

CHAPTER 41

FOLK PSYCHOLOGY

ADAM MORTON

WE describe people in terms of their beliefs, desires, emotions, and personalities, and we attempt to explain their actions in terms of these and other such states. These are folk-psychological concepts; they are the ones we use when we think of people in terms of their minds. (Strawson 1959 called them ‘M-concepts’.) Let us leave it fairly vague what is to count as a folk-psychological concept, taking as core examples the concepts of belief and desire and concepts of emotion, and then including as folk psychological concepts any concepts whose primary function is to enter into combination with the core examples to give explanations and predictions of human action. (It is important to remember, though, that explaining and predicting action is not the only purpose of ascribing states of mind. Often, for example, the ascription is part of an inference to a conclusion about the physical world, as when we say ‘She seems to be lying, so there is probably something hidden in the basement’.) Usually when there are concepts there are beliefs or theories, whose structure allows us to understand the concepts by linking them to other concepts and to experience. By ‘folk psychology’ let us mean whatever serves this linking role for folk-psychological concepts. Much philosophical writing about concepts would suggest that the linking must be done primarily by beliefs or theories; a central question is whether this is so for folk-psychological concepts.

41.1 THEORY

The idea now labelled ‘folk psychology’ arose around the 1970s when a number of philosophers, often influenced by Sellars’s description (1956) of ‘the myth of Jones’ and by Quine’s assimilation (1953) of all of our thinking to explanatory

hypotheses, began to explore the consequences of assuming that individual people acquire from their cultures an explicit theory which postulates that individuals have beliefs, desires, and other states, which interact in specific ways to produce actions. (Typical examples are Putnam 1960, Dennett 1978, Churchland 1979.) This theory would be the common-sense precursor of psychological theories developed as part of the scientific approach to human nature.

The position had many attractive features. First, it fitted very neatly with the dominant account at the time of the nature of mental states, functionalism, according to which, roughly, mental states are any states of an organism which interact in the right ways. The right ways can then be taken as those described by folk psychology. Second, the position also provided a new take on the old epistemological problem of other minds. The traditional argument from analogy was intellectually discredited, but now it became possible to see each person's reason for attributing states of mind to others as an inference to the best explanation, in which the 'hypothesis' that another has states interacting in the way required by folk psychology emerges as the best explanation of the person's actions. And, third, it became possible to formulate a new range of philosophical questions analogous to old-fashioned scepticism about the existence of other minds but much more varied and plausible. Is folk psychology true? Is it a first approximation to an eventual scientific psychology? Might it eventually be replaced by a theory based on experimental evidence? These questions are much saner than other-minds scepticism: one can quite naturally imagine the answers to any of them going either way.

But is there such a folk-psychological theory? The overwhelming consensus among philosophers and psychologists is that there is not, if what we have in mind is an explicit theory that can be expressed in ordinary language and that is learned by children from adults through its linguistic presentation. The most basic reason is that we cannot produce the theory in sufficient detail to determine, for example, the situations in which normal folk psychologists will take a stronger motive to overrule a weaker one. To that extent what Morton (1980) called the 'theory theory' is clearly false. But there are many unrefuted theory theories still in contention: for there are many more promising ways of understanding 'theory'. Instead of debating theory/no theory, it is clearer to follow Nichols and Stich (2003) and contrast information-rich and information-poor accounts of folk psychology. Information-rich accounts postulate a source of articulated principles to guide attribution, understanding, and prediction. The principles may not be formulated in spoken language or available to conscious reflection, but they guide our more explicit judgements. They may originate in innate information-processing routines, reflection on the actions of others, or innate constraints on the explanations of others. According to information-poor accounts, on the other hand, the fundamental features of our thinking about mind are determined by routines and capacities which do not embody any assumptions about how one state of mind leads to another.

41.2 SIMULATION

To see how an information-poor account can be formulated consider 'simulation' accounts of mental-state ascription. There are two influential models, one due to Jane Heal (1995) and the other to Robert Gordon (1995). Heal's model applies best to cases in which one person has to predict what another person will do given an intellectual problem. Suppose that one person, A, is predicting which of 'some fish fish for fish' and 'some dogs dog for dog' another person, B, will judge to be a grammatical English sentence, or what B will say when asked the product of 120×13 . The natural way for A to solve the problem of predicting B's response is not to think about B's thought processes directly but simply to solve the problem herself and then to suppose that B will give the correct answer; that is, the one that A has just come up with. This procedure is information-poor in that A's thinking does not need to represent any suggestion about how B thinks. A just has to think, and then to transfer the result.

The other pioneering simulation account, Gordon's, applies best to cases in which a person is predicting another person's decision. Suppose that A is driving behind B and sees a deer run into the road in front of B's car. A can anticipate B's braking heavily and swerving to the left. A will then also brake, before seeing B's brake lights, and may decide not to swerve left in the hope of missing both B and the deer. One way in which A could form this expectation about B's behaviour is by deciding what she would do in B's place, where this means not thinking about B's decision but making it, feeding into her own decision-making processes the facts about B's situation and then taking the result not as an intention for her own action but as a prediction about B's. A crucial part of the capacity to think of others as minds, on such an account, is being able to take one's own decision processes 'off line' and use them to form expectations about other people's actions. To do this one has to be able to take the other person's point of view, at least to the extent of understanding what aspects of the situation are and are not known to the other person, and one has to be able to insulate it from the processes by which one turns decisions into actions. As Currie (1995) and others have pointed out, these capacities are also required for conditional thinking, in which one feeds into one's decision processes some hypothetical facts and then stores the outcome as a conditional decision, not to be acted on unless the facts turn out to be actual.

The contrast between simulation and theory should not be made too stark. In most cases a Gordon-type simulation will have to be guided by some information about how it is to be performed. ('The driver ahead can see the deer because it is nearer to him than it is to me, and so I can take that information into the simulation; but he may not know that I am close behind, so I can leave that information out.') In many cases, too, a simulation will also require information about the target person's

desires, which may not be the same as those of the ascriber. (Suppose the driver ahead is known to hate deer more than he loves his car. Then in order to predict his actions one has to replace one's own aversion to a collision with an aim towards one.) So even an information-poor simulation account will have to make use of information about when and how to simulate, and this information is potentially very rich. Or consider how a Heal-type account can apply to situations in which the target person's problem-solving capacities are greater or less than those of the ascriber. (Their grasp of English syntax is feeble; their arithmetic is brilliant or eccentric.) Then in addition to solving the problem oneself one will have to apply some correction or transformation, to come up with the other person's solution. Most real cases will have some element of this, and will as a result require guidance from some beliefs about the other's problem-solving capacities and how to take them into account (Heal 2000; Morton 2002, ch. 2).

The line between applying a theory and reproducing the other person's thinking in your own mind is also vague. Suppose for example one is anticipating a person's actions by using a theory of belief/desire/action. Ascriptions of beliefs and desires will be needed, and there will usually be unmanageably many possibilities consistent with the person's prior behaviour. A natural approach to the problem is to use one's own desires and beliefs as a first approximation, adjusting the ascription as needed later on. In fact, we do unreflectively assume that others believe that the sky is above the earth and that death is to be avoided, in the absence of very strong evidence to the contrary. As Goldman (1989) has pointed out, this amounts to a kind of simulation: one uses one's own assessment of the situation as a guide to the thinking of the other. (The default ascriptions are information-rich in that a lot of information about the environment is taken into account, and information-poor in that one entertains fewer thoughts about the other person's thinking.)

Information-poor procedures have generally been found attractive when the task of anticipating another person's actions by tracing their beliefs and desires and reasoning is too daunting. In fact, even in the most favourable cases it is hard to see how one could anticipate actions purely by use of a theory of motivation. The reason is the combinatorial explosion of possible lines of reasoning. Given a desire for a small cup of coffee and enough coins to operate a machine that vends a large cup of decent coffee, a person might of course use the coins to buy the coffee and drink half of it, but she might also wait till someone else buys a cup and then buy half of it from them, or buy a cup and then sell half of it to another, or any of indefinitely many other actions, all of them easily rationalizable in terms of her beliefs and desires. But in many real situations we feel sure—somehow—what lines of thought people are likely to follow. It might then seem miraculous that this confidence succeeds, that people do quite often do what we expect them to. The reason, though, is that we do not blindly follow out all possible lines of reasoning available to a person. Instead, we expect people to follow reasoning that seems natural to us, the reasoning we ourselves would follow. And, of course, we adjust this expectation given what we learn about particular people's peculiarities.

The question of the exact mix of cognitively demanding information-rich theory and less demanding information-poor procedures remains very puzzling. It seems clear that we very rarely get to a prediction of what someone will do by combining explicit information about the person's beliefs and desires with explicit information about how human beings combine beliefs and desires to form intentions. On the other hand, it seems clear that we normally deploy a rich variety of information, about particular people and about people in general, in forming our expectations of what they are likely to do. Here is a suggestion about how the pieces fit together, a reasonable suggestion given the current state of the debate, but definitely a speculation.

There are three basic resources, on this suggestion. The first is a core folk-psychological theory, which specifies the basic characteristics of the basic psychological states. Call it the 'category theory', as it outlines the general categories of states that can be ascribed to people. It postulates informational states that represent the world as it is taken to be, target states that represent the world as the person aims to make it, emotions and other dispositions to types of behaviour, and processes of reasoning that lead from beliefs and desires to intentions or actions. Most of this theory is in place in four-year-old children, and parts of it emerge very early in life. But the theory does *not* specify what informational and target states there are. (It is neutral on the relation between what someone 'thinks', what she 'supposes', and what she 'believes', or between what she 'wants', 'needs', and 'would like'.) It does not specify what forms reasoning takes. (It simply says that people think and form intentions, and consider evidence and change their beliefs.) And it does not embody any ideas about rationality. (Wishful thinking is as likely as syllogisms, according to the core theory.)

The category theory becomes more powerful if it is scaffolded with some other skills. One obvious skill is linguistic. In learning to speak one acquires one's culture's vocabulary of mental-state terms and learns how to ascribe them. Belief-ascription—in particular, knowing what state is ascribed to a person when someone says 'she believes that p'—must be especially hard to learn, as the rules governing it are extremely subtle (Braun 1998). Given a command of the language, one can attribute to people the states ascribed to them by other people or themselves, as a basis for thinking about their minds. (How much of this ascription is literally true, and how much is part of a social web of mutual ascriptions which domesticated adults cannot easily act contrary to? That is not the kind of question that the philosophers or psychologists thinking about folk psychology pay a lot of attention to. For an exception see Kusch 1999.) And, given language, a great stock of platitudes, logical principles, folk generalizations, and old wives' tales becomes available, some of which will apply in almost any situation.

No amount of linguistic competence will allow one more than a primitive level of prediction and explanation without the second basic resource. That is simulation: the capacity to use one's own information and thinking as a guide to the thought of other people. One has to be able to say where a person's beliefs and desires might take her, for which the core theory gives no help. There are two related and possibly

prior capacities from which capacities of a generally simulational kind can develop. The first is the capacity, which develops early in normal children, to track the gaze of another person, to know where the other is looking. This allows a child to know which of the situations she is observing are also part of the information available to another. The second is the capacity for conditional thinking, which allows you to see what conclusions you would have come to if you had had counterfactual information and aims. Mastering counterfactual thinking requires that one learn how to separate beliefs and desires into consistent strands: wondering what would happen if the chair were put on the table involves separating off one's belief that the chair is not on the table, just as one has to when thinking what someone will do who thinks that the chair is on the table.

The capacity to follow another person's gaze is an essential prerequisite to using one's own mind to model the thinking of another. Conditional thinking on the other hand provides increasingly subtle ways of using information about another. For example, it allows one to explore the consequences of the fact that someone has not noticed something. They are both part of a third resource, neither category theory nor simulation. Let me call it the 'how-to manual'. This is a body of information about what kinds of simulation work under what conditions, and about what information about another person can be used as input to one's own thinking in modelling another's. It must also have an element of knowing how: how to use what one knows about others to fine-tune one's simulation of them. I call it a manual rather than a theory as it is likely to consist of a large number of rough generalizations and unconnected facts, and even of little tricks that work only for anticipating the thinking of particular individuals. Still, the manual is a body of rich if disparate information, largely second-order information about how to use the first-order information one has about other's minds to guide one's information-poor explanatory capacities.

The suggestion of a three-component capacity is conjectural. But the problems it addresses must be faced by any account of folk psychology. Once we accept that we use both information-rich and information-poor procedures, we must ask how they are combined. Are there standard ways of combining the basic components, which come more or less inevitably to human beings living human social lives? Or does each person work with their own improvised tool kit, made up from the same basic parts joined together in the ways that work for that person? We don't know.

41.3 DEVELOPMENT AND EVOLUTION

The capacity to understand one another as minds is a core human attribute. Human knowledge and social life would be impossible without it. Social life is the more fundamental; in terms of our abilities to know people's aims, intentions, beliefs, and personalities we can enter into cooperative activities in ways that minimize the danger of cheating or free-riding. Our capacity for knowledge is in a way a special case of this.

We negotiate a complex structure of sources of information largely by treating it as a social network in which a delicate cooperation is needed so that each person gets the information they need. Non-human animals do not have these capacities in anything like human form, and human infants have to go through a considerable process of development or learning before they can exercise adult competence. If therefore we could understand how individual humans come to acquire mind-ascribing capacities and how humanity as a species came to possess them, we would understand something very basic about what it is to be human.

To psychologists, as to philosophers, the idea that our capacity to think in mental terms is based on a theory was immediately very appealing. But psychologists immediately supposed that the 'theory of mind' must be an implicit theory that develops during a child's first few years of life. Early writers in the theory-of-mind movement, such as Wellman (1990) and Perner (1991), supposed that a child moves towards a theory the core of which postulates that human beings have representations of the environment, some in the form of beliefs and some in the form of desires. A key theoretical aim was to understand false-belief problems, the surprising difficulty children before the age of three have in attributing false beliefs. The experimental data, originally due to Wimmer and Perner (1983), concerns situations in which unambiguous evidence of a fact is available to a child but not to a 'target' person. For example candy is moved from one container to another, in the sight of the child but not of the target person. An adult would unhesitatingly ascribe to the target person a false belief, the belief that makes sense in the absence of the evidence. But small children instead ascribe to the target person a true belief, the belief that they themselves have formed in response to the evidence. They say that the target person will look for the candy in the place to which it has been moved, although she has not seen it moved.

The phenomenon is robust under a number of variations. It certainly shows that small children have difficulty with knowing when others will and will not make inferences. The difficult thought is: Evidence E is not available to X, so X will not have made the inference 'E therefore not p'. It does suggest that the concepts of knowledge and ignorance are easier to grasp than the concept of belief, for small children have little problem understanding that people can lack particular true beliefs. (That is why hide-and-seek is not a conceptual problem: the child can think 'X does not believe I am here' even though she may have difficulty with 'X believes that I am there when I am here'.) It is not clear how much trouble small children have with false beliefs when understanding inference is not required: for example when an authority simply announces that X thinks the candy is in location L (when in fact the child knows that it is in location L'). There does not seem to be a robust phenomenon of children then expecting X to look in location L'. Nor are false-belief difficulties insuperable; suitable coaching can improve the performance of children who are under the threshold. But the threshold is real and does mark a point in the development of normal human children. Some older children with developmental difficulties, notably autistic children, have the same difficulties with false-belief tasks that young children have, even though their general intelligence is on a level with

children who are well past the threshold. Great apes, such as chimpanzees, bonobos, and gorillas, do not solve analogues of classic false-belief tasks successfully, even though deliberate deception and distraction is not uncommon in their social life. It is as if they too can understand ‘p, but X does not believe p’ but cannot understand ‘p, but X believes not p’.

The rival accounts of folk psychology that we have already seen can be used to explain what is going on in false-belief tasks. Simulation accounts fit very naturally with the pre-threshold child’s behaviour: she models the target person’s beliefs on her own and predicts that the target person will do what she would do. What the child struggles to learn, on these accounts, is how to keep facts available to herself but not to the target person out of her modelling of the other’s thinking. Theory theory accounts, on the other hand, can explain very simply what the child past the threshold has learned. She has improved her theory of thinking so that it includes a clause ‘if X cannot observe that P then, other things being equal, X will not believe that p’. (This combined with the fact that X initially believes that not p, will entail that X will continue to hold the false belief that not p.)

Each account’s strength as applied to the false-belief situation links to a weakness. When a simulation account suggests that in passing the crucial threshold children learn how to be selective about which of their own beliefs to use in modelling another person’s thinking, it draws attention to the absence of any account of how one makes this selection. Since a selection is suitable or not depending on the situation of the target person, including that person’s beliefs and desires, it must be guided by some reflection on the person’s mind (at a minimum, reflection on whether the person was paying attention to the crucial inference-triggering evidence). On the other hand, when a theory theory suggests that the crucial transition consists in improving the child’s theory of inference, it draws attention to its ascription to the child of a theory of inference. But, as we have seen, explicit theories of inference are unmanageable monsters, quite useless for predicting what conclusions a person will arrive at.

The speculative account in the previous section, according to which a core theory specifies simply that people have beliefs and desires, and avoids any commitments about inference, can be used to get around these problems. Suppose that at the earlier stage of children’s development the core theory makes available to them the thought that others have beliefs, but that children rely for the content of these beliefs largely on simple rules such as: When a person can perceive something they have true beliefs about it; when something is obvious and a person has a belief about it then they believe the obvious. For updating the beliefs ascribed to a person in the face of evidence children would simply use their own belief changes as a guide to the thinking of the other person. Then, as they pass the threshold they would learn which of their own inferences can be used to model other people’s thinking under which circumstances. They would do this by forming a theory—not a theory of inference but rather a theory of how to track someone else’s reasoning with their own. Some of the applications of this theory are very subtle and complicated, and in principle the theory could continually improve throughout a person’s life, but at the threshold stage

it can consist of some fairly simple principles such as 'Model X's belief changes only using information available to X'.

If this suggestion is correct, a child emerges past the false-belief threshold with a conceptual structure characteristic of adult thinking about minds, though the range and power of her thinking about personality and motive may be much less than that of most adults. It consists of a collection of simulational capacities sandwiched between *two* theories, the category theory and the how-to manual mentioned above.

The suggestion here does not fit easily into a simple evolutionary story of how human mental-state ascription developed from the capacities of other apes. It neither points to a single crucial conceptual development which characterizes the human situation, nor allows us to see a clear series of steps leading from primate sociality to human thinking. For the basic elements of the category theory can be expected to be found in other apes, and the crucial false-belief threshold consists not in a development of this theory but in a series of small accretions to the how-to manual, which is much more dependent on language and on social interaction. And though the suggestion may of course be wrong, recent comparative primate work has marked a retreat from the once-fashionable view (Byrne and Whiten 1988; Byrne 1995) that the cooperation and deception in primate life makes it likely that many primates, and in particular chimpanzees, bonobos, and gorillas, possess many of the capacities essential to human thinking about mind, which require just a little tweaking and reconnecting to come together in their human form. Now, in contrast, the consensus (Povinelli and Eddy 1996; Sterelney 2004: ch. 4) is that ape capacities to track the attention of other apes, and to take account of information in the possession of others, are much more limited than we had for a while supposed, and that there are far more discontinuities than continuities with human cognition. That is not to say that we should see human folk psychology as emerging in a discontinuous evolutionary leap. In fact, it is much easier to give plausible accounts of its origins than for the origins of, say, human language.

It is not deeply puzzling how folk psychology could have evolved, though we do not know nearly enough to choose from among the possible routes. Our ancestors, like other apes, had the capacity to recognize dozens of individual conspecifics and remember their important characteristics; they could judge the mood of individual conspecifics by attention to their gaits and faces; they could form long-term and temporary coalitions for purposes as diverse as hunting and raising offspring; they could commit themselves to cooperative behaviour in situations in which it was in the interests of each individual to defect from such behaviour, and estimate the likelihood that another individual would defect; they could participate in routines for distribution of food and other resources and detect individuals who were abusing the routines. Many of these capacities were based on special-purpose cognitive processes aimed at specific types of situation, which, if evolutionary psychologists are correct (Barkow et al. 1992), still operate in human thought. But when we consider the effect of the slow accretion of capacities such as these we can see many ways in which they can scaffold the development of more general-purpose capacities, and in

particular the attribution of information-bearing states like those the philosophers and developmental psychologists have focused on.

Part of one story might go as follows. Proto-humans engage in cooperative hunting and foraging which requires them to keep track of what situations are perceivable to which others. This has a conceptual requirement, that they be able to represent the relation ‘X sees that p’, and a ‘pragmatic’ requirement, that they be able to figure out what one might be able to see from position p under current conditions. (These might be taken as among the precursors of the two folk-psychological theories mentioned above.) As these capacities improve, over thousands of years, language develops and with it the capacity to register reports of what people have seen, or take themselves to have seen, from places outside one’s field of view. Then the eventual combination of these three things—the concept of perception, the ability to take a perspective, and the understanding of verbal reports—results in the roots of an understanding of knowledge or belief. (The co-evolution of folk psychology and language is something we have to understand in order to get much further with these questions. The topic is very promising, but also very hard to handle responsibly.)

A similar story could be told for the evolution of the capacity to use attributions of belief and desire to anticipate behaviour. It would be a separate story, though, requiring a different, if overlapping, set of social skills as its starting point. (I would conjecture that crucial to this second story might be the capacity to manipulate others with misleading information, and to resist such manipulation.) Even when we combine the origin-of-belief story and this practical-reasoning story we do not have a full account, though. We need to explain how the conceptual and pragmatic components of these and other metacognitive skills come together to form a relatively integrated body of cognitive routines for connecting information, desire, emotion, and all the other states we attribute. It is plausible that this integration could not happen without language and its specific vocabulary for states of mind and more particularly its resources for using attributions in explanations of actions.

There are alternative stories, and the present evidence is not going to decide between them. The important point, though, is that we can see in principle how the evolution of folk psychology is possible. The crucial element is ‘niche construction’ (Sterelny 2004: ch. 8): the way in which a set of attributes can permit an animal, particularly a social animal, to change its environment in ways that allow other attributes to evolve. In the case at hand the first attributes are those that allow specific forms of cooperative activity, the niche is stable social life in which more effective cooperation and resistance to exploitation pays off, and the other attributes are the components of folk-psychological competence.

41.4 EPISTEMOLOGICAL QUESTIONS

Folk psychology allows us to attribute states of mind to others, and to use these attributions to predict what others will do. Do we *know* what states of mind others are

in, or what they will do? Are our beliefs about these things reasonable, or justified, or worth holding on to? It is not immediately obvious that these questions get answered with Yes. Compare crude folk meteorology. People look at the sky and smell the wind and consider the behaviour of plants and animals, and by instinct and traditional lore come up with opinions about whether it is a good time for the harvest. Sometimes they are right and sometimes they are wrong. They are wrong significantly often. The explanations they give have little to do with cold fronts or wind patterns in the stratosphere. Often supernatural or simply superstitious elements play an important role. It is far from obvious that the old-time farmer has a reasonable belief that the crops should be taken in early this year. Are our beliefs about one another any different?

Two basic considerations pull in opposite directions here. On the one hand, our attempts to understand one another are largely successful, in that human social life and human science flourish, and without an effective folk psychology they would be impossible. Isn't success its own justification? On the other hand, there is a now enormous body of social-psychological experiment which shows how wrong we often are about one another and ourselves. Among our failings, we think that people are more likely to repeat patterns of behaviour than in fact they are (Nisbett and Ross 1980); we underestimate how much people's opinions are affected by their social situations (Festinger 1964); we are often completely wrong about the reasons for which we have made choices (Nisbett and Wilson 1977; Kornblith 1989; Gopnik 1993); and we overestimate the accuracy of people's memory for details of events they have experienced (Loftus 1979; Conway 1997). We seem to make many, and systematic, mistakes.

Focus first on true beliefs. Human social competence entails that very often we anticipate one another's actions correctly. And it is not pushing this fact very much further to conclude that we are often accurate about other people's emotions (e.g. whether they are enraged or conciliatory) and about the information on which they are acting (e.g. whether someone has seen the letter which would bring about rage and confrontation). Accuracy about these things is consistent with error in other attributions that we make along the way to them. In particular, we may be right about what someone will do and wrong about why they will do it. The reasoning we attribute to someone in the course of predicting their action may contain thoughts the person will not think, and more profoundly may represent as causes of the person's action thoughts that are merely incidental. In fact, the psychological data on errors of attribution suggests that we are much less accurate about *why* we (and others) think what we do than about *what* we think. In the domain of mind, our explanations are even more suspect than our beliefs. (A potentially sceptical possibility is that our social competence relies less than we think on our beliefs, also, and instead on specific social skills for whose success folk psychology narrowly construed takes undeserved credit, see Morton 2002: ch. 1; Bermúdez 2004.)

These too-brief considerations suggest that it is important here to separate questions of justification from questions of knowledge. One of the original motives for postulating folk psychology was to uncover a body of beliefs that would result from a reasonable inference from our observations of other's actions. That motive was in

accordance with an 'internalist' epistemology, according to which beliefs are to be held to standards of evidence and reasonable inference. It does not now seem very promising. (The nearest contemporary account to such an ambition may be Gopnik and Meltzoff's account (1997) of the child as little scientist, going from one systematic explanation of the world and human nature to another.) But the possibilities of alternative 'externalist' approaches to epistemology are much clearer now than they were in the early days of folk psychology. On an externalist approach the central aim of beliefs is that they reliably represent the world, whether or not they result from acceptable reasoning. Knowledge becomes a more central concept than justification. (For the contrast between internalism and externalism in epistemology see Goldman 1986.) Information-poor models of attribution, notably simulationist models, fit the externalist approach very well. There is little information in the model to constitute a justification of the conclusion arrived at, but that conclusion is very often right. For example, on a co-cognitive picture of predicting someone's solution to a problem, as described by Heal, one does almost no attributing of any intermediate states to the target person. One simply has learned a routine of using one's own thinking as a guide to the other person's, and this leads one to attribute to the other a belief that, most of the time, is what in fact the person will be thinking.

Epistemological considerations are not going to determine which account of folk psychology is true. And although information-poor accounts have a natural affinity to externalist epistemology, many information-rich accounts can be incorporated into both internalist and externalist approaches. The facts that make an epistemology of folk psychology sustainable or not are deeply connected, though, with questions about the integration of folk psychology into the rest of our scientific and other beliefs. There are two extreme possibilities. At one extreme there is the combination of folk psychology as explicit theory with a conviction of its basic incompatibility with the truth about the mind-brain. That is eliminative materialism, and leads us to expect either that large parts of folk psychology should be replaced with beliefs better grounded in scientific fact or that it should be treated as a useful fiction. Beliefs derived from folk psychology are unjustified and thus untenable. At the other extreme there is the combination of folk psychology as simulation with a conviction that it makes very few factual assertions about states of mind or the causes of behaviour. That view (we might call it 'quietist compatibilism') leads us to expect that folk psychology can coexist with almost any scientific view of mind, since it makes so few substantive claims. Folk-psychological assertions are reliable and thus should be relied on. Neither position will find many defenders, since folk psychology makes more true assertions and is more essential to our social and intellectual lives than is claimed by a crude eliminativism, and makes more substantive claims about our states of mind and our reasons for action than is claimed by a simple quietism. The truth must lie somewhere between these extremes. And anywhere between these extremes we are going to be faced with a delicate mixture of true and false beliefs, very often with true beliefs derived from false beliefs, and very often with roughly reliable routes to conclusions about other people producing numbers of false beliefs as by-products. Is this an acceptable situation? Is it one that we should or can

try to change? If one is committed to an internalist epistemology one will find the situation uncomfortable, even if perhaps inevitable. For an externalist, though, it is a fairly typical example of the price we have to pay to get beliefs that serve reliably for particular purposes.

REFERENCES

- Barkow, J., Cosmides, L., and Tooby, J. (1992) (eds.), *The Adapted Mind: Evolutionary Psychology and the Development of Culture* (Oxford: Oxford University Press).
- Bermúdez, J. (2004), 'The Domain of Folk Psychology', in A. O'Hear (ed.), *Mind and Persons* (Cambridge: Cambridge University Press), 25–48.
- Braun, D. (1998), 'Understanding Belief Reports', *Philosophical Review*, 107: 555–96.
- Byrne, R. (1995), *The Thinking Ape: The Evolutionary Origins of Intelligence* (Oxford: Oxford University Press).
- Byrne, R., and Whiten, A. (1988) (eds.), *Machiavellian Intelligence: Social Expertise and the Evolution of Intelligence in Monkeys, Apes, and Humans* (Oxford: Oxford University Press).
- Churchland, P. (1979), *Scientific Realism and the Plasticity of Mind* (Cambridge: Cambridge University Press).
- Conway, M. (1997), *Recovered Memories and False Memories* (Oxford: Oxford University Press).
- Currie, G. (1995), 'Imagination and Simulation: Aesthetics Meets Cognitive Science', in M. Davies and T. Stone (eds.), *Mental Simulation: Applications and Evaluations* (Oxford: Blackwell), 151–69.
- Dennett, D. C. (1978), *Brainstorms* (Cambridge, Mass.: MIT).
- Festinger, L. (1964), *Conflict, Decision and Dissonance* (Palo Alto, Calif.: Stanford University Press).
- Goldman, A. (1986), *Epistemology and Cognition* (Cambridge, Mass.: Harvard University Press).
- (1989), 'Interpretation Psychologized', *Mind and Language*, 4: 161–85.
- Gopnik, A. (1993), 'How We Know Our Minds: The Illusion of First-person Knowledge of Intentionality', *Brain and Behavioral Sciences*, 16: 1–14.
- Gopnik, A., and Meltzoff, A. (1997), *Words, Thoughts, and Theories*, (Cambridge, Mass.: MIT Press).
- Gordon, R. (1995), 'Simulation Without Introspection, or Inference from Me to You', in M. Davies and T. Stone (eds.), *Mental Simulation: Evaluations and Applications* (Oxford: Blackwell), 53–67.
- Heal, J. (1995), 'How to Think About Thinking', in M. Davies and T. Stone (eds.), *Mental Simulation: Evaluations and Applications* (Oxford: Blackwell), 33–52.
- (2000), 'Other Minds, Rationality, and Analogy', *Proceedings of the Aristotelian Society*, suppl. vol. 74: 1–19.
- Kornblith, H. (1989), 'Introspection and Misdirection', *Australasian Journal of Philosophy*, 67: 410–22.
- Kusch, M. (1999), *Psychological Knowledge: A Social History and Philosophy* (London: Routledge).
- Loftus, E. (1979), *Eyewitness Testimony* (Cambridge, Mass.: Harvard University Press).
- Morton, A. (1980), *Frames of Mind: Constraints on the Common-sense Conception of the Mental* (Oxford: Oxford University Press).

- Morton, A. (2002), *The Importance of Being Understood: Folk Psychology as Ethics* (London: Routledge).
- Nichols, S., and Stich, S. P. (2003), *Mindreading* (Oxford: Oxford University Press).
- Nisbett, R., and Ross, L. (1980), *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, NJ: Prentice-Hall).
- Nisbett, R. E., and Wilson, T. (1977), 'Telling More Than We Can Know: Verbal Reports on Mental Processes', *Psychological Review*, 84: 231–59.
- Perner, J. (1991), *Understanding the Representational Mind* (Cambridge, Mass.: MIT Press).
- Povinelli, D., and Eddy, T. (1996), 'What Young Chimpanzees Know About Seeing', *Monographs of the Society for the Study of Child Development*, 61: 1–152.
- Putnam, H. (1960), 'Minds and Machines', in S. Hook (ed.), *Dimensions of Mind* (New York: New York University Press), 148–79.
- Quine, W. V. (1953), 'Two Dogmas of Empiricism', in Quine, *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press), 20–46.
- Sellars, W. (1956), 'Empiricism and the Philosophy of Mind', in H. Feigl and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis* (Minneapolis, Minn.: University of Minnesota Press), 253–329.
- Sterelney, K. (2004), *Thought in a Hostile World* (Oxford: Blackwell).
- Strawson, P. (1959). *Individuals: An Essay in Descriptive Metaphysics* (London: Methuen).
- Wellman, H. (1990), *The Child's Theory of Mind* (Cambridge, Mass.: MIT Press).
- Wimmer, H., and Perner, J. (1983), 'Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception', *Cognition*, 13: 103–28.