do it. Even if the simulation theory fits the phenomenology, phenomenology proves to be a poor guide to *how it's really done*.

We now have occupants for three of the cells in a two-by-two array (see Table 1). Heal, in the top left cell, is a personal-level simulation theorist who regards the (co-cognition) theory as being a priori in character. Diagonally opposite is Goldman who is frankly cognitive-scientific. By his lights, the simulation theory is a straightforwardly empirical subpersonal-level theory. In the top right cell is Gordon, similar to Heal on level of description but to Goldman on epistemological status. The bottom left cell is empty. As far as we know, no one thinks that the simulation theory is a subpersonal-level theory about information-processing machinery but that its correctness can be established by a priori means.

²⁴ We left it open whether the a priori considerations in favour of co-cognition can establish that the simulation method is actually used for predicting beliefs and decisions. But what matters for the contrast between Heal and Gordon is that, for Heal, a priori facts about grasp of thought contents explain the success of the simulation method if it is used for predicting beliefs and decisions. In contrast, Gordon relies on an a posteriori fact about emotion in order to explain the effectiveness of the simulation method for predicting emotional responses.

We note that Heal says that the prediction of behaviour that is expressive of emotions 'falls within the remit of simulationism' (1996, p. 58). Actions that arise from emotions can often be understood 'from the inside' as rational or intelligible. She also says that emotions are closely connected with value judgements and that, 'As far as rationalizing and making intelligible are concerned it is the value judgments that do the work' (ibid., p. 60). So, if we are concerned with predicting the rational or intelligible consequences of emotions, then there need be no more to simulating an emotion than 'entertaining the content of the associated value judgments' (ibid.). It is not clear whether this is to be transposed into an account of the prediction of emotional responses themselves.

²⁵ Gordon, personal communication.

	A priori	Empirical
Personal level	Heal	Gordon
Subpersonal level		Goldman

Table 1: Classification of simulation theorists

3. The role of theory in mental simulation

In mental simulation, I simulate, replicate or re-enact aspects of the mental life of another person. These aspects may include, for example, the other person's thinking, decision-making or emotional responses. In this section, we consider the role that elements of psychological theory play in the use of mental simulation.

3.1 Simulation in reality and in imagination

In some cases of mental simulation, I am able to place myself in just the same situation as the other person and to go through the same thoughts, reach the same decisions, and have the same emotional responses as he does. We call this 'simulation in reality'.²⁶ Simulation in reality is sometimes a useful way of learning about aspects of the mental life of another person, but it is apt to be an expensive and impractical methodology. If I want to know about your beliefs, decisions or feelings when your cat died then I do not buy a cat and arrange for its death. What I may do, however, is to imagine first owning a cat and then learning of its demise; and the term 'mental simulation' is usually taken to include both simulation in reality and simulation in imagination.²⁷

The difference between simulation in reality and simulation in imagination is important for an assessment of the role of theory in simulation. Thus, consider cases in which the behaviour of an object under particular actual or hypothetical circumstances is simulated by the behaviour of another object of the same kind under the same or similar circumstances. For example, suppose that I simulate the increase of pressure in a cylinder of gas that would result if its temperature were raised by heating another cylinder of the same gas, starting from the same initial temperature and pressure. I can make a prediction about what will or would happen to the first cylinder by observing what actually happens to the second. Similarly, I can simulate the effect of a drug on one person by having another person ingest the same substance and, once again, the simulation may generate a prediction. In fact, I can simulate the effect of a drug on another person by taking the drug myself.

²⁶ Stich and Nichols, 1997, p. 302, call it 'actual-situation-simulation'.

²⁷ Stich and Nichols, *ibid.*, p. 303, use the term 'pretence-driven-off-line-simulation' where we use 'simulation in imagination'.

In these cases of the cylinder and the drug, the simulation takes place in reality: the cylinder is really heated, and I really take the drug. The same processes occur in the simulation as would be operative in generating the behaviour of the object being simulated. This aspect of the exercise is vital if simulation is to provide an alternative to the deployment of theoretical knowledge about gases or drugs. If I were merely to imagine heating the cylinder or taking the drug then, in order to develop the simulation exercise, I would need to draw on a body of information about temperature and pressure in gas cylinders or about the effects of the drug on human beings. Antecedent knowledge would be needed to answer the question, ‘What happens next?’

In these examples, simulation in reality is, in Goldman’s terms,²⁸ ‘process-driven’ while simulation in imagination is ‘theory-driven’. It is only simulation in reality that constitutes a genuine alternative to the use of empirical theory in making a prediction. If this is the general pattern then there is a problem for the idea that mental simulation offers an alternative to the use of theory for arriving at psychological predictions. As we have already noted, mental simulation in reality is apt to be expensive and impractical. So mental simulation will usually be simulation in imagination. But then, the pattern that we have observed in the cases of the cylinder and the drug suggests that mental simulation will, for the most part, be theory-driven simulation.²⁹

However, a distinctive claim of the simulation theory can be expressed by saying that, unlike gas-cylinder simulation in imagination, mental simulation in imagination can be process-driven rather than theory-driven. This is because at least some mental processes operate in the same way when I imagine being in a particular situation as they would if I were really in that situation. For example, thinking and decision-making seem to proceed in the same way from imagined or hypothetical premises about my situation as from premises that I really believe to be true. So, by engaging in thinking or decision-making within the scope of the pretence or supposition that I am in a specific situation, I am able to reach a view about what I would think or decide if I were really in that situation. And I do this, it seems, without making use of an antecedently known theory about what people typically think or decide in that kind of situation.

According to the simulation theory, mental simulation in imagination can be process-driven, rather than theory-driven, simulation. But there remains the question whether prediction by mental simulation in imagination must rely on at least some elements of psychological theory.

3.2 The minimal theoretical background for mental simulation

The simulation theory and the theory theory are supposed to be opposed to each other and the theory theory says that our folk psychological practices draw on a body of stored information about psychological matters. But it would not be right to characterise simulation theorists as denying outright that knowledge of psychological generalisations is implicated in everyday psychological prediction, interpretation and explanation. To see why there is still a

²⁸ Goldman, 1989/1995, p. 85.

²⁹ Thus Dennett, 1981, p. 000: ‘How can [simulation] work without being a kind of theorizing in the end?’

role for psychological principles, suppose that I imagine being in your situation. I take on in imagination first, ownership of a cat and then, receipt of news of its death and from this starting point I engage in a simulation exercise. Suppose that things proceed just as the simulation theory says. Without my having to draw on any antecedent knowledge about the typical mental lives of cat-bereaved people, the simulation unfolds in a particular way. Within the scope of the simulation exercise, I experience a feeling of sadness, I have the thought that I shall need to buy less milk in future, and I reach the decision not to own any more pets. As a result of this simulation in imagination, I come to understand something of what it was like for you when your cat died,³⁰ and I attribute the feeling, the thought and the decision to you.

Strictly speaking, however, simulating another person's mental life is one thing and attributing mental states to another person is something else. I could engage in the simulation exercise and, within its scope, I could experience an emotion, have a thought and reach a decision. But then I might make no attribution at all. Or I might just make a hypothetical attribution to myself. That is, I might judge that if I were really in the simulated situation then I would really experience that emotion, have that thought and reach that decision.

If I do take the further step of attributing feelings, thoughts and decisions to you then this must rest on my accepting (or not calling into question) some psychological principle. The principle might link other people's mental lives (including your mental life) to my own. Thus, perhaps:

(Me–You) If, in circumstances C, my mental life would be thus-and-so then if O is in circumstances C then *ceteris paribus* O's mental life is thus-and-so.

Such a principle would, of course, have to be used in conjunction with another one linking my mental life within the scope of simulation exercises with my mental life in hypothetical circumstances:

(Sim–Me) If, within the scope of a simulation that starts from imagined circumstances C, my mental life is thus-and-so then, in circumstances C, my mental life would *ceteris paribus* be thus-and-so.

Alternatively, I might rely on a single principle that links other people's mental lives (including yours) directly with the way that my mental life unfolds within the scope of simulation exercises:

(Sim–You) If, within the scope of a simulation that starts from imagined circumstances C, my mental life is thus-and-so then if O is in circumstances C then *ceteris paribus* O's mental life is thus-and-so.

Since I myself am a possible value of 'O', (Sim–You) actually subsumes (Sim–Me).

Someone might stress that, as our mental lives unfold, it is not the case that one mental state brutally and unintelligibly follows another. This is especially clear in the case of

³⁰ Here we touch on the point that, according to the simulation theory, understanding of a third person has a first-personal component in it. Mental simulation offers understanding of another person 'from the inside'. See Heal, 1998b; see also Gordon, 1992a/1995, p. 117; 1995, p. 56.

theoretical reasoning. One thought follows another because the thinker appreciates that it follows *from* the other. I think B after thinking A because I appreciate that A entails B; because, given that I already think A, B is the thing to think. When I engage in mental simulation, I think and reason critically just as I do in reality. So a simulation exercise may yield normative knowledge. The upshot may be not merely that within the scope of a simulation that starts from imagined circumstances C my mental life is thus-and-so, and not only that in circumstances C my mental life would be thus-and-so, but that in circumstances C one's mental life should be thus-and-so. If this is indeed the kind of knowledge that mental simulation yields then the principle that is needed to underwrite attributions of mental states to other people is different from (Me–You) or (Sim–You). It is, rather, along the following lines:

(Norm–You) If, in circumstances C, one's mental life should be thus-and-so then if O is in circumstances C then *ceteris paribus* O's mental life is thus-and-so.

In a slogan: *Ceteris paribus*, people think the thing that is the thing to think.³¹ This is, perhaps, an unfamiliar kind of psychological principle, with its explicit use of normative vocabulary. But it is still a generalisation about psychological matters and, if mental simulation yields normative knowledge, then this principle should be acknowledged as a theoretical component within the mental simulation approach.³²

Goldman is explicit, from the outset, that something like (Me–You) figures as a background assumption for prediction by simulation (1989/1995, p. 89):

Is this impressive success [in predicting human behavior as well as we do] fully accounted for by the simulation theory? Only, I think, with an added assumption, viz., that the other people, whose behavior we predict, are psychologically very similar to ourselves.

In a recent paper (2000, p. 000), he mentions a background assumption that sounds more like (Sim–Me) or (Sim–You):

At the heart of the simulation hypothesis is the notion that pretend mental states have some sort of intimate similarity, or homology, to their non-pretend counterparts. That is how they can succeed, on many occasions at least, in generating the same outputs as the non-pretend states would do when fed into the same processing system.

In her first paper on mental simulation, Heal is also explicit about the need for a background assumption (1986/1995, p. 47):

I can harness all my complex theoretical knowledge about the world and my ability to imagine to yield an insight into other people *without any further elaborate theorizing about them*. Only one simple assumption is needed: that they are like me in being

³¹ See Stone and Davies, 1996, pp. 136–7; Davies and Stone, 1998.

³² Heal, 1986/1995, p. 50, says that, to the extent that mental simulation requires knowledge of principles, these are normative or semantic rather than causal.

thinkers, that they possess the same fundamental cognitive capacities and propensities that I do.

According to Heal, then, there are three things that I need in order to adopt the simulation approach. First, I need to be able to imagine. I begin the simulation exercise by ‘imagining the world as it would appear from his [the other person’s] point of view’ (ibid.). Second, I need to be able to think – ‘deliberate, reason and reflect’ – and in doing so to draw on my knowledge about the world that the other person and I both inhabit. And third, I need to make an assumption about psychological similarity between the other person and myself.

This sounds like (Me–You).³³ But in later writings it is clearer that, for Heal, the crucial background assumption is about rationality³⁴ and so is closer to (Norm–You) (1998a, pp. 486–7):

The idea that connections in thought follow connections between states of affairs is the idea that we are rational – in some sense of that term. . . . Given the assumption of such very minimal rationality, we can show why reliance on co-cognition is a sensible way to proceed in trying to grasp where another’s reflections will lead.

Our view is that it is almost inevitable that the simulation approach to mental state attribution must invoke some such psychological principles as those that we have been discussing. But perhaps it is not absolutely inevitable.

Gordon says (1995, p. 60): ‘to ascribe to O a belief that *p* is to assert that *p* within the context of a simulation of O’.³⁵ The proposal is, as Gordon acknowledges, problematic and in need of development. But if it could be made to work then, it seems, attributions based on mental simulation would not need to be underwritten by acceptance of principles like the ones that we have been discussing. The reason, in essence, is that (at least on one construal) Gordon does not accept the starting point of our discussion of the role of theory in simulation, namely, that simulating another person’s mental life is one thing and attributing mental states to another person is something else. According to Gordon’s proposal (so construed), my apparent attribution to you of the belief that you will need to buy less milk in future is not distinct from my asserting or judging, within the scope of my simulation of you, that I shall need to buy less milk in future.

This should clearly not be taken as the proposal that the truth conditions of an attribution, ‘O believes that *p*’, are simply those of the attributor’s own assertion that *p*, nor that they are those of ‘I [the simulator] asserted that *p* within the context of a simulation of O’. But it is fairly natural to read Gordon as saying that apparent attributions of beliefs are not really *fact-stating* at all; instead they are *expressions* of the simulator having asserted that *p* within the scope of a simulation of O, rather as ‘ouch’ is an expression of pain.

³³ Ibid.: ‘To get good 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.’

³⁴ Heal stresses that the rationality assumption that is needed is very weak. See also Heal, 2000, p. 14: ‘I credit [another thinker] with ability to see obvious entailments, to add up short columns of figures, to avoid clearly evident fallacies, to follow rules of inference in which she has been trained, and the like.’

³⁵ An earlier expression of the same idea is this: ‘to attribute a belief to another person is to make an assertion, to state something as a fact, *within the context of practical simulation*’ (1986/1995, p. 68).

If apparent attributions of mental states do not have truth conditions but are expressions of what takes place in a simulation exercise then they do not need to be underwritten by appeal to such psychological principles as (Sim–You). But unless we are prepared to take this radical step of denying that apparent attributions of mental states are fact-stating, it does seem inevitable that the simulation approach to attribution must invoke some psychological principles.³⁶ As Heal says (2000, p. 7): ‘There is . . . no way of reconstructing the reasoning [that leads via a simulation exercise to attribution of a belief to another person] on which it both avoids any sort of rationality assumption and also avoids being a version of the argument from analogy.’

This is not a striking new discovery, or admission, about mental simulation. As we have seen, it has been explicit since the beginning of the debate that attributions based on mental simulation must rely on at least some elements of psychological theory. The contrast between the theory-theory and the simulation-theory approaches is not between the knowledge-rich and the knowledge-free but between the knowledge-rich and the knowledge-poor.³⁷

In fact, simulation theorists usually allow that rather more in the way of psychological knowledge is actually used than just the principles that we have been discussing. Goldman, for example, is very clear about this:³⁸

I am not saying . . . that simulation is the *only* method used for interpersonal mental ascriptions, or for the prediction of behavior. Clearly, there are regularities about behavior and individual differences that can be learned purely inductively.

It may be said that this step beyond the minimal theoretical background is a step away from the pure simulation theory and towards a hybrid theory. But the more important point is that the simulation theory itself (apart, perhaps, from Gordon’s version) affirms that the use of mental simulation relies on theoretical assumptions about similarity or rationality.

³⁶ Gordon (personal communication) has pointed out another way to construe his proposal. What he intends is that simulation provides an understanding of others as ‘subjects to whom things are a certain way’. So my simulation of you leads to an understanding of you as a subject to whom it is the case that you yourself will need to buy less milk in future. Apparent attributions of beliefs are, then, fact-stating. But the facts that they state are not ‘reducible to facts congenial to the natural sciences’. They are facts about how things are from a simulated point of view or through a simulated ‘peephole’ and they cannot be grasped or understood ‘except in the first person’. Gordon would resist the claim that facts about ‘subjects to whom things are a certain way’ are still psychological facts. But, in our view, what is really crucial to his position is that they are not facts in the domain of scientific psychology. Similarly, he would reject the idea that the simulation approach must invoke some psychological principles. But, in our view, he only needs to reject the idea that the simulation approach invokes principles that could figure in scientific psychology. So we would suggest that attributions based on Gordon’s version of the simulation approach must rely on a principle along the following lines: If, within the scope of a simulation that starts from imagined circumstances C, I assert that *p* then if O is in circumstances C then *ceteris paribus* O is a subject to whom it is the case that *p*. We would also suggest that it is harmless to describe this as a psychological principle provided that we also stress that the notion of a subject to whom it is the case that *p* is not a notion that could figure in a scientific, or otherwise wholly third-personal, theory.

³⁷ Goldman, 1992/1995, p. 191.

³⁸ Goldman, 1989/1995, p. 83. See also e.g. Heal, 1986/1995, p. 48; 1995, pp. 46–7; Gordon 1986/1995, pp. 105–6.

3.3 *What remains at issue in the debate*

Frank Jackson says that ‘the theory-theory account of how we should predict human behaviour and human mental states must be the correct one’ (1999, p. 77). This announcement rests in the first instance on the fact that, when I answer a question about what some other person will think or do, I draw on my own view or belief about what that person will think or will do. So I draw on a little piece of theory. In response to that claim, Goldman or Heal can say that it is not in dispute that a verbalised folk psychological prediction is the expression of a belief about what another person will think or do. The debate is about how such predictions are reached. If the theory theory of folk psychological prediction is correct then specific predictions about what a person will think or do follow from a body of stored information about psychological matters, given specific information about the circumstances of the person in question. The immediate build-up to Jackson’s dramatic announcement does not seem to address the question whether predictions draw on stored information about psychological matters. But that question is addressed later, when Jackson discusses Kahnemann and Tversky’s example of Mr. Crane and Mr. Tees.³⁹

Jackson agrees with Goldman that people reach a view about whether Mr. Crane or Mr. Tees would be more upset about missing his flight by placing themselves, in imagination, into the same situation.⁴⁰ He can agree with Goldman (1989/1995, p. 83) that ‘people do not possess a tacit folk-psychological *theory* that warrants any particular answer to this question [Who is more upset?].’ But, Jackson says, in order to reach our view about who is more upset we do use two pieces of theory (1999, p. 88):

One is a view about the capacity of a certain kind of mental exercise to reveal how you would feel in some given situation. The other is a view about the relevance of information about what you would feel to what [Mr. Crane and Mr. Tees] would feel.

We rely, in short, on something like the two principles (Sim–Me) and (Me–You). What mental simulation yields is a premise that leads, via those two principles, to the conclusion that Mr. Tees would be more upset.

As we have already noted, it has been explicit since the beginning of the debate that, in order to make a prediction by using mental simulation, I must assume such principles as (Sim–Me) and (Me–You), or (Sim–You), or (Norm–You). There is a minimal theoretical background for mental simulation. Jackson’s argument confirms that some such principles are needed. But he also claims that neither the simulation theory nor anything else can be a genuine competitor to the theory theory. All that can remain at issue is what theory it is that we use in our folk psychological practices.

³⁹ Kahnemann and Tversky, 1982. The example was introduced into the theory versus simulation debate by Goldman, 1989/1995, p. 83.

⁴⁰ Jackson, 1999, p. 88: ‘[The opinion that Mr. Tees will be more upset] comes from imagining themselves in the relevant situation, discovering how they’d feel, and then inferring that [Mr. Crane or Mr. Tees] will feel the same way.’ Cf. Goldman, 1989/1995, p. 83: ‘How did [subjects] severally arrive at the same answer? Clearly, by simulation. They imagined how they would feel in Mr. Crane’s and Mr. Tees’s shoes, and responded accordingly.’

It is important here to distinguish two conceptions of the theory theory. On Jackson's conception, the theory theory maintains only that folk psychological predictions rely on some theoretical principles. So the theory theory has to be correct, because a minimal theoretical background is acknowledged even by the simulation theory. On our conception, in contrast, the theory theory maintains that a prediction about what a subject, O, will do in circumstances C draws on a stored body of psychological information that is logically adequate to yield an answer to the question what O will (all else equal) do in circumstances C. So, even when the minimal theoretical background for simulation is acknowledged, it is still possible to regard the simulation theory and the theory theory as competing accounts of our folk psychological practices. In the example of Mr. Crane and Mr. Tees, we can conceive of the kinds of information about normal psychological processes that would permit an answer to the question, 'Who is more upset?', to be derived. If, as Goldman and Jackson seem to agree, subjects answer the question without drawing on antecedently available knowledge of such kinds then, on our conception, the theory-theory account of the example is not correct. Both conceptions of the theory theory are legitimate but, in our view, the second conception is better suited to commentary on the theory versus simulation debate as it exists in the literature.⁴¹

The contrast between the simulation theory and the theory theory is not undermined simply by the acknowledgement that predictions based on mental simulation rely on theoretical assumptions. The simulation theory's core claim is that, over a significant range of cases, the methodology of mental simulation allows us to avoid the need for detailed antecedent knowledge about how psychological processes typically operate. The plausibility of this claim rests on the point (Section 3.1) that mental simulation in imagination can be process-driven rather than theory-driven. But Heal (1994) argues that, if the simulation theory is conceived as Goldman conceives it, then the cognitive mechanism that is used for process-driven simulation will turn out to embody tacit knowledge of a theory about how the processes that it is used to simulate typically operate. It is time to turn directly to this threat of collapse, beginning with the notion of tacit knowledge.

4. Tacit knowledge

Stich and Nichols claim that the 'dominant explanatory strategy' in cognitive science is to credit people with tacit knowledge of theories (Stich and Nichols, 1992/1995, p. 123). So, by their reckoning, the dominant cognitive-scientific version of the theory theory says that our engagement in folk psychological practices is explained by the presence of a 'largely tacit psychological theory' (ibid., p. 124).

Goldman (1989) raises a number of problems that apply most directly to versions of the theory theory that postulate ordinary verbalisable knowledge of psychological generalisations. We shall mention just two. One objection starts from the fact that ordinary folk are not very good at articulating principles of psychological theory: '[W]hy, one

⁴¹ Jackson, 1999, pp. 90–1, quotes some editorial comments on the simulation theory by the present authors and notes that the correctness of the simulation theory as we describe it would not count against the theory theory as he conceives it. He also quotes a passage by the principal participants on the theory-theory side of the debate (Stich and Nichols) and says that it 'starts the debate off on the wrong foot' (ibid., p. 93).

wonders, should it be so difficult to articulate laws if we appeal to them all the time in our interpretive practice?’ (1989/1995, pp. 80). The other objection begins from the point that, if folk psychological practice depends on knowledge of psychological theory, then this knowledge would need to be present in children of around five years of age. So that raises a problem about acquisition: ‘Are such children sophisticated enough to represent such principles? And how, exactly, would they acquire them?’ (ibid.).

As Goldman anticipates, a natural reaction to these two objections is that if they were really telling against the theory theory then there would be correspondingly telling objections against the claim that our linguistic practices rest on knowledge of principles of grammar. But almost nobody now thinks that there are good objections to the whole enterprise of Chomskyan linguistics starting from the fact that ordinary folk are not very good at articulating grammatical principles. Nor is linguistics threatened by a problem about early acquisition. The linguist can respond to the two putative objections by saying, first, that knowledge of language is partly tacit and, second, that it is partly innate.

Similarly, the theory theorist about folk psychology can respond to Goldman’s objection about lack of articulation by saying that the psychological theory is tacitly known, and can respond to the objection about early acquisition by saying that some elements of the psychological theory are innate. So Goldman allows that it is open to the cognitive-scientific theory theorist to postulate tacit knowledge of a psychological theory although, of course, he argues that the tacit-theory theory is not well motivated given the availability of the simulation alternative.⁴² But it would be seriously problematic for his overall position if Heal’s argument were to establish that the cognitive-scientific simulation ‘alternative’ is really no alternative at all to the tacit-theory theory.

4.1 Two worries about trivialisation

There is a challenge that the theory theorist must meet if he invokes the notion of tacit knowledge; namely, that the notion of a tacitly known theory should not be trivial. He needs to meet the charge of ‘promiscuity’ levelled at the theory theorist by Simon Blackburn (1992/1995, p. 275):

If we are good at something . . . then we can be thought of as making tacit (very tacit) use of some set of principles that could, in principle, provide a description of a device, or possibly a recipe for the construction of a device, that is also good at it.

According to this notion of tacit knowledge, ‘any activity that ends up with a belief counts as forming it by a process of theorizing’ (ibid.). The point is not that the belief that is arrived at itself counts as a little piece of theory.⁴³ It is, rather, that anyone who is able to arrive at particular beliefs on the basis of various considerations will count as having tacit knowledge

⁴² Goldman, 1989/1995, p. 80. In a later paper he explicitly rejects the idea that an adequate motivation for favouring the tacit theory theory is provided by existing practice in cognitive science (1992/1995, p. 191–3). We are not concerned here with the issue of nativism. In fact, Goldman, 1992/1995, p. 194, agrees with Fodor (1987; see also 1992) that it is more plausible to suppose that folk psychological knowledge is innate than that it is either extracted from experience or culturally transmitted.

⁴³ Cf. Jackson, 1999; and see above Section 3.3.

(‘tacit (very tacit)’) of any and every set of principles that could be used to mediate the derivation of those beliefs from those considerations.

So consider someone who carries around a cylinder of gas which he heats in order to simulate the increase of pressure in other cylinders of gas when they are heated. This person is good at something, namely, predicting pressure given temperature. On the very thin notion of tacit knowledge that Blackburn mentions, this person counts as having tacit knowledge of a generalisation relating temperature and pressure in gas cylinders. For someone with explicit knowledge of that generalisation would be able to arrive at the same beliefs that the predictor reaches by gas-cylinder simulation. If this were the only available notion of tacit knowledge then, of course, the simulation theory would turn out to be wholly consistent with the tacit-theory theory.⁴⁴

Before we turn to refinements of the notion of tacit knowledge that avoid this first worry about trivialisation, we should briefly consider an even more basic worry. It might be said that, for the notion of tacit knowledge that Blackburn mentions, we attribute tacit knowledge of principles to someone when it is, in certain respects, *as if* that person were making use of those principles. But John Searle says (1990, p. 587): ‘Everything in the universe . . . behaves with a certain degree of regularity, and for that reason everything behaves *as if* it were following a rule . . . For example, suppose I drop a stone. The stone . . . *follows the rule* $S = \frac{1}{2}gt^2$.’ So the worry that looms is that planets and rocks and billiard balls will all turn out to embody tacit knowledge of generalisations about the behaviour of objects of their kind: ‘it makes everything in the universe mental’ (ibid.).

The worry that we took from Blackburn was that a person who engages in gas-cylinder simulation will count as having tacit knowledge of the generalisation relating temperature and pressure. But now the concern is that the gas cylinder itself will embody tacit knowledge of the generalisation relating temperature and pressure, and will do so quite independently of whether anyone engages in gas-cylinder simulation. Similarly, the cognitive machinery involved in making decisions will turn out to embody tacit knowledge of a psychological theory about how decision-making proceeds.

The response to this second, more basic, worry about trivialisation is that, just as ordinary knowledge of a principle allows a thinker to arrive at beliefs, so tacit knowledge of a principle is supposed to explain the generation of, or transitions between, thoughts or other states with representational contents. Knowledge of the generalisation relating temperature and pressure allows a thinker to move from a premise about temperature to a conclusion about pressure. So tacit knowledge of that generalisation would contribute to an explanation of transitions from a state that represents a gas cylinder’s temperature to a state that represents its pressure. Since a gas cylinder’s temperature and pressure are not themselves representations of temperature and of pressure there is no prospect of an unwanted attribution of tacit knowledge.⁴⁵

⁴⁴ Blackburn, *ibid.* He notes that the tacit-theory theorist may impose further conditions on tacit knowledge, but he does not pursue the issue.

⁴⁵ See Davies, 1995b, p. 382.

Similarly, knowledge of generalisations about how decision-making proceeds would allow a thinker to proceed from premises about someone having particular beliefs and desires to a conclusion about that person reaching a particular decision. So tacit knowledge of a theory about decision-making would be implicated in transitions from states that represent people believing and desiring things to states that represent people deciding things. But the representations that are inputs and outputs of the machinery that is involved in making decisions are not representations of people believing, desiring and deciding things. The representation that corresponds to a belief, for example, concerns the world, not the believer. So there would be no basis for attributing tacit knowledge of a psychological theory about decision-making to the mechanisms of decision-making.

Blackburn's worry about triviality is not Searle's. Blackburn begins, not from a gas cylinder reaching a certain pressure but from someone reaching a belief about the pressure in the gas cylinder. In fact, Blackburn's worry about tacit (very tacit) knowledge is not very different from Quine's challenge (1972) to the notion of tacit knowledge when it was introduced by Chomsky (1965). Quine challenges Chomsky's introduction of the notion of tacit knowledge by making use of the distinction between behaviour that conforms to a rule and behaviour that is guided by a rule. A subject can behave in a way that conforms to a rule without using the rule to guide his behaviour for, as Quine uses the notion of guidance (1972, p. 442): '[T]he behavior is not *guided* by the rule unless the behavior knows the rule and can state it.' Guidance requires verbalisable knowledge.

Chomsky's tacit knowledge is supposed to require less than verbalisable knowledge; but it cannot be equated with mere conformity.⁴⁶ There will always be alternative sets of rules that require just the same behaviour for conformity; but it is part of the idea of tacit knowledge that a speaker's actual behaviour might be correctly explained in terms of tacit knowledge of one set of rules rather than the alternatives. Viewed in the context of Quine's challenge, the notion of tacit (very tacit) knowledge that Blackburn mentions can be seen as requiring no more than conformity. But Blackburn's focus on the ability to arrive at beliefs makes it clearer than does Quine's talk of behaviour that the relevant notion of conformity concerns transitions amongst representations.

It is because tacit knowledge is supposed to be explanatory that it is not trivial. But it is at this point that Quine (1972) presses his challenge. He insists that, if an attribution of tacit knowledge is an empirical claim that goes beyond a summary of conforming behaviour, then it should be possible to indicate what kinds of evidence would count in favour of or against that empirical claim.⁴⁷

⁴⁶ In fact, conformity to rules is neither necessary nor sufficient for tacit knowledge of those rules. It is not necessary, since the presence of tacit knowledge of rules does not guarantee perfect deployment of that knowledge in actual performance. It is not sufficient, since a tacit knowledge claim is not offered as a summary description of behaviour but as a putative explanation of behaviour.

⁴⁷ He also insists that this evidence should involve the subject's behaviour. But it is reasonable to reply that there can be no a priori limit on the kinds of evidence that might be relevant to an empirical claim (Fodor, 1981, p. 199). So it is not legitimate to restrict evidence to the behaviour of the very subject to whom the attribution of tacit knowledge is being made.

4.2 *An account of tacit knowledge*

Quine's point about evidence is a fair one. But more fundamental than the question of what evidence would support an attribution of tacit knowledge is the question what the correctness of such an attribution consists in.

We can sketch one answer to this fundamental question about tacit knowledge by using an example that involves very simple letter-sound rules or correspondences of the kind that could be employed in reading words aloud. Suppose that one of these rules states that if a letter-string begins with the letter 'b' then its pronunciation begins with the sound /B/. It may be said that if a subject's pronunciation behaviour conforms to this rule then it displays a pattern. Whenever a presented letter-string begins with 'b', the subject's pronunciation begins with /B/. But, in line with our response to Searle's worry about trivialisation, the transitions that concern us are not these transitions from presentation to pronunciation but rather transitions amongst beliefs, representations or states of information. Whenever the subject starts out with the information that the presented letter-string begins with 'b', the subject ends up with the information that the letter-string's pronunciation begins with /B/.

If these states of information were beliefs, then the subject's pattern of transitions from belief state to belief state would be accounted for if the subject possessed verbalisable knowledge of the 'b'-/B/ rule. Possession of this piece of knowledge would provide a common causal explanation of a family of transitions from belief to belief. In contrast, there would be no such common causal explanation if the subject had merely memorised the pronunciation of each of a large number of letter-strings beginning with 'b'. The difference between having knowledge of the rule and having an independent piece of knowledge for each of the instances that fall under the rule corresponds to a difference in causal-explanatory structure. In fact, we do not assume that the subject has beliefs about letter-strings and their pronunciations; the transitions may involve states of the kind that figure in information-processing psychology. Nor do we assume that the subject either has verbalisable knowledge of pronunciation rules or else has explicitly memorised the pronunciations of a host of letter-strings. But we do make use of the notion of causal-explanatory structure and, in particular, of the idea that a battery of transitions amongst states with representational content may have a common causal explanation.

An attribution of tacit knowledge of the 'b'-/B/ rule can be construed as the claim that there is a single state of the subject that figures in a common causal explanation of a battery of transitions that conform to the rule.⁴⁸ These are the transitions that the subject makes between representations of letter-strings beginning with 'b' and representations of their pronunciations as beginning with /B/. Since these letter-strings also contain letters other than 'b', it is not correct to say that the presence of the state of tacit knowledge of the 'b'-/B/ rule is sufficient to explain the transitions from representations of complete strings to representations of their pronunciations. Rather, tacit knowledge of that rule is a common factor in the explanations of those transitions. It accounts for that aspect of the transitions

⁴⁸ Evans, 1981. Davies, 1981a,b, 1987, 1989, 1995a and Peacocke, 1986, 1989 develop Evans's proposal in slightly different ways.

that is described by the rule. In general, tacit knowledge of several different rules will be implicated in the explanation of the generation of a pronunciation for any given letter-string.

Even when we restrict attention to the aspect of the transitions that is described by the tacitly known rule, the presence of the state of tacit knowledge is not, strictly speaking, sufficient to account for the transitions. For an explanation that appeals to states of tacit knowledge presumes a background of enabling conditions. This is another respect in which a state of tacit knowledge is just one factor in the explanation of representation-to-representation transitions. But if we say that a state of tacit knowledge is a state that figures as a common factor in the causal explanations of a battery of transitions, it is important that we should not underdescribe the causal-explanatory role of this state. A state does not count as a state of tacit knowledge of the 'b'-/B/ rule simply in virtue of figuring somewhere in the explanations of all the transitions that are described by the rule.⁴⁹

Once an account of tacit knowledge in terms of causal-explanatory structure has been given, it is a relatively straightforward matter to meet Quine's challenge. Where different sets of rules or principles impose the same conditions for conformity, the attribution to a subject of tacit knowledge of a particular theory is made correct by the presence of a particular structure in the causal explanations of a battery of transitions. So we can indicate what kinds of evidence would count for or against the attribution to a subject of tacit knowledge of a particular set of rules or principles such as a grammar. What is needed is evidence of causal-explanatory structure. Indeed, some of this evidence meets Quine's additional requirement of concerning the behaviour of the subject to whom the attribution is being made. Relevant behavioural evidence could come from experimental studies of language acquisition, language perception, and language dysfunction following brain damage. Further evidence, not behavioural but still concerning the subject of the attribution, would be available from neural imaging.

4.3 Tacit knowledge and formal theories

The suggested account of tacit knowledge has an important consequence that concerns formal theories of a cognitive task. In order to explain this consequence, we need to say something about formal theories.

In order to specify a theory formally, we have to say what the axioms are and what the rules of inference are. Perhaps the most familiar cases are those in which the rules of inference are purely logical while the axioms are non-logical or *proper*. But in general, some of the logical rules of inference can be replaced by logical axioms provided that other rules of inference remain.⁵⁰ So, for example, the background logic of a theory might be formulated with the rule of *modus ponens* and the rule of &-introduction. But, in the presence of the rule

⁴⁹ See Davies, 1987, for some refinements that are needed to rule out common factors that are merely enabling conditions such as the presence of appropriate nutrients. Here we have focused on an example that involves explanations of transitions amongst representations. But states of tacit knowledge may also figure in explanations of the generation of representations. For example, tacit knowledge of a generalisation can figure in the explanation of the generation of representations whose contents are instances of that generalisation.

⁵⁰ If there were no rules of inference at all then no theorems could be proved from the axioms.

of *modus ponens*, the rule of &-introduction could be replaced by the axiom (schema): $A \rightarrow (B \rightarrow (A\&B))$.

Equally, conditional axioms, whether logical or proper, can be replaced by rules of inference.⁵¹ In the case of logical axioms, this kind of replacement is familiar. But there is no reason why there cannot also be non-logical or proper rules of inference. So, in principle, a theory of the pronunciation task could be formulated with a proper rule of inference for each letter-sound correspondence, and a psychological theory could be formulated with a proper rule of inference such as:

From: x 's thumb has just been hit by a hammer
Infer: x is experiencing pain.⁵²

We begin by considering theories that have no non-logical or proper rules of inference and then remove that restriction.

The letter-sound correspondences that figure in the pronunciation task can be stated by the axioms of a formal theory and the consequences of those correspondences for the pronunciation of letter-strings can then be derived as theorems. Furthermore, it is possible to specify a canonical proof procedure that is to be followed in deriving those theorems from the axioms. An individual axiom, stating a specific letter-sound correspondence, can then figure as a derivational common factor in the (canonical) proofs of several theorems. For example, the axiom stating the 'b'-/B/ correspondence would figure as a derivational common factor in the proofs of all the theorems that state the pronunciations of letter-strings beginning with the letter 'b'.

We can now state the important consequence of the suggested account of tacit knowledge. If the principles stated by the axioms of a formal theory of a task are embodied as tacit knowledge in a processing system that performs the task, then the causal structure of the processing *mirrors* the derivational structure of the (canonical) proofs in the formal theory. Suppose that a single axiom functions as a derivational common factor in the proofs in the theory that concern certain items in the task domain, certain letter-strings in our example. Then a single state (embodying tacit knowledge of the principle stated by that axiom) functions as a causal common factor in the processing in the system that begins from input states representing those items (those letter-strings).⁵³ In fact, this notion of causal structure mirroring derivational structure provides a way of capturing the main idea of the present account of tacit knowledge (Davies, 1987).

⁵¹ We assume that the rule of \rightarrow -introduction or conditional proof is present as either a primitive or a derived rule of inference.

⁵² There is a complication if a *ceteris paribus* clause has to be included; but the complication is the same whether the psychological principle is formulated as an axiom or as a rule.

⁵³ In general, the salient facts about derivational structure are of the form: The derivational resources used in the canonical proofs of the key theorems (here, the pronunciation-stating theorems) for a certain set of strings, S , are jointly sufficient for the canonical proof of the key theorem for some other string, s . The salient facts about causal structure are of the form: The states implicated in the causal explanation of the generation of representations of pronunciations for the strings in S are jointly sufficient for a causal explanation of the generation of a representation of the pronunciation for s .

We have been considering theories that have no proper rules of inference. If we now remove that restriction then there are two ways of developing the account of the mirroring relation between causal structure and derivational structure. One way is to allow the mirroring relation to abstract away from the difference between axioms and rules of inference. Either an axiom or a proper rule of inference can be equally a derivational common factor and if two theories differ only in trading-off between axioms and proper rules of inference then the requirements for tacit knowledge of one are just the same as for tacit knowledge of the other.⁵⁴

The alternative way is to tighten the requirements for the mirroring relation. Where the derivational common factor is an axiom the state that is the corresponding causal common factor should be an explicit representation, and where the derivational common factor is a proper rule of inference the causal common factor should be the presence of a transition-mediating mechanism.⁵⁵

On the first approach, either an explicit representation or a transition-mediating mechanism can embody tacit knowledge of a theoretical principle, and can do so independently of whether the principle is formulated as an axiom or as a rule of inference. On the second approach, either an explicit representation or a transition-mediating mechanism can embody tacit knowledge of a theoretical principle, but this goes in step with whether the principle is formulated as an axiom or a rule. Neither approach has the consequence that tacit knowledge of a theoretical principle requires explicit representation in a language-like format. Either approach can allow, for example, that tacit knowledge might be embodied in the connections of a neural network.

We are now in a position to discharge the commitment taken on back in Section 1. We said there that the stipulation that tacit knowledge of a psychological theory requires stored sentence-like representations of a battery of psychological generalisations would not be well motivated either by the structure of the theory versus simulation debate or in its own right. We explained the point about the structure of the debate, but left for later the argument that the stipulation would not be well motivated in its own right.

The account of tacit knowledge on which Heal's 'threat of collapse' argument relies imposes no requirement of explicit representation. As on the first of the two approaches just mentioned, it abstracts away from the difference between formulations of a theory with axioms and formulations with the corresponding rules of inference. But it would be possible and probably preferable to take a step in the direction of requiring explicit representation by having the notion of tacit knowledge track the axiom-rule distinction as on the second approach.⁵⁶ However, this would not, by itself, have the consequence that seems to be needed in order to block Heal's argument, namely, that tacit knowledge of a psychological theory requires explicit representations of psychological principles. In order to secure that consequence there would have to be a ban on formulating a psychological theory with proper rules of inference. But it is not clear what the motivation for such a ban could be.

⁵⁴ This is the approach taken by Davies, 1987, p. 454.

⁵⁵ See the Informational Criterion of Peacocke, 1989, p. 116. What corresponds, in his account, to the axiom-rule distinction is the distinction between drawing on (a state of) information and using a transition-type.

⁵⁶ The reason that the second approach seems preferable is simply that it is more discriminating.

In this sub-section, we have done three things: We have shown that, where there is tacit knowledge of the axioms of a formal theory, the causal structure of processing mirrors the derivational structure of proofs. We have explained two slightly different ways of developing the account of the mirroring relation when theories include proper rules of inference. And we have discharged the obligation to argue that a stipulation to the effect that tacit knowledge of a theoretical principle requires explicit representation would not be well motivated. Though these are points of detail, each one is of some importance for the argument of this paper. But the main aim of this section as a whole has simply been to sketch a non-trivial account of tacit knowledge cast in terms of causal-explanatory structure. A state of tacit knowledge of a rule is a state that figures in a common causal explanation of a battery of representation-to-representation transitions that conform to the rule. This is the account of tacit knowledge that figures as a crucial background assumption for Heal's (1994) argument that there is really no significant difference between Goldman's subpersonal-level and empirical version of the simulation theory, on the one hand, and the tacit-theory theory, on the other.⁵⁷

5. The threat of collapse

Heal argues that if a mechanism is used to simulate the operation of other mechanisms of the same kind then, according to the account of tacit knowledge sketched in Section 4, it embodies tacit knowledge of theoretical principles about how mechanisms of that kind operate.⁵⁸ If Heal is right about this then it appears that the simulation theory, conceived as Goldman conceives it in a cognitive-scientific and mechanistic way, collapses into a version of the tacit-theory theory.⁵⁹

5.1 *The background to the argument*

The background to Heal's argument is an attempt by Davies (1994) to anticipate and respond to a version of the worry about collapse. Consider someone using mental simulation to arrive at predictions about another subject O. Suppose that the predictor already knows something about O's beliefs and desires; say, that O believes that p and desires that q . How exactly is the mental simulation method supposed to work?

One option considered by Davies (1994, p. 114) is that the simulator entertains in imagination such hypotheses about mental states as:

⁵⁷ Heal, 1994, p. 134.

⁵⁸ If the second of the two approaches in Section 4.3 is adopted, then Heal's argument requires that the principles may be formulated as proper rules of inference rather than as axioms.

⁵⁹ There is a complication here. According to the simulation theory, the cognitive machinery used for decision-making is redeployed in the service of decision-predicting. So what Heal's argument would show is that the machinery used for making decisions embodies tacit knowledge of a theory about how decision-making proceeds. This is not quite sufficient for the correctness of the theory theory as Stich and Nichols define it. They propose (1992/1995, p. 154, n. 7) that, for the theory theory to win the debate, it must be shown that 'prediction, explanation and interpretation are subserved by a tacit theory *stored somewhere other than in the practical reasoning system*'. But Stich and Nichols allow that this proposal is 'a bit odd' (ibid., p. 155) and the proposal is made against the background of the suggestion that decision-making itself might make use of a psychological theory whereas, so far as we can tell, no such suggestion would result from Heal's 'threat of collapse' argument.

I believe that p

I desire that q

and then proceeds to a conclusion about a further mental state, perhaps another belief (if what is being simulated is theoretical reasoning) or a decision or intention (if it is practical reasoning):

I believe that r

I intend to V .

These products of the simulation exercise are then used to make a prediction about O , namely, that O will believe that r or will intend to V . But, if the mechanism that is used in simulation exercises mediates transitions amongst representations of psychological states, then it is virtually bound to count as embodying tacit knowledge of a psychological theory.

In a little more detail, Goldman's idea of process-driven simulation⁶⁰ is captured by the requirement that the transitions between these thoughts in the simulator should be relevantly similar to the transitions between mental states in O , the subject being simulated. But it is consistent with this requirement that the processes in the simulator should also follow the contours of the derivations of theorems in some theory. So, the causal structure of the processing in the simulator may mirror the derivational structure of the proofs of conclusions like 'I believe that r ' or 'I intend to V ' from premises such as 'I believe that p ' and 'I desire that q ' in some theory. And since these premises and conclusions are explicitly about psychological matters, the theory in question could not fail to be a psychological theory. Thus, it seems inevitable that the cognitive machinery that is employed by the simulator will embody tacit knowledge of a psychological theory.

Davies (1994, p. 117) responded to this initial worry about collapse by suggesting that the simulation process should be conceived in a rather different way. The mental states taken on in imagination by the simulator should be states (pretend beliefs, desires, decisions and intentions) whose contents simply concern the world. The motivation for the response is this. If psychological notions do not figure in the contents of the input representations or the output representations of the simulation exercise then the machinery that is responsible for the representation-to-representation transitions in the simulator might still embody tacit knowledge of some theory. But it would be tacit knowledge of a theory about one or another aspect of the non-mental world. It would not be a tacit theory about psychological matters; and so the simulation theory would not collapse into the tacit-theory theory.

5.2 Heal's reply and a more general form of the threat of collapse

Heal replies that this response is inadequate (1994, p. 136):

⁶⁰ Goldman, 1989/1995, p. 85: 'A simulation of some target systems might be accurate even if the agent lacks [a good theory of the system]. 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.'

Unless I lose grip on the distinction between myself and others, I do start, and must start, with a representation having the content ‘So and so believes that p ’. My subsequent imagining that p (if that is what I do when I simulate) is only part of a total thought state which remains a thought about the other’s thought. And what is delivered out at the far end of my deliberations is likewise an explicit representation of the other’s future thought or action. This is the basic fact we are to explain, namely our facility in psychological prediction of others on psychological premises.

She is surely right about the fundamental point here. According to the simulation theory, mental simulation provides a method for making predictions about the mental states of other people. I begin with some information about another person, O believes that p and desires that q , and I end up making the prediction that O will also believe that r or will intend to V. In an earlier example (Section 2.2), Vincent begins with the information that Yvonne believes that $p_1 - p_n$ and that she is interested in whether q and ends up predicting that Yvonne will also believe that q . So there is no prospect of removing psychological notions from the beginning and the endpoint of a simulation exercise. But if Heal is right about this point, then it seems that we cannot avoid the conclusion that whatever cognitive mechanisms underpin mental simulation must embody tacit knowledge of a psychological theory.

In fact, Heal’s argument is completely general. If a mechanism is used to simulate the operation of other mechanisms of the same kind so as to permit predictions about them then the mechanism embodies tacit knowledge of theoretical principles about how mechanisms of that kind operate. In Heal’s own example (*ibid.*, pp. 133–5), I could use my heart as an instrument of simulation so as to generate predictions about what would happen to another person’s heart under this or that condition. Or, since that is a dangerous game, I could carry a spare heart around with me for this purpose. Either way, she says, the heart that is used for simulation turns out to embody tacit knowledge of a theory about hearts (*ibid.*, p. 134): ‘My suggestion is that I shall count as having a tacit theory of the heart in virtue of possessing a heart which I can interrogate’.

In order to explain this general argument, we can use the simpler example of gas-cylinder simulation (Section 3.1). I carry around with me a cylinder C. If I have information about the temperature of the gas in some other cylinder, D, then I adjust the temperature of C and observe the consequences in order to reach a prediction about the pressure in D.⁶¹ The initial worry about collapse is that C may turn out to embody tacit knowledge of the generalisation relating temperature and pressure.⁶² The initial response is just the same as our response to Searle’s worry about trivialisation (Section 4.1). Since C’s temperature and pressure are not themselves representations of temperature and of pressure there is no prospect of an unwanted attribution of tacit knowledge.

Heal’s reply is, in effect, that the use of this response in the present context overlooks the fact that C is being used for simulation. The simulation exercise is supposed to mediate a transition between an initial representation of D’s temperature and a representation (a

⁶¹ We assume that the cylinders are relevantly similar and that they contain the same quantity of gas.

⁶² The initial worry about collapse here is the analogue of the worry that Davies (1994) anticipated and to which he gave an inadequate response.

prediction) of D's pressure. Also, C could be used to reach predictions about the pressure in other cylinders with different temperatures. So there is a whole battery of transitions from representations of temperatures to representations of pressures that I am able to make in virtue of carrying C around with me. This is already sufficient for me to count as having 'tacit (very tacit)' knowledge of the temperature-pressure generalisation (Section 4.1), but it is not yet enough to guarantee the presence of tacit knowledge on the account sketched in Section 4.2. The crucial question, according to that account, is whether there is a single state that figures as a common factor in the explanations of all the transitions in the battery. As Heal can point out, there is a plausible candidate for such a state, namely the presence, in C, of a particular quantity of gas enclosed in a particular volume.⁶³ This does not, of course, constitute an explicit representation of the temperature-pressure generalisation in a language-like format. But the enclosed quantity of gas functions as a transition-mediating mechanism and that, Heal can say, is all that is required for tacit knowledge of the generalisation.⁶⁴

In the case of mental simulation we suppose, analogously, that Vincent carries his theoretical and practical reasoning machinery around with him. If he begins with the information that Yvonne believes, say, *A or B* and also *not-A* and that she is interested in whether *B* is true then he takes his theoretical inference mechanism off line, feeds it representations with the contents *A or B* and *not-A* as inputs, and uses the output in order to generate a prediction about Yvonne. The initial worry about collapse would be that the theoretical reasoning machinery will turn out to embody tacit knowledge of such pieces of psychological theory as that people who believe *A or B* and *not-A* typically also believe *B*. The initial response would be that an unwanted attribution of tacit knowledge can be avoided. The theoretical reasoning machinery mediates transitions amongst representations with contents that concern the world (*A or B*, *not-A*, *B*) rather than amongst representations with contents that concern what other people believe.

Heal's reply to this inadequate response is that mental simulation is here being used to mediate a transition between a premise about what Yvonne believes and a conclusion about what else she believes. It could also be used to generate predictions about the beliefs of other people. So there is a battery of transitions that Vincent is able to make concerning people who believe both a disjunction and the negation of the first disjunct. The crucial question for an attribution of tacit knowledge of the principle that people who believe a disjunction and believe the negation of the first disjunct typically also believe the second disjunct is whether those transitions have a common causal explanation. Heal can reasonably say that, given a cognitive-scientific and mechanistic view of the mind, it is plausible that those transitions do have a common causal explanation. For, given such a view, it is reasonable to suppose that the mind's theoretical reasoning machinery contains a component that is responsible for performing disjunctive syllogism inferences. But then the presence of this very component in Vincent is plausibly a common factor in the causal explanations of the transitions that he makes concerning people, like Yvonne, who believe the premises of disjunctive syllogisms.

⁶³ What she says about her own example of the heart is (1994, p. 134): 'So there is an element in me playing a causal role analogous to the logical role of each statement of such a theory, namely the actual feature of the heart which does the mediating.'

⁶⁴ On the second approach in Section 4.3, the generalisation would have to be formulated as a rule of inference.

So the cognitive machinery that Vincent uses in simulation exercises turns out to embody tacit knowledge of principles about how thinking typically proceeds.⁶⁵

5.3 Response to the threat of collapse

According to the theory theory, as we conceive it, a prediction about what a subject, O, will do in circumstances C draws on a stored body of psychological information that is logically adequate to yield an answer to the question what O will (all else equal) do in circumstances C. According to the simulation theory, there is a minimal theoretical background for folk psychological prediction, but the methodology of mental simulation allows us to avoid the need for detailed antecedent knowledge about how psychological processes typically operate. If Heal's general argument is correct then a mechanism that is used for simulation embodies tacit knowledge of a theory about the typical operation of mechanisms of that kind. So, in particular, a mental simulation mechanism embodies tacit knowledge of just the kind that the theory theory says is drawn on and just the kind that the simulation theory says is not needed.

In response to Heal's general argument, we first consider the example of gas-cylinder simulation. We maintain that gas cylinder C, even when it is being used for simulation, does not embody tacit knowledge of the temperature-pressure generalisation. The reason is that the putative state of tacit knowledge, namely the presence, in C, of a particular quantity of gas enclosed in a particular volume, does not play the right causal-explanatory role.

In the simulation exercise, I begin with a representation of the temperature of cylinder D: call this S1. As a causal result of S1, the temperature of cylinder C is adjusted to a particular level: call this S2. As a causal result of this, the pressure in C also changes: S3. Finally, I generate a prediction about D: S4. A state of tacit knowledge of the temperature-pressure generalisation would be a state whose presence is sufficient, given enabling conditions, to provide a common causal explanation of a battery of transitions between representations of temperature and representations of pressure, including the transition from S1 to S4.⁶⁶ But the

⁶⁵ Heal argues that if a mechanism is used to simulate the operation of other mechanisms of the same kind then the mechanism embodies tacit knowledge of theoretical principles about how mechanisms of that kind operate. In the case of mental simulation, mechanistically conceived, we suppose that a theory about how minds operate includes the principle that people who believe a disjunction and believe the negation of the first disjunct typically also believe the second disjunct (*ceteris paribus*). If this theory cuts nature at its joints then the mind's theoretical reasoning machinery contains a disjunctive syllogism component. In the first instance, this component mediates transitions between the premises and conclusions of a thinker's own disjunctive syllogism inferences. But the presence of this component is also a common factor in the causal explanations of a simulator's representation-to-representation transitions that concern the inferences of other people. So, argues Heal, the presence of the disjunctive syllogism component itself constitutes tacit knowledge of the principle about disjunctive syllogism inferences. In the same way, for example, a component of a particular heart (a valve, say) is argued to embody tacit knowledge of an empirical principle about the contribution that such a component makes to the typical operation of hearts.

⁶⁶ There is a complication when the tacitly known generalisation includes a *ceteris paribus* clause. If the transition-mediating mechanism has no way to represent whether all else is equal, then we have to regard it as embodying, not only tacit knowledge of the *ceteris paribus* generalisation, but also a tacit assumption (which might be false) that all else is indeed equal.

presence of the enclosed quantity of gas in C does not explain the transition from S1 to S4; it only explains the transition from S2 to S3.

At least two additional factors must figure in an explanation of the transition from S1 to S4. One factor is my awareness of what is happening in C. If, for example, I am not aware of the pressure in C then I shall not be able to make any prediction about D. The other factor is my acceptance (or at least my not calling into question) that temperature and pressure are related in D in the same way as they are related in C. If I were to think that D is different from C in this respect then my prediction would be adjusted accordingly.⁶⁷

The overall system comprising me plus cylinder C produces representations of pressure as output given representations of temperature as input. So that system embodies knowledge, tacit or ordinary, of some principles or other about the relationship between temperature and pressure in the cylinders about which predictions are generated. This is the fundamental point that we must concede to Heal. But it is uncontroversial that the system embodies (in particular, I embody) knowledge, or at least an assumption, that temperature and pressure are related in other cylinders (including D) in the same way as they are related in C. This is already enough to meet the fundamental point. And my knowledge of this principle (or my assumption), together with my awareness of the non-representational states of C, is sufficient to explain the transitions, including the transition from S1 to S4.

We reject the claim that the system, and in particular cylinder C, embodies tacit knowledge of the temperature-pressure generalisation. The presence of the enclosed quantity of gas in C explains the transition from S2 to S3. But S2 and S3 are not representations, so there are no grounds for an attribution of tacit knowledge here. There is a transition between representations to be explained. But it is explained in terms of my acceptance of a similarity principle relating C and D. Furthermore, we can note that if C (or the whole system comprising me plus C) really were to embody tacit knowledge of the temperature-pressure generalisation then there would be no explanatory work for the similarity assumption to do.

If this is the right thing to say about gas-cylinder simulation, then there is something wrong with Heal's general argument. We can also respond specifically to the argument as it applies to the example of Vincent's simulation of Yvonne's reasoning. The claim to be rejected here is that Vincent's theoretical reasoning machinery, which he takes off line and uses to simulate Yvonne's disjunctive syllogism inference, embodies tacit knowledge of the psychological principle that people typically proceed from the premises of a disjunctive syllogism to its conclusion.

Vincent begins with a thought about what Yvonne believes: call this T1. He then engages in his own thinking about the world, beginning from the thoughts (not necessarily the beliefs)

⁶⁷ It may be said that a state of tacit knowledge is never strictly sufficient for what it is supposed to explain. An explanation of the production of a representation that appeals to states of tacit knowledge always presumes a background of enabling conditions, such as the presence of nutrients or an adequate flow of blood. But this correct point will not support the attribution to the gas cylinder of tacit knowledge of the temperature-pressure generalisation unless the additional factors that we have mentioned are plausibly classified as enabling conditions. Perhaps we should allow that my awareness of the condition of C is an enabling condition for simulation-based prediction as the presence of light is an enabling condition for vision. But my acceptance that temperature and pressure are related in D in the same way as they are related in C is not so plausibly regarded in this way.

A or B and *not-A* and ending with the thought *B*: call this the transition from T2 to T3. Finally he makes a prediction about what else Yvonne will believe: call this thought T4. There is a transition here from one thought about what Yvonne believes to another. So it would be folly to try to avoid altogether the attribution to Vincent of tacit knowledge (or ordinary knowledge or acceptance) of psychological principles about what people believe. Heal is absolutely right about this fundamental point. But we disagree with Heal about what attribution of tacit knowledge is licensed by the account in Section 4.

Heal's proposal is that the presence of the component of Vincent's theoretical reasoning machinery that is responsible for performing disjunctive syllogism inferences constitutes the presence in Vincent of tacit knowledge of a psychological principle about how reasoning typically proceeds. But we say that the presence of that component does not play the right causal-explanatory role. It is not sufficient to explain the transitions that have to be explained by tacit knowledge of such a principle, including the transition from T1 to T4. The presence of the disjunctive syllogism component in Vincent only explains the transition from T2 to T3.⁶⁸ Even given the presence of that component, the transition from T1 to T4 would not occur unless Vincent accepted one of a range of other principles that connect his own thinking with the thinking of other people including Yvonne.⁶⁹

Exactly what other principle is needed depends on what Vincent learns as a result of his own transition in thought from T2 to T3. If Vincent learns that when he reasons from *A or B* and *not-A* as premises he draws the conclusion *B*, then something like (Me–You) is needed. If he learns that imagining believing *A or B* and *not-A* leads him imagine believing *B*, then something like (Sim–You) is needed. If he learns that *A or B* and *not-A* jointly entail *B*, so that *B* is the thing to think if one already thinks *A or B* and *not-A*, then something like (Norm–You) is needed. We accept Heal's fundamental point that Vincent must be credited with tacit or ordinary knowledge or acceptance of some theoretical principles in which psychological notions occur. But the principles that we have just mentioned are those which, it has already been agreed, belong in the minimal theoretical background for mental simulation. So there is no threat that the simulation theory will collapse into the tacit-theory theory.

⁶⁸ Since T2 and T3 (unlike S2 and S3) are representations, it is plausible that the presence of the disjunctive syllogism component does constitute tacit knowledge of some principle. But since T2 and T3 are thoughts about the non-mental world, this will not be a psychological principle. The presence of the disjunctive syllogism component is most naturally described as constituting Vincent's tacit knowledge of the rule of disjunctive syllogism. This is quite different from tacit knowledge of the psychological principle that people typically reason in accordance with the rule of disjunctive syllogism.

⁶⁹ If the disjunctive syllogism component (or the whole system made up of Vincent's mental machinery) really were to embody tacit knowledge of the psychological principle about how reasoning typically proceeds then there would be no explanatory work to be done by Vincent's acceptance of a principle linking his own thinking with Yvonne's. But it might be said that acceptance of a linking principle would still have an explanatory role if Vincent's disjunctive syllogism component embodied tacit knowledge of a psychological theory which was only about Vincent's own disjunctive syllogism inferences. However, such an attribution of tacit knowledge would have to begin from a description of Vincent's thoughts about the world (*A or B*, *not-A*, *B*) as being both representations of states of the world and also representations of Vincent's thinking those very thoughts. In our view, such a description would not be well motivated.

5.4 A different interpretation of the examples

Our strategy for responding to Heal's argument depends on distinguishing between states of a piece of machinery that is used for simulation (S2 and S3; T2 and T3) and states that represent the condition of the thing being simulated (S1 and S4; T1 and T4). In the case of gas-cylinder simulation, for example, there is causal distance between S1, which is my representation of the temperature of D, and S2, which is the state of C having that same temperature. And there is causal distance between S3, which is the state of C having a particular pressure, and S4, which is my predictive representation of D having that same pressure.

It would be much more difficult to resist the threat of collapse if it were allowed that the temperature and pressure of C themselves constitute representations of D's temperature and pressure. For in that case, whatever mediates between S2 and S3 ipso facto mediates between a representation of temperature and a representation of pressure. It does what a state of tacit knowledge (of a proper rule of inference) is supposed to do. Similarly in the case of mental simulation, the threat of collapse would be more serious if it began from the idea that Vincent's thoughts within the scope of his simulation are simultaneously both thoughts about the world and thoughts about what Yvonne believes.

Suppose that someone presses this 'double labelling' interpretation of the examples of simulation.⁷⁰ It is not easy to reject the claim that the transition-mediating mechanisms embody tacit knowledge once we allow the re-labelling of states of temperatures and pressure as representations of temperature and pressure. But our view is that the re-labelling should not be allowed. If I use cylinder C to simulate cylinder D then it may be said that I treat the temperature and pressure of C *as if* they were representations of the temperature and pressure of D. But 'as if' representation is not representation, and 'as if' tacit knowledge is not tacit knowledge.⁷¹

If the pressure in C were really to be a representation of the pressure in D, this could only be in virtue of my readiness to make a prediction about the pressure in D on the basis of my awareness of the pressure in C.⁷² So the re-labelling of the pressure in C as a representation would depend on my acceptance of the principle that temperature and pressure are related in D in the same way as they are related in C. But, given that I accept that principle, the labelling of the pressure in C as a representation has no work to do in explaining my prediction about D.⁷³

⁷⁰ See again Heal, 1994, p. 136: 'My subsequent imagining that *p* (if that is what I do when I simulate) is only part of a total thought state which remains a thought about the other's thought.' See also Heal, 2000, p. 3, n. 5.

⁷¹ It is important that we do not accept the view that the subpersonal-level notion of representational content used in information-processing psychology is at best 'as if' representation. Cf. Searle, 1990.

⁷² No sane theory of representation will have the consequence that the pressure in C is automatically a representation of the pressure in other gas cylinders.

⁷³ Cf. Peacocke, 1984, p. xxiii.

5.5 *What if the argument had worked?*

We have argued that Heal has not succeeded in showing that a cognitive-scientific version of the simulation theory collapses into the tacit-theory theory. But what would have been the overall dialectical situation if her argument had worked?

At the beginning of her paper, Heal says this (1994, p. 132):

I want to try to make plausible the idea that the threat of collapse is induced not so much by ways of specifying what it is to simulate as by the conception of the question as being empirical and about sub-personal mechanisms.

Her aim is precisely not to show that every version of the simulation theory inevitably collapses into the theory theory but rather to use the threat of collapse as a consideration in support of a different way of setting up the whole debate. So we need to consider how her own personal-level and a priori version of the simulation theory is supposed to be immune against the threat of collapse.

Heal herself offers two answers to this question. The first begins from her conception of the opposition between the theory theory and the simulation theory. The theory theory says that a theory of thoughts about a given subject matter is quite separate from a theory about that subject matter, while the simulation theory says that ‘the capacity to think about thoughts is (and must be) an extension of the ability to think about their objects’.⁷⁴ In order to demonstrate a collapse of the simulation theory into the theory theory, we would suppose that the simulation theory is correct and then show that its correctness also constitutes the correctness of the theory theory. But, says Heal (*ibid.*, p. 142): ‘there is no route, never mind what definition we have of tacit knowledge, from the premise that one capacity is an extension of another to the conclusion that they are separate’.

Suppose, then, that we assume that Heal’s version of the simulation theory (co-cognition theory) is correct. The capacity to think about thoughts about X is an extension of the capacity to think about X. For all that Heal has said so far, her argument about tacit knowledge of a psychological theory still does through, so that possession of the capacity to think about X constitutes possession of tacit knowledge of a theory about how thinking about X typically proceeds. So suppose that this is so. This does not amount to the two capacities being quite separate. But the capacity to think about X now has two different descriptions. The first or basic description is relevant when we consider just thinking about X, but the second description, as possession of tacit knowledge, is relevant when we consider thinking about thinking about X.

It appears then that, for all that Heal has said so far, if her version of the simulation theory is correct then thinking about thinking involves tacit knowledge of a psychological theory that is not implicated in the explanation of thinking about the world. This is surely very close to saying that if the simulation theory is correct then so is the theory theory. If the claim that thinking about thinking involves tacit knowledge of a psychological theory does not add up to the theory theory as Heal conceives it, then this is because she requires that the possession of tacit knowledge (of a theory about thinking about X) and the capacity

⁷⁴ Heal, 1994, pp. 137–8. For further development, see especially Heal, 1998a.

to think about the world (about X) should not have the same basis. But if insisting on that requirement is an adequate way of avoiding collapse here then it needs to be explained why a similar requirement would not be a satisfactory way of responding to the threat of collapse that the cognitive-scientific simulation theory is supposed to face.⁷⁵

Heal's second answer to the question why her version of the simulation theory is supposed to be immune against the threat of collapse involves withdrawing one of the background assumptions of the 'threat of collapse' argument. This is the assumption that there is 'a humanly knowable theory of thinking capable of delivering . . . specific predictions' (ibid.). If this assumption is not correct then, as she says, 'the argument lapses' (ibid.). In that case, there is no threat of collapse to be faced by her version of the simulation theory. But Goldman's cognitive-scientific version is not threatened either.

In a brief sub-section at the end of an already lengthy paper, we cannot do justice to the question whether Heal's and Goldman's versions of the simulation theory are differentially open to a threat of collapse. But provisionally, our view is that if Goldman's version were to be threatened then Heal's could not escape.

Conclusion

Our aim has been to defend the legitimacy of Goldman's conception of the theory versus simulation debate as a debate between empirical theories pitched at a subpersonal level of description. We have been concerned with the major problem that is posed for this cognitive-scientific conception of the debate by Heal's 'threat of collapse' argument. If she is right, then the cognitive-scientific simulation theory is nothing other than a version of the tacit-theory theory.

In Section 2, we set the difference of approach between Goldman and Heal within the more general framework of a two-by-two array of variations on the simulation theory. In Section 3, we identified the kinds of principles that belong in the minimal theoretical background for mental simulation. In Section 4, we explained the notion of tacit knowledge that Heal assumes. In Section 5, we began by acknowledging that Davies's earlier (1994) attempt to anticipate and respond to a version of the worry about collapse was inadequate. But we rejected Heal's general argument for the claim that if a mechanism is used to simulate the operation of other mechanisms of the same kind then it embodies tacit knowledge of theoretical principles about how mechanisms of that kind operate. Heal is correct to maintain that prediction by mental simulation draws on psychological principles. But the principles do not go beyond those that have already been recognised as belonging in the minimal theoretical background for mental simulation. So Heal's argument does not establish that Goldman's version of the simulation theory collapses into the theory theory. We ended by suggesting that it is not clear how, if Heal's argument had worked, her own version of the simulation theory would have avoided the same threat of collapse.

⁷⁵ See Stich and Nichols, 1992/1995, p. 154, n. 7; Heal, 1994, p. 136.

References

- Blackburn, S. 1992: Theory, observation and drama. *Mind and Language*, 7, 187–203. Reprinted in Davies and Stone (eds), 1995a.
- Chomsky, N. 1965: *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Currie, G. 1995: Visual imagery and the simulation of vision. *Mind and Language*, 10, 25–44
- Currie, G. and Ravenscroft, I. 1997: Mental simulation and motor imagery', *Philosophy of Science*, 64, 161–80.
- Davies, M. 1981a: *Meaning, Quantification, Necessity: Themes in Philosophical Logic*. London: Routledge and Kegan Paul.
- Davies, M. 1981b: Meaning, structure, and understanding. *Synthese*, 48, 135–61.
- Davies, M. 1987: Tacit knowledge and semantic theory: Can a five per cent difference matter? *Mind*, 96, 441–62.
- Davies, M. 1989: Tacit knowledge and subdoxastic states. In A. George (ed.), *Reflections on Chomsky*. Oxford: Blackwell, 131–52. Reprinted in C. Macdonald and G. Macdonald (eds), *Philosophy of Psychology: Debates on Psychological Explanation*. Oxford: Blackwell Publishers, 1995.
- Davies, M. 1994: The mental simulation debate. In C. Peacocke (ed.), *Objectivity, Simulation and the Unity of Consciousness: Current Issues in the Philosophy of Mind* (Proceedings of the British Academy vol. 83). Oxford: Oxford University Press, 99–127. Reprinted in W.G. Lycan (ed.), *Mind and Cognition: An Anthology*. Second Edition, Oxford: Blackwell Publishers, 1998.
- Davies, M. 1995a: Two notions of implicit rules. In J.E. Tomberlin (ed.), *Philosophical Perspectives, 9: AI, Connectionism, and Philosophical Psychology*. Atascadero, CA: Ridgeview Publishing Company, 153–83.
- Davies, M. 1995b: Consciousness and the varieties of aboutness. In C. Macdonald and G. Macdonald (eds.), *Philosophy of Psychology: Debates on Psychological Explanation*. Oxford: Blackwell Publishers, 356–92.
- Davies, M. 2000: Interaction without reduction: The relationship between personal and sub-personal levels of description. *Mind and Society*, 1(2), 87–105.
- Davies, M. and Stone, T. (eds) 1995a: *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell Publishers.
- Davies, M. and Stone, T. (eds) 1995b: *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell Publishers.
- Davies, M. and Stone, T. 1998: Folk psychology and mental simulation. In A. O'Hear (ed.), *Contemporary Issues in Philosophy of Mind*. Cambridge: Cambridge University Press, 53–82.
- Dennett, D.C. 1981: Making sense of ourselves', *Philosophical Topics*, 12, 63–81. Reprinted in *The Intentional Stance*. Cambridge, MA: MIT Press, 1987.
- Evans, G. 1981: Semantic theory and tacit knowledge. In S. Holtzman and C. Leich (eds), *Wittgenstein: To Follow a Rule*. London: Routledge and Kegan Paul, 118–37. Reprinted in *Collected Papers*. Oxford: Oxford University Press, 1985.

- Fodor, 1981: Some notes on what linguistics is about. In N. Block (ed.), *Readings in Philosophy of Psychology, Volume 2*. London: Methuen, 197–207.
- Fodor, J. 1987: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- Fodor, J. 1992: A theory of the child's theory of mind. *Cognition*, 44, 283–96. Reprinted in Davies and Stone (eds), 1995b.
- Gallese, V. and Goldman, A.I. 1998: Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493–501.
- Goldman, A.I. 1989: Interpretation psychologized. *Mind and Language*, 4, 161–85. Reprinted in Davies and Stone (eds), 1995a.
- Goldman, A.I. 1992: In defense of the simulation theory. *Mind and Language*, 7, 104–19. Reprinted in Davies and Stone (eds), 1995a.
- Goldman, A.I. 1995: Empathy, mind, and morals. In Davies and Stone (eds), 1995b, 185–208.
- Goldman, A.I. 2000: The mentalizing folk. In D. Sperber (ed.), *Metarepresentations*. Oxford: Oxford University Press, 000–00.
- Gopnik, A. and Wellman, H.M. 1992: Why the child's theory of mind really *is* a theory. *Mind and Language*, 7, 145–71. Reprinted in Davies and Stone (eds), 1995a.
- Gordon, R.M. 1986: Folk psychology as simulation. *Mind and Language*, 1, 158–71. Reprinted in Davies and Stone (eds), 1995a.
- Gordon, R.M. 1992a: The simulation theory: Objections and misconceptions. *Mind and Language*, 7, 11–34. Reprinted in Davies and Stone (eds), 1995a.
- Gordon, R.M. 1992b: Reply to Stich and Nichols. *Mind and Language*, 7, 174–84. Reprinted in Davies and Stone (eds), 1995a.
- Gordon, R.M. 1995: Simulation without introspection or inference from me to you. In Davies and Stone (eds), 1995b, 53–67.
- Gordon, R.M. 1996: 'Radical' simulationism. In P. Carruthers and P.K. Smith (eds), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 11–21.
- Heal, J. 1986: Replication and functionalism. In J. Butterfield (ed.), *Language, Mind and Logic*. Cambridge: Cambridge University Press, 135–50. Reprinted in Davies and Stone (eds), 1995a.
- Heal, J. 1994: Simulation vs. theory theory: What is at issue? In C. Peacocke (ed.), *Objectivity, Simulation and the Unity of Consciousness: Current Issues in the Philosophy of Mind* (Proceedings of the British Academy vol. 83). Oxford: Oxford University Press, 129–44.
- Heal, J. 1996: Simulation and cognitive penetrability. *Mind and Language*, 11, 44–67.
- Heal, J. 1998a: Co-cognition and off-line simulation: Two ways of understanding the simulation approach. *Mind and Language*, 14, 477–98.
- Heal, J. 1998b: Understanding other minds from the inside. In A. O'Hear (ed.), *Contemporary Issues in Philosophy of Mind*. Cambridge: Cambridge University Press, 83–99.
- Heal, J. 2000: Other minds, rationality and analogy. *Proceedings of the Aristotelian Society, Supplementary Volume 74*, 1–19.

- Jackson, F.C. 1999: All that can be at issue in the theory-theory simulation debate. *Philosophical Papers*, 28, 77–96.
- Kahnemann, D. and Tversky, A. 1982: The simulation heuristic. In D. Kahnemann, P. Slovic and A. Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 201–8.
- McDowell, J. 1985: Functionalism and anomalous monism. In E. LePore and B.P. McLaughlin (eds), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 387–98.
- Meltzoff, A.N. 1995: Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838–50.
- Meltzoff, A.N. and Moore, M.K. 1983: Newborn infants imitate adult facial gestures. *Child Development*, 54, 702–9.
- Meltzoff, A.N. and Moore, M.K. 1995: Infants' understanding of people and things: From body imitation to folk psychology. In J. Bermudez, A. Marcel, and N. Eilan (eds), *The Body and the Self*. Cambridge, MA: MIT Press, 43–69.
- Nichols, S., Stich, S., Leslie, A. and Klein, D. 1996: Varieties of off-line simulation. In P. Carruthers and P.K. Smith (eds), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 39–74.
- Nichols, S. and Stich, S. 1998: Rethinking co-cognition: A reply to Heal. *Mind and Language*, 13, 499–512.
- Peacocke, C. 1986: Explanation in computational psychology: Language, perception and level 1.5. *Mind and Language*, 1, 101–23.
- Peacocke, C. 1989: When is a grammar psychologically real? In A. George (ed.), *Reflections on Chomsky*. Oxford: Basil Blackwell, 111–30.
- Peacocke, C. 1994: Introduction: The issues and their further development. In C. Peacocke (ed.), *Objectivity, Simulation and the Unity of Consciousness: Current Issues in the Philosophy of Mind* (Proceedings of the British Academy vol. 83). Oxford: Oxford University Press, xi–xxvi.
- Quine, W.V.O. 1972: Methodological reflections on current linguistic theory. In D. Davidson and G. Harman (eds), *Semantics of Natural Language*. Dordrecht: Reidel, 442–54.
- Ravenscroft, I. 1998: What is it like to be someone else?: Simulation and empathy. *Ratio*, 11, 170–85.
- Searle, J. 1990: Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences*, 13, 585–96.
- Stich, S. and Nichols, S. 1992: Folk psychology: Simulation or tacit theory? *Mind and Language*, 7, 35–71. Reprinted in Davies and Stone (eds), 1995a.
- Stich, S. and Nichols, S. 1995: Second thoughts on simulation. In Davies and Stone (eds), 1995b, 87–108.
- Stich, S. and Nichols, S. 1996: How do minds understand minds? Mental simulation versus tacit theory. In S.P. Stich, *Deconstructing the Mind*. Oxford: Oxford University Press, 136–67.
- Stich, S. and Nichols, S. 1997: Cognitive penetrability, rationality and restricted simulation. *Mind and Language*, 12, 297–326.

- Stone, T. and Davies, M. 1996: The mental simulation debate: A progress report. In P. Carruthers and P.K. Smith (eds), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 119-37.
- Stone, T. and Davies, M. 1999: Autonomous psychology and the moderate neuron doctrine. *Behavioral and Brain Sciences*, 22, 849-50.
- Walton, 1997: Spelunking, simulation and slime: On being moved by fiction. In M. Hjort and S. Laver (eds), *Emotion and the Arts*. Oxford: Oxford University Press, 37-49.