Folk Psychology and Mental Simulation

MARTIN DAVIES AND TONY STONE

This paper is about the contemporary debate concerning folk psychology – the debate between the proponents of the *theory theory* of folk psychology and the friends of the *simulation alternative*.[1] At the outset, we need to ask: What should we mean by this term 'folk psychology'?

Shall we perhaps say that folk psychology is just what the folk know (or believe) about psychological matters? The problem with this putative definition is that, if folk psychology is a body of known or believed propositions about psychology, then it may be said that folk psychology is a psychological theory. This would threaten to render invisible even the possibility of an alternative to the theory theory of folk psychology.

Someone might respond to this problem by saying that not just any collection of propositions about psychology deserves to be called a *theory*. Only a set of propositions organized around generalizations that support counterfactuals and are appropriately

Some of the material in this paper was presented at the Pacific Division meeting of the American Philosophical Association and at the University of Michigan, as well as at seminars in Canberra, Melbourne, Oxford, Paris and Sydney. We are grateful to many friends and colleagues, including Ned Block, Greg Currie, Allan Gibbard, Robert Gordon, Paul Harris, Jane Heal, Frank Jackson, Janet Levin, Christopher Peacocke, Philip Pettit, Huw Price, Peter Railton, Ian Ravenscroft, Michael Smith, Dan Sperber, Stephen Stich and Kendall Walton, for comments and conversations. MD is pleased to acknowledge financial support from the Australian National University and the Humanities Research Board of the British Academy and is especially grateful to members of the Philosophy Department at the University of Michigan for the opportunity to visit as the James B. and Grace J. Nelson Philosopher in Residence.

[1] Much of the relevant literature is gathered in three collections: *Folk Psychology: The Theory of Mind Debate*, ed. M. Davies and T. Stone (Oxford: Blackwell Publishers, 1995); *Mental Simulation: Evaluations and Applications*, ed. M. Davies and T. Stone (Oxford: Blackwell Publishers, 1995); and *Theories of Theories of Mind*, ed. P. Carruthers and P. K. Smith (Cambridge: Cambridge University Press, 1996).

objective will earn that title.[2] So, folk psychology will be a theory only if what the folk know or believe about psychology has something of the character of a science. This response has some plausibility. There is surely something to be said for this restrictive use of the term 'theory', and it will be important in Section III of this paper, when we consider explanation and understanding. But many of the participants in the debate between the theory theory and the simulation alternative have used the term 'theory' in an extremely inclusive way. For example, Stephen Stich and Shaun Nichols adopt a 'wide interpretation' of the term on which 'just about any internally stored body of information about a given domain [counts] as an internally represented theory of that domain'.[3] Our initial aim is to describe the debate – or at least one aspect of the debate – in a way that takes account of the use of the term 'theory' to include any body of knowledge, belief or information.

Instead of beginning with folk psychology as what the folk know or believe about psychology, we do better to start with folk psychological *practice* – a practice in which we all engage on an everyday basis. We *describe* people as bearers of psychological states. We *explain* people's behaviour (or decisions, or judgements or other psychological states) by appeal to their psychological states. We *predict* people's behaviour (or decisions, or judgements or other psychological states) by relying on assumptions about their psychological states. The debate between the theory theory and the simulation alternative can then be seen as a debate about this three-stranded practice.[4]

[2] T. Nagel, *The View from Nowhere* (Oxford: Oxford University Press, 1986), p. 5: 'A view or form of thought is more objective than another if it relies less on the specifics of the individual's makeup and position in the world, or on the character of the particular type of creature he is.'

[3] S. Stich and S. Nichols, 'Folk Psychology: Simulation or Tacit Theory?' in *Folk Psychology*, ed. Davies and Stone, p. 133. See also S. Stich and S. Nichols, 'Second Thoughts on Simulation', in *Mental Simulation*, ed. Davies and Stone; S. Nichols, S. Stich, A. Leslie and D. Klein, 'Varieties of Off-Line Simulation', in *Theories of Theories of Mind*, ed. Carruthers and Smith.

[4] The debate (particularly in its early stages) seems to have been conducted under two assumptions. One is that there is a single question to be asked about folk psychology. The other is that the theory theory and the simulation alternative offer the only two viable approaches to answering that question. But both of these assumptions are flawed. As against the first assumption, we would say that there are many different, and fairly independent, questions to be asked about folk psychological practice, each one of which might be given a theory theory or a simulation theory style of answer. (See T. Stone and M. Davies, 'The Mental

Amongst the many questions that can be asked about folk psychological practice, one question that has been central in much of the recent literature is *the basis question*: What is the basis of our ability to engage in folk psychological practice?[5] Indeed, a great deal of attention has been focused on the basis question applied to just the prediction strand of folk psychological practice. The greater part of this paper shares this relatively narrow focus (sections I and II). Only in the final section do we move to consider explanation and understanding.

I Prediction, Theory and Simulation

What would be the theory theory's account of folk psychological prediction, and what alternative account would the simulation theory offer? We approach the question indirectly by considering first a case of prediction in a straightforwardly physical domain. How could someone predict the change in pressure of the gas in a cylinder when its temperature is raised?

Prediction in a physical domain

One possibility would be to use an empirical generalization about the way in which the pressure of a volume of gas increases as its

[5] Elsewhere (Stone and Davies, 'The Mental Simulation Debate: A Progress Report', p. 120), we have put the question this way: 'What resources do mature adult humans draw upon as they go about the business of attributing mental states, and predicting and explaining one another's mental states and actions?' We called it the *explanatory question about normal adult folk psychological practice*. We have now opted to call it the basis question lest the term 'explanatory' suggest that the question relates only to the explanation strand of folk psychological practice.

Simulation Debate: A Progress Report', in *Theories of Theories of Mind*, ed. Carruthers and Smith, pp. 119–20, for nine such questions. No doubt there are more.) As against the second assumption, we would make two brief points. One point is that we cannot simply assume that the two terms, 'theory theory' and 'simulation theory', even when quite generously construed, cover the whole space of possible answers to the questions that are at issue. The other point is that, even for a single question, and even when the theory theory and the simulation alternative are the only approaches in view, the correct answer might be a hybrid, drawing on both approaches.

temperature increases.[6] In this case, the predictor would be drawing on a body of information about gases, in line with a theory theory account. Another possibility would be to draw on a theory about the movement and energy of gas molecules. By considering the forces exerted on the walls of the cylinder, the predictor might arrive at a prediction of increased pressure without actually having antecedent knowledge of the temperature-pressure law. Or again, the predictor might not draw on any knowledge about gases in general, but simply make use of a formula relating the temperature and pressure of the gas in this particular cylinder. Given the inclusive notion of theory that is in play, this would count as use of a theory.

There is, of course, an alternative to these theory-based strategies for arriving at a prediction about the pressure of the gas in a cylinder, A, after its temperature is raised. We could take another similar cylinder of gas, B, heat it to the temperature in question, and measure the pressure. Provided that the cylinder B really is relevantly similar to cylinder A, this method is liable to yield an accurate prediction. By using the behaviour of the second cylinder of gas as a *simulation* of the behaviour of the first cylinder, we can make a prediction about cylinder A in the absence of any antecedent empirical knowledge about changes in the behaviour of gases under increases in temperature.

In order to use simulation in this way to predict the pressure in gas cylinder A, we need to use another real gas cylinder and we need to raise its temperature in reality. This *simulation in reality* provides for prediction in the absence of antecedent empirical knowledge about the behaviour of gases. A predictor who did not have a second cylinder to hand could, of course, *imagine* having a second gas cylinder. Or a predictor who was armed with a second gas cylinder but did not want to heat it could imagine its temperature being raised. But in order for either of these imaginative strategies to yield a prediction about the pressure in cylinder A, the predictor would need to develop the imagined gas cylinder narrative beyond its starting point ('There is a cylinder of gas. It is heated up. And then...'); and to do this, the predictor would need to draw on some theory – some body of information – about the behaviour of gases.[7]

[6] The general principle is that pressure is proportional to (absolute) temperature and inversely proportional to volume. In the present context, the volume is constant. If, instead, the temperature is regarded as constant then the resulting principle, that volume is inversely proportional to pressure, is known as Boyle's law.

[7] This kind of prediction by simulation in imagination is closely connected with the use of thought experiments in science. Thought experiments are often important in the development of theory, and so it may

As a strategy for predicting the pressure in cylinder A, *simulation in imagination* must deploy essentially the same resources as those that are used according to the theory theory account. So, in this case at least, simulation in imagination is *theory-driven* simulation.[8] It is only simulation in reality that constitutes a genuine alternative to the use of empirical theory in prediction.

Psychological prediction

In the folk psychological case, it is clear enough how knowledge of an empirical theory about psychological matters can yield predictions. The body of theory drawn on might consist of some relatively superficial generalizations about (personal level) psychological properties (cf. the laws relating temperature, pressure and volume of gases) or postulates about (subpersonal level) information processing apparatus (cf. postulates about the movement and energy of gas molecules); or it might be information that is specifically about a particular individual (e.g. someone whom the predictor knows well; cf. a formula linking temperature and pressure in cylinder A).

It is also clear that, in the psychological case, simulation in reality can be an effective way of generating predictions without relying on knowledge of empirical theory. Suppose that I want to predict (i) how a person C will feel (or how soon C will fall over) after drinking a pint of whisky, or (ii) how the Müller-Lyer illusion will look to C, or (iii) how C will feel and what he will decide to do if he is suspended over a cliff on a rope and he cannot find a foothold and his hands are starting to slip, or (iv) whether C will draw the conclusion that something is white from his belief that snow is white.[9]

[8] See A. I. Goldman, 'Interpretation Psychologized', in *Folk Psychology*, ed. Davies and Stone, p. 85, for the distinction between theory-driven and process-driven simulation.

[9] The whisky example is discussed by Jane Heal, 'How to Think About Thinking', in *Mental Simulation*, ed. Davies and Stone, p. 48, and by Richard Moran, 'Interpretation Theory and the First Person',

seem implausible to say that simulation in imagination draws on theory. We need to note, once again, that an inclusive notion of theory is in play, and that in some cases the propositions drawn on will simply be intuitive assumptions about what kinds of thing do, or do not, tend to happen in the physical world. See, for example, R. Sorenson, *Thought Experiments* (Oxford: Oxford University Press, 1992), pp. 52-4, for an account of Stevinus's use (in 1605) of a thought experiment to determine the force needed to keep a ball from moving down an inclined plane. One of the assumptions at work in this case was that perpetual motion does not happen.

In each case, I can use the strategy of placing another person, D, into the same situation and observing D's reactions. This may well yield a correct prediction about C, provided that C and D are relevantly similar ((i) in the way that alcohol affects their bodily constitution; (ii) in the way that their visual systems work; (iii) in the way that they experience and act on emotions; (iv) in the way that they reason). To the extent that I, myself, am relevantly similar to C, I have an option that is not available to me in the case of gas cylinder simulation in reality; namely, I can place myself into those situations and observe my own reactions. I drink a pint of whisky, or look at the two lines, or dangle perilously over a cliff, or draw out some simple inferences from my belief that snow is white. Indeed, in discussions of mental simulation in reality, it is usually this option of using oneself in a simulation that is considered.[10]

But it is *mental simulation in imagination* that is central for the simulation theory. We saw that gas cylinder simulation in imagination needs to be driven by empirical theory. Does the same go for

[10] S. Stich and S. Nichols, 'Cognitive Penetrability, Rationality and Restricted Simulation', *Mind and Language*, 12 (1997), p. 302, call this 'actual-situation-simulation'. It is important to avoid a possible confusion here. In some important examples, a protagonist has a false belief about her situation: there is a difference between the situation as it actually is and the situation as the protagonist takes it to be. A subject who is asked to predict what the protagonist will think or do may make an incorrect prediction by focusing on the situation as it actually is rather than the situation as the protagonist takes it to be. (This is what very young children tend to do. There is a substantial empirical literature on the *false belief task*. See, for example, H. Wimmer and J. Perner, 'Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception', *Cognition,* 13 (1983), pp. 103–28.) But this predictive strategy is not what Stich and Nichols mean by 'actual-situation-simulation' (and not what we mean by 'simulation in reality'). Rather, actual-situation-simulation would involve placing myself into the same situation as the protagonist and making myself (perhaps *per impossibile*) subject to the same false belief.

Philosophical Quarterly, 44 (1994), p. 163. The Müller-Lyer illusion is discussed by Robert Gordon, 'Reply to Stich and Nichols', in *Folk Psychology*, ed. Davies and Stone, pp. 175–6. The example of emotional response to a story is discussed by Kendall Walton, 'Spelunking, Simulation and Slime: On Being Moved by Fiction', in *Emotion and the Arts*, ed. M. Hjort and S. Laver (Oxford: Oxford University Press, 1997), and Ian Ravenscroft, 'What Is It Like to be Someone Else?: Simulation and Empathy', *Ratio*, 11 (1998). The case of inference is central in Heal's discussions. We take the example from Allan Gibbard, 'Brains, Thoughts, and Norms', unpublished manuscript.

mental simulation? It seems clear that if, with a view to making a prediction about C, I imagine placing another person D into the same situation, then I shall need to draw on theory in order to develop the simulation beyond this starting point. But if what I imagine is actually being in the situation,[11] then simulation in imagination might allow a prediction about C to be generated. What this prospect seems to depend on is the possibility that my imagining being in a situation engages the same psychological or mental processes in me as would be operative if I were really in that situation.

Consider, then, the conditions under which simulation in imagination would yield correct predictions in the four sample cases that we have mentioned. (i) If simulation in imagination is to yield a correct prediction about how C will feel after drinking a pint of whisky, then imagining drinking a pint of whisky must produce in me feelings of giddiness leading to a fall – or at least imagined feelings of giddiness leading to an imagined fall. (ii) In the case of the Müller-Lyer illusion, imagining the two lines and the arrowheads must lead to a visual experience – real or imagined – as of one line being longer than the other. (iii) When I imagine being suspended over a cliff on a rope, this act of imagination must lead to real or imagined feelings of fear and panic. (iv) When I imagine believing that snow is white (or, more to the point, when I imagine believing that, say, butter is white – something that I do not, in reality, believe), this must lead to the real or imagined act of judging that something is white.

We take it that the facts about these cases are roughly as follows. (i) Imagining drinking a pint of whisky does not, in and of itself, produce real or imagined feelings of giddiness. The bodily processes that lead up to a feeling of giddiness are not engaged by the imagined consumption of alcohol in the same way that they would be engaged by the real consumption of alcohol. If my simulation in imagination does move forward from the drinking to the feelings, then this is because I am bringing to bear some empirical knowledge about how people typically feel – or about how I usually feel – after consuming large quantities of alcohol.

(ii) Imagining the lines and the arrowheads does not, in and of itself, generate the Müller-Lyer illusion in imagination. The visual processes that give rise to the illusion are not engaged by the imagined confrontation with that array of lines and arrowheads in the same way that they would be by the real presentation of the array. (iii) On the other hand, imagining being in that dangerous situation,

[11] B. A. O. Williams, 'Imagination and the Self', in *Problems of the Self* (Cambridge: Cambridge University Press, 1973).

dangling at the end of a rope, may well lead to real feelings of fear or panic, without my drawing on any empirical theory about how people in that kind of situation typically feel. Imagined danger may engage a range of bodily and emotional processes in somewhat the same way that real danger does.

(iv) Finally, imagining believing the premises of an argument (that butter is white) certainly can lead me to an imagined judgement of the conclusion (that something is white), without my using any antecedently known empirical theory about how people typically reason. There is an important contrast between the case of reasoning from imagined beliefs and the case of emotional response to imagined danger. The bodily symptoms of fear or panic may well be real, even though the danger is only imagined. But in the case of reasoning, if my commitment to the premises is only an imagined commitment, then my judgement of the conclusion is similarly imagined rather than real. The process leading from one to the other is, however, real, and not merely imaginary, reasoning; and that real reasoning may also prompt a real judgement, namely, the conditional judgement that if the premises were true then so would be the conclusion.

What all this suggests is that the prospects for psychological prediction by simulation in imagination, without the use of empirical theory, are not utterly forlorn. It may also seem to suggest that we need to set about the task of cataloguing which psychological processes are engaged in the same way by imagined inputs as by real inputs. But while real interest and importance would attach to that cataloguing project, it is also important to note that it is not just a brute fact that imagining premises engages our reasoning abilities in the same way that really believing the premises does.[12] Rather, the explanation of this fact is that reason relations (such as entailment relations) obtain, and are known by any thinker to obtain, amongst imagined or hypothesized thought contents, in just the same way that they obtain amongst believed thought contents. When I simulate C's reasoning in imagination, a theory may well be used. But it is not an empirical theory about how people happen typically to reason. Rather, it is a normative theory about right reasoning; and it is the very same theory that I can use when I engage in reasoning from premises that I actually believe.[13]

Although the simulation of reasoning may involve deployment of normative principles, the simulation theory is not (even when

[12] This point is stressed, for example, by Goldman, 'Interpretation Psychologized', p. 85, and by Heal, 'How to Think About Thinking', pp. 34–5.

[13] See Stone and Davies, 'The Mental Simulation Debate: A Progress Report', pp. 136–7.

restricted to reasoning) to be equated with what might be called the *normative theory theory*. It is possible to know normative principles relating to an activity in which one does not, oneself, engage. But the simulation theory is clearly not proposing that we make predictions by the disengaged use of a set of normative principles about reasoning.[14] Rather, normative principles may be used in simulation because they are already available to us when we ourselves engage in reasoning. When we use those normative principles, our reasoning becomes what Tyler Burge describes as critical reasoning.

> Critical reasoning is reasoning that involves an ability to recognise and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such. As a critical reasoner, one not only reasons. One recognises reasons as reasons...
>
> Essential to carrying out critical reasoning is using one's knowledge of what constitutes good reasons to guide one's actual first-order reasoning.[15]

Not all reasoning is critical reasoning. But it is arguable that the possibility of critical reasoning is an essential part of normal adult reasoning as we know it.[16]

The point we have reached is that predicting the conclusions of another person's (theoretical or practical) reasoning appears to be a particularly favourable case for a simulation theory answer to the basis question about the prediction strand of folk psychological practice. Of course, in order to reach a correct prediction about C's conclusions by simulating his reasoning in imagination, I need to take account of differences between C and myself. If C believes that butter is white, while I do not, then C may arrive at the judgement that butter and snow are the same colour, given that snow is white,

[14] See S. Blackburn, 'Theory, Observation and Drama', in *Folk Psychology*, ed. Davies and Stone, p. 283: 'Theorizing under a normative umbrella is still *theorizing*. It could, it seems, be done quite externally, in the light of a sufficient stock of principles telling what it would be right or wrong to think or feel in some situation...'. Janet Levin, 'Folk Psychology and the Simulationist Challenge', *Acta Analytica*, 14 (1995), p. 91, also makes the point that if we use a normative theory to predict what inferences a person will make then this does not yet seem to involve anything that is 'in any serious sense *a simulation*'.

[15] T. Burge, 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society*, 96 (1996), pp. 98–9.

[16] Burge, 'Our Entitlement to Self-Knowledge', p. 99: 'A non-critical reasoner reasons blind, without appreciating reasons as reasons. Animals and children reason in this way.'

whereas I would not myself draw that conclusion. But I can take account of this difference between C and me within the simulation, without needing to draw on any empirical information about how people who believe that butter is white tend to reason. Rather, in imagination I take on the belief that butter is white and then, given the premises that snow is white and that butter is white, I conclude that butter and snow are the same colour. That is what right reasoning requires.

Predicting how someone will feel after drinking a pint of whisky, in contrast, is a good case for a theory theory answer to the basis question. Consequently, predicting the conclusions that will be reached by someone reasoning after drinking a pint of whisky also depends on at least some contribution from empirical theory. If C has just drunk a pint of whisky and I have not, then I need to take account of this difference between him and me when I try to simulate his reasoning in imagination. Even if I correctly take on C's premises in imagination *and* imagine drinking a pint of whisky, still my predictions about C's conclusions are liable to be incorrect, unless I bring to bear some empirical information about how whisky affects (C's) reasoning. Here, correct prediction requires an intrusion of theory. But this is not to say that, in the case of the inebriated C, my prediction strategy must owe everything to empirical theory and nothing to mental simulation. The empirical information that I draw on might take the form of information about the ways in which someone in C's condition is liable to depart from right reasoning. In that case, I could first use my own reasoning ability to work out what would be a correct conclusion to draw from C's premises and then tweak my prediction in the light of that empirical information.

The epistemology of prediction by simulation

Let us now consider, in a little more detail, how prediction by simulation would work. We have already noted that, in the case of the gas cylinders, prediction by simulation in reality relies on some assumption of relevant similarity between cylinder B and cylinder A. One form that this assumption can take is that cylinder B is a typical member of a class, G, of gas cylinders of which A is also a member. Heating the gas in cylinder B and measuring its pressure can then be conceived as an *experiment*, licensing a general claim about temperature and pressure in gas cylinders in the class G. Since cylinder A is assumed to be a member of this class, the experimentally licensed generalization can be applied to it. Essentially the same kind of account could be given, in the psy-

chological case, of the role of the assumption of relevant similarity between person D and person C. And if, in a case of mental simulation in reality, I use myself instead of another person D, then an assumption of relevant similarity between me and C plays the same role again. Placing myself in the situation can be conceived as an experiment.

It would seem plausible, then, that there is no very deep difference between the epistemological status of predictions based on simulation and predictions that rely on experimentally licensed knowledge of empirical generalizations. Furthermore, it would appear that, in the case of mental simulation in imagination, much the same account would be given, but with an extra empirical assumption to the effect that the processes in me that are engaged by imagined inputs work in the same way as the processes in me, and in C, that are engaged by real inputs. The cataloguing project mentioned on p. 60 above would then be seen as the project of assessing the extent to which that empirical assumption is warranted.

However, the account that we have sketched of simulation of reasoning in imagination may open the possibility of a distinctive epistemology of psychological prediction. What the normative theory of right reasoning tells the simulator is that the conclusion – say, that something is white – is the right thing to think, given the premise – say, that snow is white, or that butter is white. This normative judgement about what is the thing to think does not, by itself, yield a prediction about C, of course. The simulator also needs an assumption that C will think the thing that is the thing to think. That is a defeasible assumption in any given case. But it may enjoy a default status, nevertheless, since unexplained departures from these normative requirements of reasoning call in question our attributions to C of thoughts with such contents as that snow is white or that butter is white.[17] This route to prediction goes via a normative judgement (this is the thing to think in such-and-such a situation) and an assumption about interpretation (C will think the thing that is the thing to think). It is to be contrasted with a route

[17] The general idea here is familiar from discussions of the principles involved in radical interpretation. Some advocates of mental simulation contrast the simulation approach with the rationality approach, and so would not adopt the account of the epistemology of psychological prediction that is sketched here. See, for example, Goldman, 'Interpretation Psychologized'. On the other hand, R. M. Gordon, 'Simulation Without Introspection or Inference from Me to You', in *Mental Simulation*, ed. Davies and Stone, can be seen as resisting the idea that the epistemology of prediction by simulation is the same as that of prediction by way of empirical theory.

that goes via an empirical judgement (this is what I think when placed experimentally in such-and-such a situation) and an assumption of similarity (C is relevantly like me).

II Prediction Failure

We have distinguished between simulation in reality and simulation in imagination as methods of prediction. Simulation in reality can certainly be an effective way of generating predictions without relying on empirical knowledge. But the prospects for prediction by simulation in imagination depend on the possibility that imagining being in a situation should engage the same psychological or mental processes as would be operative if one were really in that situation. We considered a range of examples and concluded that predicting the results of another person's reasoning is a good case for simulation in imagination while predicting how someone will feel after drinking a pint of whisky is not. But while it might be agreed that predicting the conclusions of reasoning could be achieved by mental simulation, this does not settle the question whether prediction is in fact achieved in that way. Perhaps, despite the availability of simulation, we ordinarily make such predictions by relying on an empirical theory about how people reason.

The basis question with which we began is an empirical question about our three-stranded folk psychological practice, and we have been focusing on the question as it applies to the prediction strand. But we have so far said nothing about the kinds of empirical evidence that would count in favour of one or another answer to the basis question. In a series of important papers, Stich and Nichols have urged that the phenomenon of *prediction failure* is strong evidence in support of a theory theory answer to the question about the basis of our prediction practice.[18]

In our everyday folk psychological practice, we sometimes make wrong predictions. Stich and Nichols argue that this happens because our prediction method is *cognitively penetrable* – that is, our psychological predictions are influenced by our antecedent knowledge or beliefs about the psychological domain. This kind of explanation of prediction failure is available to the theory theorist but unavailable, Stich and Nichols say, to the friend of mental simulation. So the existence of prediction failure is a crucial test of the empirical adequacy of the two competing accounts of the causal

[18] 'Folk Psychology: Simulation or Tacit Theory?', 'Second Thoughts on Simulation', and 'Varieties of Off-Line Simulation'. We note again that Stich and Nichols use the term 'theory' in an extremely inclusive sense.

basis of our prediction practice, and favours the theory theory account. Thus, on the one hand:

> One virtue of using a simulation to predict the behavior of a system is that you need have no serious idea about the principles governing the target system. You just run the simulation and watch what happens ... In predictions based on simulation, what you don't know won't hurt you ... If there is some quirk of the human decision-making system, something quite unknown to most people that leads the system to behave in an unexpected way in certain circumstances, the accuracy of prediction based on simulations should not be adversely affected. If you provide the simulation with the right pretend input, it should simulate (and thus predict) the unexpected output.[19]

But, on the other hand:

> Just the opposite is true for predictions that rely on a theory. If we are making predictions on the basis of a set of laws or principles, and if there are some unexpected aspects of the system's behavior that are not captured by our principles, then our predictions about those aspects of the system's behavior should be less accurate. Theory based predictions are sensitive to what we know and don't know about the laws that govern the system; they are cognitively penetrable.[20]

The dialectical situation that Stich and Nichols sketch is especially clear when we contrast theory-based prediction and prediction by simulation in reality. Thus, consider again our prediction of the pressure in gas cylinder A. If someone has a false theory about the behaviour of gases, then a theory-based prediction about cylinder A is liable to be false. But, if the predictor uses the behaviour of cylinder B as a simulation of the behaviour of cylinder A, then the prediction arrived at should be correct. Because the prediction method does not draw on any antecedently believed empirical theory about the behaviour of gases, the prediction can, in principle, be insulated from any false theoretical beliefs that are antecedently held by the predictor.[21] If someone makes an incorrect prediction about the pressure of the

[19] 'Folk Psychology: Simulation or Tacit Theory?' p. 150.

[20] Ibid.

[21] In section I, we noted the similarity between gas cylinder simulation in reality and the use of experiments to establish generalizations about how gas cylinders in a certain class generally behave. The present point, that simulation in reality yields predictions that are insulated from antecedently held theory, is analogous to the point that experiments are apt to yield results that conflict with antecedently held theory.

gas in cylinder A after it has been heated then either the predictor is not using simulation as the prediction method or else the simulation is flawed in one of two ways. It may be that cylinder B is not relevantly similar to cylinder A or it may be that the gas in cylinder B was not heated to the correct temperature.[22]

In the psychological case, just the same points can be made. If, in order to arrive at a prediction about C, I use D (or myself) for a simulation in reality, then the prediction should be correct. If it is incorrect then either D is not relevantly similar to C (or I am not similar to C), or else D (or I) was not placed into the correct situation (that is, the simulation was not provided with the correct inputs). But the central case of mental simulation is simulation in imagination. Is the dialectical situation the same here? There is some reason to allow that it is. Someone who claims that mental simulation provides even a possible account of folk psychological prediction relies on the idea that imagining being in a situation may engage the same psychological or mental processes as would be operative if one were really in that situation. For some examples, such as the situation in which one drinks a pint of whisky, the idea has no plausibility. But the advocate of mental simulation has to maintain that there are other cases where the idea is plausible, and we have suggested that these would include cases of theoretical and practical reasoning. So, it appears that prediction failure relating to reasoning would present a problem for anyone offering a mental simulation answer to the basis question about folk psychological prediction. Certainly, this is what Stich and Nichols have argued; and they have gone on to present examples of this kind of prediction failure.

Examples of prediction failure

There is no shortage of surprising experimental psychological data about conclusions that people draw and decisions that they take. The very fact that we find the data surprising indicates, of course, that we ourselves would have made incorrect predictions about what the subjects in the experiments would conclude or what they would decide. We shall describe two of these examples.[23]

[22] Someone using simulation in reality as a prediction method may, of course, refuse to accept the result of a simulation if it conflicts with an antecedently held theory, and may judge that the simulation must be flawed in some way. The same goes for experimentation.

[23] These two examples are discussed by Stich and Nichols, 'Folk Psychology: Simulation or Tacit Theory?' p. 151, along with the example of belief perseverance; see R. Nisbet and L. Ross, *Human Inference* (Englewood Cliffs, NJ: Prentice Hall, 1980), pp. 175-9. In 'Second

Position effects: right bias in selecting goods

Shoppers are presented with a display of what are, in fact, identical samples of some product. They are asked to assess the quality of these samples and then – by way of payment for participating in the survey – to select one sample to keep. The result is that the shoppers' selections show a bias towards samples near the right-hand end of the display over samples near the left-hand end.[24]

Most people are surprised to hear the result of this experiment; they would predict that shoppers' selections would be random. If these predictions are arrived at by mental simulation, then simulation is generating incorrect predictions. Yet it is reasonable to assume that the people who are asked to predict the outcome of the experiment are relevantly similar to the subjects in the experiment (the shoppers).

The Langer effect

Two groups of subjects are sold lottery tickets for $1 each. Subjects in one group are allowed to choose their lottery ticket (choice condition); subjects in the other group are simply given a ticket (no-choice condition). Subjects are then (under some pretext or other) asked to be ready to sell their ticket back to the experimenter, and are asked to set a sell-back price. The result is that subjects in the choice condition set very much higher prices on average than subjects in the no-choice condition (over $8 versus just under $2).[25]

Most people are surprised to hear the result of this experiment. For example, Stich and Nichols report anecdotal evidence of presenting undergraduate students with a description of the experiment and asking them to predict what the subjects would do. The students failed to predict the difference between the sell-back prices set by subjects in the choice condition and subjects in the no-choice condition. If these predictions are arrived at by mental simulation – the students simulating first being in one condition and then in the other – then simulation is generating incorrect predictions. Yet it is reasonable to assume that the students who are

[24] Nisbet and Ross, *Human Inference*, p. 207.

[25] E. Langer, 'The Illusion of Control', *Journal of Personality and Social Psychology*, 32 (1975), pp. 311–28. The example is discussed at length in Nichols et al., 'Varieties of Off-Line Simulation'.

Thoughts on Simulation', pp. 101–2, they introduce the further example of failure to predict how subjects will behave in Milgram's obedience experiment. S. Nichols, S. Stich and A. Leslie, 'Choice Effects and the Ineffectiveness of Simulation', *Mind and Language,* 10 (1995), pp. 442–4, also discuss an example of subjects' failure to predict how they themselves will behave when asked to put a price on an article that they own.

asked to predict the outcome of the experiment are relevantly similar to the original subjects.

Response on behalf of the simulation theory

Given the way that the argument about prediction failure has been set up, it will appear that the defender of mental simulation is bound to make a move analogous to saying that gas cylinder B was not heated to the correct temperature. That is, the defender of simulation must say that, in these cases of prediction failure, the (pretended) inputs to the predictor's simulation in imagination are in crucial respects different from the inputs that engaged the psychological processes of the subjects in the real experiments. This is, indeed, the way in which advocates of mental simulation have responded.

Thus, for example, Robert Gordon comments on the example of right bias in selecting goods that, 'unlike the subjects in the original experiment, the subject in the imagination experiment [the person trying to predict how shoppers will behave] must be told that the items on display are identical (and thus of equal quality)'.[26] In a similar vein, Paul Harris notes that a person trying to predict the outcome of the Langer experiment using simulation:

> needs to simulate the vacillation and eventual commitment of the free-choice subjects. Moreover, in making that simulation they must also set aside the tacit reminder embedded in a narrative that juxtaposes the two groups of subjects, namely that any lottery ticket whether selected or allocated, has the same likelihood of winning. Subjects in the experiment who were offered a free choice had no knowledge of the other group, and by implication, no such tacit reminder.[27]

This is a good initial move to make on behalf of the simulation theory. Someone who is aiming to make a prediction by simulation in imagination must imagine being in the very situation that the subjects in the original experiment were in. And this must be done in such a way as to offer the simulator's psychological processes inputs that are equivalent to the inputs that engaged the original subjects' psychological processes. In a case of simulation of reasoning, the simulator must take on in imagination the very same premises that were available to the original subject. But, as Gordon and Harris point out, the way in which the experimental situation is described may prevent this condition from being met.

[26] Gordon, 'Reply to Stich and Nichols', p. 176.

[27] P.L. Harris, 'From Simulation to Folk Psychology: The Case for Development', in *Folk Psychology*, ed. Davies and Stone, p. 218.

There is a quite general point here; namely, that the way in which the situation to be imagined is described can make a huge difference to the prospects for successful simulation. Consider the case of a lexical decision experiment. Letter strings are flashed up on a computer screen – some strings form real words, and some form (pronounceable) non-words. The subject has to decide whether each letter string is a word or a non-word and press one or another button to indicate this decision. Suppose that I am asked to predict what decisions subjects will make. Simulation in reality is no problem here: I can just sit in front of the screen myself. But if I have to simulate this experimental regime in imagination, then some ways of describing the input make my task nearly impossible. I might, for example, be given a description of the screen display in terms of the pattern of light and dark pixels that form the image of the letter string. If, on the other hand, the screen display is described by the letters being named in order, then I may very well be successful in simulating the performance of subjects in the experiment and thus predicting their responses. This successful prediction would not seem to depend on antecedent knowledge about how normal subjects respond to this or that letter string in a lexical decision experiment. Rather, I would just make what I take to be the correct decision about each imagined letter string, and then assume that other subjects would make the correct decision too. In doing this, I would make use of stored information; but it would be information about lexical items, not information about normal subjects' lexical decisions.

In our view, this line of response (in terms of *wrong inputs*) enables the simulation theorist to fend off the initial versions of the objection from prediction failure. But it does not resolve the debate in favour of either side because the response simply invites the theory theorist to improve the design of the prediction experiment so as to rule out the wrong inputs response. Thus, for example, Nichols, Stich, Leslie and Klein report a prediction experiment in which subjects in one group watch a videotape of a subject in the choice condition of a Langer-style experiment and are asked to predict the subject's sell-back price, while subjects in another group similarly watch a videotape of a subject in the no-choice condition.[28] As in the original Langer experiment, subjects in the choice condition set significantly higher sell-back prices than subjects in the no-choice condition. But there was no significant difference between the prices predicted by subjects shown the choice condition videotape and the prices predicted by subjects shown the video of the no-choice condition. Thus, even with a videotape to help them imagine the experimental situation, subjects are not reliably able to reach correct predictions.

[28] Nichols et al., 'Varieties of Off-Line Simulation', pp. 49–52.

There is no doubt that discussion of these examples can be continued, with the defender of prediction by mental simulation in imagination deploying variations on the wrong inputs theme.[29] But there is a slightly different kind of response to these examples that is suggested by our earlier reflections on the prospects for psychological prediction by simulation in imagination (pp. 57–62 above).

The circumscribed domain of prediction by mental simulation

There are all kinds of factors that may affect a person's theoretical or practical reasoning, such as whether the person believes that butter is white, or whether the person has just drunk a pint of whisky. Some of these factors can readily be taken into account by someone attempting a prediction by mental simulation in imagination, while others cannot. Showing me a videotape of a subject drinking a pint of whisky before engaging in some reasoning will not enable me to predict the outcome of the subject's reasoning, however accurately I may imagine the subject's situation. As we noted on p. 59 above, what I need is empirical information about the effects of whisky drinking.[30] (Recall, too, that the use of this empirical information need not wholly supplant engagement in mental simulation.)

The fact that prediction by mental simulation in imagination

[29] See for example, A. Kühberger, J. Perner, M. Schulte and R. Leingruber, 'Choice or No Choice: Is the Langer Effect Evidence Against Simulation?', *Mind and Language*, 10 (1995), p. 433: '[I]t is difficult to ensure that simulator participants are provided with sufficient information about exactly the right combination of factors that produces the Langer effect.' Kühberger et al. refer to the requirement that 'the imagined situation captures the relevant features of the simulated person's actual situation' as the requirement of *imaginative adequacy* (p. 424).

A theory theorist may object that the use of the wrong inputs response by the friend of mental simulation is ad hoc and that the defender of the simulation theory in the face of examples of prediction failure should be willing to state in advance under what conditions the requirement of imaginative adequacy would be met. (See Stich and Nichols, 'Second Thoughts on Simulation', p. 102.) But it is not clear that the theory theorist's own approach to examples of prediction failure is any more principled. The theory theorist's explanation of prediction failure is in terms of the predictor's use of an incomplete or false theory about psychological matters, or the predictor's use of incorrect initial conditions to instantiate correct generalizations. But no independently motivated account of the exact nature of the predictor's failure is provided. (This point is made in an unpublished paper by Ian Ravenscroft, and also by Stich and Nichols, 'Cognitive Penetrability, Rationality and Restricted Simulation', p. 323, who credit it to Meredith and Michael Williams.)

[30] Alternatively, I could drink a pint of whisky myself, combining simulation of the subject's beliefs in imagination with simulation of the subject's

requires an intrusion of theory in such cases of 'non-rational' influences has been recognised from the beginning of the contemporary debate.[31] Furthermore, it seems quite likely that some of the factors at work in producing the Langer effect or the right bias in selecting goods may be more like whisky than like reasons. For example, most people who are told about the position effects experiment find it surprising that the shoppers' selections show a bias towards samples near the right-hand end of the display. They would predict a random distribution of selections. A plausible explanation for this prediction is that there is no evident reason to make one selection rather than another; the fact that a sample is towards the right-hand end of the display scarcely constitutes a justification for selecting that sample rather than any of the others.[32] It is not *ad hoc*, then, to maintain that these examples of prediction failure fall outside the proper domain of prediction by mental simulation unaided by empirical theory.

A narrow circumscription of this domain is explicit in Heal's work:

> The kind of simulationism I would like to defend says that the only cases that the 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

[31] J. Heal, 'Replication and Functionalism', in *Folk Psychology*, ed. Davies and Stone, p. 48; Harris, 'From Simulation to Folk Psychology: The Case for Development', p. 219.

[32] So-called non-rational influences may have their effects in a very direct way – by-passing reasoning altogether – as, perhaps, in the case of the shoppers. But they may also work by making something that is not in fact a reason for acting in a certain way nevertheless appear to be a reason. We are not committing ourselves to any specific account of the various examples of prediction failure. Indeed, we are not even committed to the idea that the examples of prediction failure all involve non-rational influences. Perhaps subjects in the Langer-style experiment have good reasons for setting their sell-back prices, but those reasons are somehow obscured from subjects in the prediction experiment. In that case, a defender of simulation will, in the end, be right to use some version of the wrong inputs response. What we are pointing out is just that there is a different kind of response – in terms of non-rational influences – that is, in principle, available to the simulation theorist. See J. Heal, 'Simulation and Cognitive Penetrability', *Mind and Language*, 11 (1996), pp. 60–6.

---

whisky drinking in reality. This might enable me to make a correct prediction, if whisky has the same effect on my reasoning from hypothesised contents as on the subject's reasoning from believed contents. However, it is important to note that the effects of my drinking the whisky will not be restricted to my simulation of the subject; my reasoning in my own person will be perturbed as well. This might be a disadvantage if I need to think carefully and accurately about how best to act towards the subject.

(c) the latter content is rationally or intelligibly linked to that of the earlier item(s).[33]

But her proposal faces an objection. In many cases of prediction failure, the subjects about whom the predictions are made seem to depart in some way from right reasoning. But, in some other cases of equally flawed reasoning, correct prediction seems to be quite straightforward. In these latter cases, why do not the non-rational influences put the reasoning beyond the range of prediction by mental simulation?

Consider an example discussed by Daniel Kahnemann and Amos Tversky.[34] At a flying school, instructors adopt a policy of responding positively to good performance (such as successful execution of complex manoeuvres) and negatively to bad performance. When reviewing this policy, they note that pilots who do particularly well and are praised are likely to perform less well next time, while pilots who perform particularly badly and are criticized are likely to do better at their next attempt. The instructors conclude that, contrary to what psychologists tell us about positive reinforcement, rewarding good performance is not an effective training method.

Most people find the flight instructors' conclusion to be quite unsurprising; it is just as they would predict. Yet the instructors' reasoning is flawed; it overlooks the phenomenon of regression towards the mean. (A pilot who has reached a certain level of competence and performs outstandingly well on one trial is likely to perform less well on the next trial, independently of the reaction of the instructor.) Is this not a problem, the objector asks, for the idea that the proper domain of prediction by mental simulation is the domain of rational linkages?

It is clearly relevant to note, here, that the reasoning of the people who successfully predict the instructors' conclusion is flawed in just the same way as the reasoning of those instructors. But that point is liable to suggest, again, that there is something wrong with the proposal to circumscribe the proper domain of prediction by mental simulation in terms of the contrast between right reasoning and non-rational influences. What matters for mental simulation, the objector may say, is not rationality but similarity.[35] Prediction by

[33] Heal, 'Simulation and Cognitive Penetrability', p. 56.

[34] D. Kahnemann and A. Tversky, 'On the Psychology of Prediction', in *Judgment Under Uncertainty: Heuristics and Biases*, ed. D. Kahnemann, P. Slovic and A. Tversky (Cambridge: Cambridge University Press, 1982), pp. 67–8. The example was used by Ned Block (in conversation) to make the objection under discussion here. Essentially the same objection against Heal's circumscribed version of simulation theory is pressed by Stich and Nichols, 'Cognitive Penetrability, Rationality and Restricted Simulation'.

[35] Dan Sperber (in conversation) pressed the objection in this form.

mental simulation will be successful just where a process that operates in imagination in the predictor is relevantly similar to the process operating in reality in the subject about whom the prediction is being made. By that account, probabilistic reasoning that overlooks regression towards the mean falls squarely within the proper domain of prediction by mental simulation, since the error is one that virtually everyone is disposed to make.

Our view is that it is possible to defend Heal's circumscription of the proper domain of mental simulation by drawing on two ideas from section I: the idea of a normative theory and the idea of a distinctive epistemology of psychological prediction. First, as critical reasoners, we are each in possession of a normative theory of right reasoning (p. 61 above). We are also subject to non-rational influences and so we are all liable, on occasion, to reason in ways that are out of line with our normative principles. However, some departures from right reasoning may actually be sanctioned by our normative principles; that is, our normative theory may itself be flawed. Second, in virtue of our possession of a normative theory, we can arrive at judgements about what is the thing to think in a certain situation; and we can use those judgements, coupled with an assumption that the subject will think the thing that is the thing to think, in order to arrive at predictions. This predictive strategy can bestow a distinctive kind of epistemic warrant (p. 63 above). When a subject departs from right reasoning in a way that is out of line with our normative theory, this strategy will yield a wrong prediction, and will need to be augmented by empirical information about the non-rational influences at work on the subject. When the subject departs from right reasoning in a way that is sanctioned by our normative theory, in contrast, this strategy will yield a correct prediction. But it will be a prediction that does not constitute knowledge, since it is based on two false claims – that this is the thing to think, and that the subject will think the thing that is the thing to think – where the error in the second claim compensates for the error in the first.

On this account, if the predictor and the subject share an incorrect normative theory then the predictions arrived at will be fortuitously, rather than knowledgeably, correct. The narrowly circumscribed domain of mental simulation is the domain of knowledgeable predictions that are arrived at by that epistemologically distinctive route.

However, we should also consider what happens if the predictor learns about the importance of not ignoring regression towards the mean. For now the predictor will, provided that he or she is properly attentive, arrive at a correct judgment about what is the thing for the flight instructors, for example, to think. But the predictor may also recognize that, in this kind of case, most people are apt not

to think the thing that is the thing to think. So, the predictor will take this empirical information into account when arriving at a prediction about the flight instructors. It may be that the informed predictor characterizes the way in which most people depart from right reasoning simply as the way in which he or she used to reason before learning about regression towards the mean. Perhaps, indeed, the predictor still finds it all too easy to slip back into that flawed pattern of reasoning. In that case, the predictor may suspend his or her recently acquired normative knowledge, and engage in a piece of not wholly critical reasoning, so as to arrive at a correct – and knowledgeably correct – prediction about the flight instructors.

This is quite properly regarded as a piece of prediction by mental simulation. But the route that it takes is via an empirical judgement (this is what I used to think – how I used to reason) and an assumption of similarity (the flight instructors are relevantly like me as I used to be). So, while the distinctive epistemology of prediction that goes with the idea of a normative theory of right reasoning is of some importance, it would be wrong for us to suppose that all knowledgeable prediction by mental simulation exhibits that distinctive epistemology.

III Simulation, Explanation and Understanding

We began with the three strands of folk psychological practice – description, explanation, and prediction – but we have been almost exclusively concerned with folk psychological prediction, and with the basis question concerning that strand of our practice. In this final section, we turn briefly to folk psychological explanation.

Explanation and generalizations

Suppose that we want to explain the increase of pressure in our gas cylinder that results from an increase in temperature. The theory theory account of prediction (pp. 55–6 above) can readily be converted into an account of explanation by subsumption.[36] The conjunction of an increase in temperature and an increase in pressure is subsumed under the temperature-pressure law. The truth of this generalization is not, however, something brute. The relatively superficial temperature-pressure law belongs, not only with a more

[36] C. Hempel, *Aspects of Scientific Explanation* (New York: The Free Press, 1965), who provides the seminal account of the deductive-nomological model of explanation, regards the distinction between prediction and explanation as being merely pragmatic.

general principle relating temperature, pressure and volume, but also with a body of empirical theory about the movement and energy of gas molecules, and about forces exerted on the walls of the cylinder. In terms of this theory, it is possible to give a mechanistic account of how it is that the relatively superficial law is true – of how the temperature-pressure connection is implemented. In short, according to the theory theory account, prediction and explanation go naturally together, and a predictor who knows not only the superficial generalization but also the broader body of theory is able to achieve an explanatory understanding of the predicted increase in pressure.

In the folk psychological case, too, the theory theory account of prediction goes along with an account of explanation. Knowledge of a body of psychological theory provides the resources for explanations that work by subsuming events under causal generalizations. There may be variations on this theme. Some theory theorists will regard knowledge of relatively superficial psychological generalizations as the visible tip of an iceberg of more elaborated tacit knowledge, while others will commit themselves only to knowledge in the ordinary sense of the term. Some theory theorists will regard cognitive scientific theories about subpersonal level information processing machinery as offering deeper explanations of psychological matters, while others will hold hard to the personal level. But the general picture is clear.

Given that familiar picture of explanation by subsumption, it may seem obvious that the basis question about the explanation strand of folk psychological practice has a ready answer in terms of the theory theory, but cannot be answered in terms of the simulation alternative. Explanation requires generalizations; but mental simulation is supposed not to depend on antecedent knowledge of psychological generalizations. However, what seems obvious is not quite correct.

Consider again the case of the gas cylinders. We have noted already (p. 62 above) the similarity between prediction by simulation in reality and the use of experiments to license generalizations. So, gas cylinder simulation, carried out without antecedent knowledge about the behaviour of gases, could yield knowledge of generalizations that could, in turn, be used in subsumptive explanations. Gas cylinder simulation in reality would naturally be called *black box simulation*; we simply give the simulation device (gas cylinder B) a temperature as input and receive back from it a pressure as output. Consistently with that description, the experimentally licensed generalizations would be superficial and, because of the lack of explanatory depth, the simulation would scarcely provide any explanatory understanding of the predicted increase in pressure. But still, the basic point remains. Simulation, conceived as experi-

ment, may yield knowledge of generalizations under which events can be subsumed. We could call this *simulation-driven theory*.

So also, in the folk psychological case, simulation can be regarded as experiment and may yield knowledge of empirical generalizations. This is particularly clear in the case of simulation in reality. By drinking pints of whisky, looking at pairs of lines, dangling on ropes and drawing inferences, I may not only arrive at predictions about another person C (pp. 57–62 above). I may also, by induction from these bouts of simulation considered as experiments, arrive at empirical generalizations under which events in the mental life of C may be subsumed. This is also true – though over a circumscribed domain – for mental simulation in imagination. When I simulate C's reasoning in imagination, I draw on a normative, rather than an empirical, theory about reasoning. But I may arrive at empirical generalizations by induction on the results of simulation in imagination; and, to the extent that mental simulation may yield correct predictions, it may also yield correct generalizations.[37]

Simulation and understanding

If explanation is conceived as subsumption under generalizations, then the debate initiated by the basis question about the explanation strand of folk psychological practice will take a course that is essentially parallel to the debate over the prediction strand. But in fact, many advocates of the simulation alternative would defend the idea that there is a distinctive – not straightforwardly subsumptive – kind of explanation involved in folk psychological understanding. Thus, for example, Heal says:

> The difference between psychological explanation and explanation in the natural sciences is that in giving a psychological explanation we render the thought or behaviour of the other intelligible, we exhibit them as having some point, some reason to be cited in their defence.[38]

[37] We are committed to the possibility that there may be both normative and empirical principles cast in very similar terms. Both kinds of principle would make use of *ceteris paribus* clauses; but those clauses would be interpreted differently in the two cases.

[38] Heal, 'Replication and Functionalism', p. 52. See also, J. McDowell, 'Functionalism and Anomalous Monism', in *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, ed. E. LePore and B. McLaughlin (Oxford: Basil Blackwell, 1985), p. 389: '[T]he concepts of the propositional attitudes have their proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be.'

This kind of normative explanation reveals what someone thought or did as having been the rational thing to think or do, or the thing that it made sense to think or do, given the circumstances and the agent's beliefs and preferences. Clearly, explanation in this style fits together with our account of prediction by mental simulation (in a circumscribed domain) in somewhat the way that explanation by subsumption is the natural companion of prediction that draws on empirical generalizations.

But we do not get an adequate view of the distinctive kind of psychological understanding that might be provided by mental simulation if we focus only on the normative aspect. For, as we have noted (p. 61 above), it is possible to deploy a normative theory about an activity in which one does not, oneself, engage. What mental simulation promises is a kind of understanding that is not only normative but also *first personal*.[39] We see the combination of these two aspects most vividly in the simulation of reasoning in imagination; and the idea that mental simulation can provide a distinctive kind of understanding of another person's reasoning is strikingly similar to R. G. Collingwood's claim that historical understanding is to be achieved by the re-enactment of the historical character's thought:

> But how does the historian discern the thoughts which he is trying to discover? There is only one way in which it can be done: by rethinking them in his own mind. The historian of philosophy, reading Plato, is trying to know what Plato thought, when he expressed himself in certain words. The only way in which he can do this is by thinking it for himself. This, in fact, is what we mean when we speak of 'understanding' the words. So the historian of politics or warfare, presented with an account of certain actions done by Julius Caesar, tries to understand these actions, that is, to discover what thoughts in Caesar's mind determined him to do them. This implies envisaging for himself the situation in which Caesar stood, and thinking for himself what Caesar thought about the situation and the possible ways of dealing with it. The

[39] Thus, for example, Gordon, 'Simulation Without Introspection or Inference from Me to You', p. 56 quotes Kant (*Critique of Pure Reason*, A353) approvingly: 'It is obvious that, if I wish to represent to myself a thinking being, I must put myself in his place, and thus substitute, as it were, my own subject for the object I am seeking to consider (which does not occur in any other kind of investigation).' For an illuminating discussion of issues not far removed from those of the present section, see Moran, 'Interpretation Theory and the First Person'.

> history of thought ... is the re-enactment of past thought in the historian's own mind.[40]

Indeed, just as the domain of prediction – and correlatively of understanding – by mental simulation may be narrowly circumscribed (pp. 71–2 above), so also understanding by re-enactment may seem to be restricted to right thinking. This would be a severe limitation on historical understanding.

Patrick Gardiner considers this objection to Collingwood in a recent paper:

> [I]t may ... be objected that the re-enactment conception of understanding remains unrealistically restrictive in the amount it seems to exclude from the historian's proper scope. However scrupulous the care taken to judge an action from the agent's own standpoint, there can be no *a priori* guarantee that the reasoning ascribable to him will turn out to have been cogent or sound; as Francis Bacon once remarked, 'it is a great mistake to suppose men too rational'. It is always conceivable in principle, and it is surely often the case in practice, that there is a lack of coincidence between the conclusions people actually draw on the basis of their beliefs and purposes and the conclusions that rationally they should have drawn. Thus in history as elsewhere people may engage in faulty practical thinking, whether because of such things as haste or unimaginativeness or as a result of underlying emotional factors that sway or distort their judgement. But when that happens – the objection may continue – it does not follow that their behaviour is unintelligible in terms of reasons, only that the reasons are liable to be poor or inadequate ones.[41]

Gardiner's response to this objection is to note that 'Collingwood would be less vulnerable to some of the criticisms brought against him on the present score if his conception [of re-thinking] were interpreted in a more flexible manner.' On such an interpretation, re-enactment of thought would cover not only right reasoning but also, for example, 'empathetically appreciating how an agent could

[40] R. G. Collingwood, 'Human Nature and Human History', in *The Idea of History* (Oxford: Oxford University Press, Revised Edition 1992), p. 215. As is quite widely remarked, the simulation approach to psychological understanding has marked affinities with the hermeneutic tradition of Vico, Herder, Dilthey, Weber and Croce, as well as Collingwood. See *Verstehen and Humane Understanding*: *Royal Institute of Philosophy Supplement 41*, ed. A. O'Hear (Cambridge: Cambridge University Press, 1997).

[41] P. Gardiner, 'Collingwood and Historical Understanding', in *Verstehen and Humane Understanding*, ed. O'Hear, pp. 117–18.

have been tempted or misled into accepting a particular practical conclusion without recognising the faultiness of the reasoning involved'.[42] Is it possible for a friend of mental simulation to expand the domain of understanding by simulation in a similar way?

In the case of prediction by simulation (p. 62 above), we saw that there could be an intrusion of empirical theory without the prediction strategy coming to owe everything to theory and nothing to mental simulation. The possibility that we mentioned there was that the empirical information drawn on might be information about how particular influences (such as drinking a pint of whisky) lead to departures from right reasoning. However, there is no guarantee that, if we modulate the re-enactment of thought in the light of empirical information, then the resulting first person narrative (in imagination) will be one that we find intelligible. Thus, for example, Simon Blackburn considers the case of deliberating about what is the thing to do if one is a subject in Milgram's obedience experiment, and then taking account of the empirical evidence about what subjects actually tend to do. The simulator can modify his or her narrative in the light of the empirical information. But, 'this need have no tendency to make the behaviour of Milgram's subjects intelligible. I might still feel quite baffled, both by them, and if I am like them, by me.'[43]

An intrusion of empirical theory may, then, bring with it a loss of intelligibility. But it would not be right to conclude that there is no prospect of a more flexible conception of the domain of understanding by simulation. Consider, for example, the predictor who now knows about regression towards the mean but who still finds it all too easy to slip back into flawed reasoning (p. 74 above). This predictor will surely not be baffled by the reasoning of the flight instructors. Their reasoning does not measure up to the informed predictor's normative theory; but their first person narrative is nevertheless one that the predictor will find intelligible.

There are other cases, too, in which it may be possible, without simply being driven by an empirical theory, to re-enact thinking that departs from right reasoning. Let us return to one of our earlier examples. I want to predict how C will feel and what he will decide to do if he is suspended over a cliff on a rope and he cannot find a foothold and his hands are starting to slip (example (iii) on p. 57 above). Seized by fear or panic, C may not think or do the best thing, the most rational thing. Yet, by simulating C's situation in imagination, I might reach a correct prediction about C without drawing on any empirical theory about how people dangling over

[42] Ibid., p. 118.

[43] Blackburn, 'Theory, Observation and Drama', p. 283.

cliffs on ropes tend to think. For imagining the situation might be enough to produce in me physiological and emotional responses that perturb my reasoning in imagination in just the way that C's reasoning in reality is perturbed.[44] I might arrive at a correct prediction about C; and if I regard the simulation exercise as an experiment I might arrive by induction at some generalizations about how people think and act in dangerous situations.[45] But there is something more. By re-enacting C's desperate thinking, struggling to maintain a grip, deciding to make another attempt to find a foothold – all in imagination, of course – I surely gain some kind of empathetic understanding of the thoughts, feelings and decisions that I predict. This is not a case of theory-driven simulation; and it is not black box simulation either. It is simulation that, in Gordon's phrase, 'essentially engages [my] own practical and emotional responses'.[46]

There is an alternative way in which I can gain a measure of first personal understanding of C's thoughts, feelings and decisions – a way that does not require actual physiological and emotional responses in me at the moment of understanding. I may take into account my own remembered similar experiences.[47] In this case, I draw on stored information – about how I felt, physically and emotionally, and about how this affected my thinking and decision taking – and I use this information to help me imagine what it is like to be C. (I may also draw on memories of imaginings in which I was fully physiologically and emotionally engaged.) Producing a correct narrative about another person is not always sufficient for finding what that person thinks and does to be intelligible. But it is plausible that in some cases we can make sense of what someone thinks and does by drawing on memory to help us imagine being in the

[44] These responses may have consequences, not only for my reasoning within the scope of my simulation of C, but also for my reasoning in my own person. Cf. footnote 30 above.

[45] If my prediction about C's thoughts and actions is to count as knowledge then it should not depend on the flawed normative judgement that this is the thing to think, or to do, in these circumstances. In this case, knowledgeable prediction seems to require some recognition of the fact that one's reasoning is indeed being perturbed.

[46] R. M. Gordon, 'The Simulation Theory: Objections and Misconceptions', in *Folk Psychology*, ed. Davies and Stone, p. 103. Since understanding is a kind of knowledge, there will once again be a need for me not to be wholly in the grip of the re-enactment (cf. footnote 45 above).

[47] Nichols et al., 'Varieties of Off-Line Simulation', pp. 59–67, discuss empathy and in particular the role of memory in empathetic emotion. What we are concerned with here, however, is remembered emotion, not emotion aroused by memory.

other person's situation – indeed, to help us imagine being that person. This is an intrusion of empirical theory, given the inclusive way in which the term 'theory' has been used. But it does not obstruct first personal understanding, and it does not move us back towards explanation by subsumption.

Conclusion

In the first two sections of this paper we were concerned with the prediction strand of folk psychological practice. The theory theory and the simulation alternative agree about what folk psychological prediction is; but they disagree about its basis. According to the theory theory, the predictor draws on a body of information about psychological matters. According to the simulation alternative, prediction is sometimes possible by simulation in imagination without the aid of empirical psychological theory. However, the domain of prediction by mental simulation – particularly if the epistemology is to be different from the epistemology of empirical theory – is rather closely circumscribed: it is the domain of reason.

When we turn to the explanation strand of folk psychological practice, we find that the contours of the debate are very different. For there is a disagreement about what folk psychological explanation is. According to the theory theory, it is explanation by subsumption under causal generalizations. So, if the basis of the explanation strand of folk psychological practice is to be knowledge of a psychological theory, then that theory must contain generalizations of the right kind – objective, counterfactual supporting – to figure in subsumptive explanations. It is a theory in a more restricted sense. According to the simulation alternative, folk psychological explanation is normative and first personal; it is a matter of finding the other person's life intelligible 'from the inside'.[48] This is an imaginative project; and understanding involves not only reasoning in imagination but also emotion and memory. What is remembered is, of course, information about psychological matters. So, if psychological understanding is to range beyond the domain of reason then, even by the lights of the simulation account, it must draw on psychological theory. But this does not constitute a victory for the theory theory, because the psychological theory on which simulation and understanding draw is theory in the inclusive sense, but not in the restricted sense that is relevant to the theory theory's account of psychological explanation.

If we do not distinguish the inclusive sense of 'theory' which is

[48] See Jane Heal's paper in this volume.

relevant to the debate about prediction from the restricted sense of 'theory' which is relevant to the debate about explanation, then we may obscure from ourselves the role of empathy and emotion in commonsense psychological understanding.