

Notice: This material may be
protected by copyright law.
(Title 17, U.S. Code)

2

Folk Psychology as Simulation

ROBERT M. GORDON

Recently I made a series of predictions of human behavior, using the meager resources allotted to a non-scientist. Having nothing to rely on but 'common-sense' or 'folk' psychology and being well forewarned of the infirmities of that so-called theory, I had reason to anticipate at best a very modest rate of success.

These were the predictions:

I shall now pour some coffee.
I shall now pick up the cup.
I shall now drink the coffee.
I shall now switch on the word processor.
I shall now draft the opening paragraphs of a paper on folk psychology.

My predictions, as I think no one will be surprised to learn, proved true in every instance. Should anyone doubt this, I recommend spending a few minutes predicting from one moment to another what you are 'about to do'. Such predictions, if not quite as reliable as 'night will follow day' or 'this chair will hold my weight', are at least among the most reliable one is likely to make. Of course, one would have to allow for unforeseen interventions by 'nature' (sudden paralysis, a coffee cup glued to the table) and for ignorance (the stuff you pour and drink isn't coffee). But that seems a realistic limitation on any *psychological* basis for prediction.

This paper offers an account of the nature of folk psychology. Sections 1 and 2 focus on the prediction of behavior, beginning with reflections on my little experiment in prediction. Section 3 concerns the interaction of explanation and prediction in what I call hypothetico-practical reasoning. Finally, a new account of belief attribution is proposed and briefly defended in section 4.

1 Predicting One's Own Behavior

At least one lesson can be drawn from my prediction experiment. Discussions of the nature of 'folk psychology' and of its own adequacy, particularly as a basis for predictions of overt human behavior, ought to begin by dividing the question: *one's own* behavior or *another's*; behavior in the *immediate* or in the *distant* future; behavior under *existing* conditions or under specified *hypothetical* conditions? For such a division uncovers a little-known and unappreciated success story: our prodigious ability to foretell what we ourselves are 'about to do' in the (actual) immediate future. We have in this department a success rate that surely would be the envy of any behavioral or neurobehavioral science.

The trick, of course, is not to predict until one has 'made up one's mind' what to do: then one simply declares what one 'intends' to do. We display our confidence in the *predictive reliability* of these declarations by the way we formulate them: one typically says, not 'I intend now to . . .', but simply 'I shall now . . .' or 'I will now . . .' Somehow, in learning to 'express our (immediate) intention' we learn to utter sentences that, construed as statements about our own future behavior, prove to be extremely reliable.¹ (Normally, apart from the conditions mentioned earlier, the only errors occur when something 'makes us change our mind': the telephone rings before we have poured the coffee, we see that the stuff isn't coffee, and so on.) A plausible explanation of this reliability is that our declarations of immediate intention are causally tied to some actual precursor of behavior: perhaps tapping into the brain's updated behavioral 'plans' or into 'executive commands' that are about to guide the relevant motor sequences.² In any case, these everyday predictions of behavior seem to have an anchor in psychological reality.

One might have thought all predictions of human behavior to be inferences from theoretical premises about beliefs, desires, and emotions, together with laws connecting these with behavior: laws of the form: 'if A in states S1, S2, S3, etc., and conditions C1, C2, C3 obtain, then A will (or will probably) do X'. Thus one would have a *deductive-nomological* or *inductive-nomological* basis for prediction. This is plainly not so: declarations of immediate intention – 'I shall now X' – are not products of inference from such premises.

Moreover, if they were, one could not account for either their predictive reliability or our *confidence* in their predictive reliability. We are not self-omniscient: we do not keep tabs on all of the relevant beliefs and attitudes, and *a fortiori* we do not keep a *reliable* inventory of these. But even if we knew all the relevant beliefs and attitudes, our predictions would at best be qualified and chancy. Folk psychology, on most accounts, doesn't specify a deterministic system; it specifies only the probable or 'typical' effects of mental states. Using it as my basis I should have to qualify my predictions by saying, e.g. 'Typically, I would now pick up the cup.' And actions that are *atypical*, *exceptional*, or *out of character* – my wearing a tie to class, or my heckling the commencement speaker – would defy prediction altogether, even seconds before I take action. Whereas in fact I feel confident that I can predict what I

am about to do now, whether the act is typical or not; and my confidence seems well-founded: I predict imminent atypical actions about as reliably as any others.

Although they are not based on nomological reasoning, declarations of immediate intention – these ultra-reliable predictors of behavior – are often products of *practical* reasoning: reasoning that provides the basis for a decision to do something.³ I shall now write a letter' may express a decision based on certain salient facts (a student asked me to write a letter of recommendation), salient norms and values (I have a duty to write letters for good students who request letters, and she's a good student), and a background of other facts, norms, and values that I am unable to list exhaustively. The important point is that declarations of the form: 'I shall now do X' offer a bridge between such practical reasoning and prediction.

This bridge introduces a very interesting possibility: that of using *simulated* practical reasoning as a *predictive* device. First of all, it is easy to see how, by simulating the appropriate practical reasoning, we can extend our capacity for self-prediction in a way that would enable us to predict *our own behavior in hypothetical situations*. Thus I might predict, for example, what I would do if, right now, the screen of the word processor I am working on were to go blank; or what I would do if I were now to hear footsteps coming from the basement.

To simulate the appropriate practical reasoning I can engage in a kind of *pretend-play*: pretend that the indicated conditions *actually obtain*, with all other conditions remaining (so far as is logically possible and physically probable) as they presently stand; then – continuing the make-believe – try to 'make up my mind' what to do given these (modified) conditions. I imagine, for instance a lone modification of the actual world: the sound of footsteps from the basement.⁴ Then I ask, in effect, 'what shall I do now?' And I answer with a declaration of immediate intention, 'I shall now ...' This too is only feigned. But it is not feigned on a *tabula rasa*, as if at random: rather, the declaration of immediate intention appears to be formed in the way a *decision* is formed, *constrained* by the (pretended) 'fact' that there is the sound of footsteps from the basement, the (unpretended) fact that such a sound would now be unlikely if there weren't an intruder in the basement, the (unpretended) awfulness of there being an intruder in the basement, and so forth.

What I have performed is a kind of *practical simulation*, a simulated deciding *what to do*. Some simulated decisions in hypothetical situations include acting out: e.g. rehearsals and drills. The kind I am interested in, however, suppress the behavioral output. One reports the simulated decision as a *hypothetical prediction*: a prediction of what I would do in the specified hypothetical circumstances, other things being as they are. For example: if I were now to hear footsteps from the basement, (probably) I would reach for the telephone and call an emergency number.⁵

I noted earlier that one could not account for either the *confidence* or the *reliability* with which I predict what I am about to do now, if such predictions were based on attributions of beliefs, desires, etc., together with laws. The

same holds for *hypothetical* self-predictions. Once again I don't know enough about my beliefs and desires; and the laws would at best yield only the *typical* effects of those states, anyway.⁶ In real life we sometimes surprise ourselves with *atypical* responses: 'I certainly wouldn't have thought I'd react that way!' Practical simulation imitates real life in this respect, giving us the capacity to surprise ourselves *before* we confront the actual situation. If I pretend *realistically* that there is an intruder in the house I *might* find myself surprisingly brave – or cowardly.⁷

2 Predicting the Behavior of Others

In one type of hypothetical self-prediction the hypothetical situation is one that some *other* person has actually been in, or at least is described as having been in. The task is to answer the question, 'What would I do in *that* person's situation?' For example, chess players report that, playing against a human opponent or even against a computer, they visualize the board from the other side, taking the opposing pieces for their own and vice versa. Further, they pretend that their *reasons for action* have shifted accordingly: whereas previously the fact that a move would make White's Queen vulnerable would constitute a reason *for* making the move, it now becomes a reason *against*; and so on. Thus transported in imagination, they 'make up their mind what to do.' *That*, they conclude, is what I would do (have done). They are 'putting themselves in the other's shoes' in one sense of that expression: that is, they project themselves into the other's *situation*, but without any attempt to project themselves into, as we say, the other's 'mind'.

A prediction of how I would act in the other's situation is not, of course, a prediction of how the other will act – unless, of course, the other should happen to be, in causally relevant respects, a *replica* of me. But people claim also that by 'putting themselves in the other's shoes', in a somewhat different sense of that expression, they can predict the *other's* behavior. As in the case of hypothetical self-prediction, the methodology essentially involves *deciding what to do*; but, extended to people of 'minds' different from one's own, this is not the same as deciding *what I myself would do*. One tries to make *adjustments for relevant differences*. In chess, for example, a player would make not only the imaginative shifts required for predicting 'what I would do in his shoes', but the further shifts required for predicting what *he* will do in his shoes. To this purpose the player might, e.g. simulate a lower level of play, trade one set of idiosyncrasies for another, and above all pretend ignorance of *his own* (actual) intentions. Army generals, salespeople, and detectives claim to do this sort of thing. Sherlock Holmes expresses the point with characteristic modesty:

You know my methods in such cases, Watson. I put myself in the man's place, and, having first gauged his intelligence, I try to imagine how I should myself have proceeded under the same circumstances. In this

case the matter was simplified by Brunton's intelligence being quite first-rate, so that it was unnecessary to make any allowance for the personal equation, as the astronomers have dubbed it. (Doyle, 1894)

The procedure serves cooperative as well as competitive ends: not to go far afield, bridge players claim they can project themselves into their *partner's* shoes.

Several earlier philosophers claimed that interpersonal understanding depends on a procedure resembling what I call simulation. Precursors of simulation include historical reenactment (Collingwood, 1946) and *Verstehen* or 'empathetic understanding' (Schutz, 1962, 1967; von Wright, 1971, ch. 1).⁸ But little attention has been given to prediction. Nor have these authors appreciated the methodological importance of hypothesis-testing and experimentation in practical simulation: the fact that at its heart is a type of reasoning I characterize as *hypothetico-practical*. Finally, they have not tried to explain the very concept of belief in terms of practical simulation, as I shall.

3 *Hypothetico-practical Reasoning*

Let me illustrate with an extended example of hypothetico-practical reasoning. A friend and I have sat down at a table in a fashionable international restaurant in New York. The waiter approaches. He greets me effusively in what strikes me as a Slavic language. He says nothing to my friend. I do not speak any Slavic language.

I wish to understand the waiter's behavior. I wish also to predict his future behavior, given various responses I might make to his greeting. As a first step, I shift spatiotemporal perspectives – I am standing over there now, where the waiter is; not sitting here. In some cases, shifting spatiotemporal perspectives might be enough: e.g. for predicting, or explaining, the behavior of a person I see in the path of an oncoming car. This would be woefully inadequate for the restaurant example, of course. As a further experiment, I might switch institutional roles. I suppose (and perhaps imagine) myself to be a waiter, waiting on a customer sitting here in this restaurant. Such counterfactual suppositions raise difficult questions: for example, shall I suppose myself to be a waiter who has read Quine (as I have)? Shall I suppose the *customer* to have read Quine?⁹ Fortunately, I do not ordinarily have to ask these questions, since they would make no difference in my behavior in this situation; and when they do make a difference, the situation is likely to alert me to their relevance.

Donning my waiter's uniform is clearly not enough: to have what I see as a basis for greeting the customer in a Slavic language, supposing I could, I shall have to alter other facts. As a first stab, I might see myself as an *émigré* from a Slavic country, working as a waiter. I seem to recognize the customer: he is a countryman of mine who used to eat at the restaurant many years ago. It pleases him, as I recall, to be greeted in our native tongue. That would be a reason for doing so. There being no reason not to, that is what I shall do.

Other modifications of the world would lead to the same decision. Suppose that, before the restaurant episode, I had read a cheap spy thriller. Under its corrosive influence I hypothesize as follows: I am a counterintelligence agent posing as a waiter. The customer I am waiting on is a known spy from a Slavic country, and there is good reason to get him to reveal that he knows the language of this country. One way to do this is to watch his reaction as I address him in the language of his country. Given this background, I would indeed address him in that language, if I could.

To choose between the two hypotheses would require further tests. Suppose that in my real role of customer, I look puzzled and respond: 'I don't understand that language. You must be making a mistake.' On the countryman hypothesis, the waiter will probably apologize – in English – and explain that he had mistaken me for someone else. On the counterspy hypothesis, he may either persist in speaking in the foreign tongue or turn to more subtle devices for getting the customer to reveal his knowledge of the language. If in fact the waiter apologizes, then the counterspy hypothesis will have suffered one perhaps small defeat.

Ideally, the hypothesis-testing would continue until the subject appeared to be, as it were, *the puppet of my (simulated) intentions*. In actuality, when I persist in my effort to find a pretend-world in which the other's behavior would accord with my intentions, I usually find myself, after a number of errors, 'tracking' the other person fairly well, forming a fairly stable pretend-world for that person. Of course, I cannot predict or anticipate exactly what he will do, to any fine-grained description. But, by and large, I will not be very much surprised very often, at least in matters that are important to interpersonal coordination.

No matter how long I go on testing hypotheses, I will not have tried out all candidate explanations of the waiter's behavior. Perhaps some of the unexamined candidates would have done at least as well as the one I settle for, if I settle: perhaps indefinitely many of them would have. But these would be 'far-fetched', I say intuitively. Therein I exhibit my inertial bias: the less 'far-fetched' (or 'stretching', as actors say) I have to do to track the other's behavior, the better. I tend to *feign* only when necessary, only when something in the other's behavior doesn't fit. This inertial bias may be thought of as a 'least effort' principle: the 'principle of least pretending'. It explains why, other things being equal, I will prefer the less radical departure from the 'real' world – i.e. from what I myself take to be the world.

Within a close-knit community, where people have a vast common fund of 'facts' as well as shared norms and values, only a minimum of pretending would be called for. (In the limit case – a replica – the distinction noted earlier between 'what I would do in the other's situation' and 'what *the other will* do in his situation' would indeed vanish, except as a formal or conceptual distinction: what *I would* do and what *the other will* do would invariably coincide.) A person transplanted into an alien culture might have to do a great deal of pretending to explain and predict the behavior of those around him. Indeed, one might eventually learn to *begin* all attempts at explanation and

prediction with a stereotypic set of adjustments: pretending that dancing causes rain, that grasshoppers taste better than beefsteak, that blue-eyed should never marry brown-eyed, and so on. This 'default' set of world-modifications might be said to constitute one's 'generalizations' about the alien culture.

Whether or not practical simulation begins with such stereotypes, it does not essentially involve (as one might think) an implicit *comparison to oneself*. Although it does essentially involve *deciding what to do*, that, as I have noted, is not the same as deciding *what I myself would do*. To predict another's behavior I may have to pretend that there is an Aryan race, that it is meta-physically the master race, and that I belong to it; finally, that I was born in Germany of German stock between 1900 and 1920. To make decisions within such a pretend-world is not to decide what *I myself* would do, much less to reliably know what *I myself* would do 'in that situation'. First, it is not possible for *me* to be in that situation, if indeed it is a *possible* situation (for anyone); second, it is not possible for me even to *believe* myself to be in that situation – not, at least, without such vast changes in my beliefs and attitudes as to make all prediction unreliable. Hence in such a case I cannot be making an implicit *comparison to myself*.¹⁰

4 Attributions of Beliefs

I do not deny that explanations are often couched in terms of *beliefs, desires*, and other propositional attitudes; or that predictions, particularly predictions of the behavior of others, are often made on the basis of attributions of such states. Moreover, as functionalist accounts of folk psychology have emphasized, common discourse about beliefs and other mental states presupposes that they enter into a multitude of causal and nomological relations. I don't want to deny this either. A particular instance or 'token' of belief, such as Smith's belief that Dewey won the election, may be (given a background of other beliefs, desires, etc.) a *cause* of Smith's doing something (joining the Republican party) or undergoing something (being glad, being upset); it may have been *caused by* his reading in the newspapers that Dewey won and his believing that newspapers are reliable in such matters, or by his having taken a hallucinogenic drug.

There are in addition certain formally describable regularities that might be formulated as laws of typical causation: e.g. a belief that *p* and a belief that (if *p* then *q*) will typically cause a belief that *q*; a desire that *p* and a belief that (if and only if I bring it about that *q*) will typically cause a desire to bring it about that *q*. And there are more specific regularities that obtain for particular individuals, classes, communities, or cultures: e.g. when some tennis players believe their opponents aren't playing at their best they typically get angry; when members of a certain tribe see a cloud they think inhabited by an animal spirit they typically prepare for the hunt. Sometimes it helps to remember such regularities when predicting or explaining behavior – even one's own.

One mustn't apply such generalizations too mechanically, however. For there are indefinitely many circumstances, not exhaustively specifiable in advance, in which these general or specific regularities fail to hold. Those generalizations that do not explicitly concern only 'typical' instances should be understood to contain implicit *ceteris paribus* clauses. (This point is developed, along with much else that is congenial, in Putnam, 1978, lecture VI.) How does one know how to recognize atypical situations or to expand the *ceteris paribus* clause? An answer is ready at hand. As long as one applies these generalizations in the context of *practical simulation*, the unspecifiable constraints on *one's own* practical reasoning would enable one to delimit the application of these rules. This gives one something to start with: as one learns more about others, of course, one learns how to modify these constraints in applying generalizations to them.

Moreover, the *interpretation* of such generalizations, as indeed of all common discourse about beliefs and other mental states, remains open to question. In the remainder of this paper I sketch and at least begin to defend a way of interpreting ordinary discourse about beliefs in terms of pretend play and practical simulation. The idea isn't wholly new. In *Word and Object*, Quine explained indirect quotation and the ascription of propositional attitudes in terms of what he called 'an essentially dramatic act':

We project ourselves into what, from his remarks and other indications, we imagine the speaker's state of mind to have been, and then we say what, in our language, is natural and relevant for us in the state thus feigned. (1960, p. 92)

That is, we first try to simulate, by a sort of pretending, another's state of mind; then we just 'speak our mind'. In Quine's view, this is essentially an exercise in translation and heir to all its problems. Stephen Stich develops the idea further, using a device introduced by Davidson: in saying, e.g. 'Smith believes that Dewey won', one utters the content sentence 'Dewey won', pretending to be asserting it oneself, as if performing a little skit (Stich, 1983). To ascribe such a belief to Smith is to say that he is in a state *similar* to the one that might typically be expected to underlie that utterance – had it not been produced by way of play-acting.

As Stich portrays the play-acting device, it is merely a device for producing a specimen utterance which in turn is used to specify a particular theoretical state. The attribution of such a state is supposed to play a role in nomological reasoning roughly analogous to that of attributions of theoretical states in the physical sciences, and in that role to serve in the tasks of explaining and predicting the object's behavior.¹¹ Rather than treating the observer as an *agent* in his own right, as one who might form intentions to *act* on the basis of pretend inputs, it calls upon him merely to *speak* as he would given those inputs.

Stich's assumption that the methodological context for such attributions is nomological reasoning leads him, I believe, to misrepresent the role of pre-

tending in folk psychology. I shall sketch very briefly a different role for pretending in belief attribution. On this account, the methodological context for such attributions is not nomological reasoning but practical simulation.

A chess player who visualizes the board from his opponent's point of view might find it helpful to *verbalize* from that point of view – to assert, for example, 'my Queen is in danger.' Stepping into Smith's shoes I might say: 'Dewey won the election.' Such assertions may then be used as premises of simulated practical inference. But wouldn't it be a great advantage to us practical simulators if we could *pool our resources*? We'll simulate Smith *together*, cooperatively, advising one another as to what premises or inputs to practical reasoning would work best for a simulation of Smith. That is, give the best predictions and the most stable explanations, explanations that won't have to be revised in the light of new evidence. Of course, I couldn't come *straight out* with the utterance: 'Dewey won.' I need to flag the utterance as one that is being uttered *from within a Smith-simulation mode* and addressed to *your* Smith-simulation mode. I might do this by saying something like the following:

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

The same task might be accomplished by saying:

- 2 Smith believes that Dewey won the election.

My suggestion is that (2) be read as saying the same thing as (1), though less explicitly.

It is worth noting that unlike Stich I am not characterizing belief as a relation to any linguistic entity or speech act, e.g. a sentence or an assertion. Nor, as far as I can see, does my suggestion involve explicating the contents of belief in terms of possible worlds. Rather than specifying in a *standard non-pretending mode of speech* a set of possible worlds, one says something about the *actual world*, albeit in a *non-standard, pretending mode of speech*. Needless to say, the exposition and defense of this account of belief are much in need of further development. But it is interesting to note that, given the 'principle of least pretending' mentioned earlier, our belief attributions would be in accord with something like the 'principle of charity' put forward by Quine and Davidson: roughly, that one should prefer a translation that maximizes truth and rationality. More precisely, our attributions would conform to an improved version of this principle: Grandy's more general 'principle of humanity' according to which one should prefer a translation on which 'the imputed pattern of relations among beliefs, desires, and the world be as similar to our own as possible' (Grandy, 1973, p. 443).

If I am right, to attribute a belief to another person is to make an assertion, to state something as a fact, *within the context of practical simulation*. Acquisition of the capacity to attribute beliefs is acquisition of the capacity to make assertions in such a context. There is some experimental support for this view. Very young children give verbal expression to predictions and explanations

of the behavior of others. Yet up to about the age of four they evidently lack the concept of belief, or at least the capacity to make allowances for false or differing beliefs. Evidence of this can be teased out by presenting children with stories and dramatizations that involve dramatic irony: where we the audience know something important the protagonist doesn't know (Wimmer and Perner, 1983).¹²

In one such story (illustrated with puppets) the puppet-child Maxi puts his chocolate in the box and goes out to play. While he is out, his mother transfers the chocolate to the cupboard. Where will Maxi look for the chocolate when he comes back? In the box, says the five-year-old, pointing to the miniature box on the puppet stage: a good prediction of a sort we ordinarily take for granted. (That is, after all, where the chocolate had been before it was, without Maxi's knowledge, transferred to the cupboard.) But the child of three to four years has a different response: verbally or by pointing the child indicates the cupboard. (That is, after all, where the chocolate is to be found, isn't it?) Suppose Maxi wants to mislead his gluttonous big brother to the *wrong* place, where will he lead him? The five-year-old indicates the cupboard, where (unknownst to Maxi) the chocolate actually is; often accompanying the response with what is described as 'an ironical smile'. The *younger* child indicates, incorrectly, the box.¹³

From this and other experiments it appears that normal children around age four or five vastly increase their capacity to predict the behavior of others. The child develops the ability to make allowances for what the other isn't in a position to know. She can predict behavioral failures, e.g. failure to look in the right place, failure to mislead another to the wrong place, that result from *cognitive* failures, i.e. *false beliefs*. At an earlier age she makes all predictions in an egocentric way, basing them all on the *actual facts*, i.e. the facts as she herself sees them. She either lacks the concept of belief altogether or at least lacks the ability to employ it in the prediction of behavior. One may even say that the young child attributes knowledge – by default – before she has learned to attribute belief.

It is the position of many philosophers that common-sense terms such as 'believes' are *theoretical* terms, the meanings of which are fixed in the same way as theoretical terms in general: by the set of laws and generalizations in which they figure. This view is widely (but not universally) assumed in functionalist accounts of folk psychology. (It is the offspring of the dispositional theories that were popular in the days of philosophical behaviorism.) Presumably, mastery of the concept of belief would then be a matter of internalizing a sufficiently large number of laws or generalizations in which the term 'belief' (and related verb forms) occurs. The term 'belief' would be used in something like the way biologists used the term 'gene' before the discovery of DNA.¹⁴

But suppose that mastery of the concept of belief did consist in learning or in some manner internalizing a system of laws and generalizations concerning belief. One would in that case expect that *before* internalizing this system, the child would simply be unable to predict or explain human action. And *after* internalizing the system the child could deal indifferently with actions

caused by *true* beliefs and actions caused by *false* beliefs. It is hard to see how the semantical question could be relevant.

Suppose on the other hand that the child of four develops the ability to make assertions, to state something as fact, *within the context of practical simulation*. That would give her the capacity to overcome an initial egocentric limitation to the *actual facts* (i.e. as *she* sees them). One would expect a change of just the sort we find in these experiments.

There is further evidence. Practical simulation involves the capacity for a certain kind of systematic pretending. It is well known that *autistic* children suffer a striking deficit in the capacity for pretend play. In addition they are often said to 'treat people and objects alike'; they fail to treat others as subjects, as having 'points of view' distinct from their own. This failure is confirmed by their performance in prediction tests like the one I have just described. A version of the Wimmer-Perner test was administered to autistic children of ages *six to sixteen* by a team of psychologists (Baron-Cohen, Leslie, and Frith, 1985). *Almost all* these children gave the wrong answer, the three-year-old's answer. This indicates a highly specific deficit, not one in general intelligence. Although many autistic children are also mentally retarded, those tested were mostly in the average or borderline IQ range. Yet children with Down's syndrome, with IQ levels substantially below that range, suffered no deficit: almost all gave the right answer. My account of belief would predict that only those children who can engage in pretend play can master the concept of belief.¹⁵ It is worth noting that autistic children do at least as well as normals in their comprehension of *mechanical* operations – a distinct blow to any functionalists who might think mastery of the concept of belief to consist in the acquisition of a theory of the functional organization of a mechanism.

I suspect that, once acquired, the capacity for practical simulation operates primarily at a sub-verbal level, enabling us to *anticipate* in our own actions the behavior of others, though we are unable to say *what* it is that we anticipate or *why*. The *self-reported* pretending I have described would then only be the tip of the iceberg. Something like it may happen quite regularly and without our knowledge: our decision-making or practical reasoning system gets partially disengaged from its 'natural' inputs and fed instead with suppositions and images (or their 'subpersonal' or 'subdoxastic' counterparts). Given these artificial pretend inputs the system then 'makes up its mind' what to do. Since the system is being run off-line, as it were, disengaged also from its natural output systems, its 'decision' isn't actually executed but rather ends up as an anticipation, perhaps just an unconscious *motor* anticipation, of the other's behavior.

One interesting possibility is that the readiness for practical simulation is a prepackaged 'module' called upon automatically in the perception of other human beings. One might even speculate that such a module makes its first appearance in the useful tendency many mammals have of turning their eyes toward the target of another's gaze. Thus the very sight of human eyes might *require* us to simulate at least their spatial perspective – and to this extent, at least, to put ourselves in the other's shoes. This would give substance to

the notion that we perceive one another primarily as *subjects*: as world-centers rather than as objects in the world. It is pleasant to speculate that the phenomenology of *the Other* – particularly the Sartrean idea that consciousness of the Other robs us of our own perspective – might have such humble beginnings.

It remains the prevailing view of philosophers and cognitive scientists that mental states, as conceived by naïve folk psychology, are constructs belonging to a pre-scientific theory of the inner workings of the human behavior control system housed, as we now know, in the brain. One problem with this conception of folk psychology is that mastery of its concepts would seem to demand a highly developed theoretical intellect and a methodological sophistication rivaling that of modern-day cognitive scientists. That is an awful lot to impute to the four-year-old, or to our savage ancestors. It is also uncanny that folk psychology hasn't changed very much over the millennia. Paul Churchland writes:

The [folk psychology] of the Greeks is essentially the [folk psychology] we use today, and we are negligibly better at explaining human behavior in its terms than was Sophocles. That is a very long period of stagnation and infertility for any theory to display. (1981, p. 75)

Churchland thinks this a sign that folk psychology is a bad theory; but it could be sign that it is no theory at all, not, at least, in the accepted sense of (roughly) a system of laws implicitly defining a set of terms. Instead it might be just the capacity for practical reasoning, supplemented by a special use of a childish and primitive capacity for pretend play. I hope that I have shown that to be a plausible and refreshing alternative.¹⁶

Notes

- 1 The qualifying phrase is added because I am concerned with *assertive* reliability, not *commissive* reliability: that of predictions, rather than that of promises, vows, and expressions of intention. Construing 'I shall now X' as a mere expression of intention, if the speaker does not X he will have 'failed to carry out' his intention: his *action* would in a (non-moral) sense be 'at fault'. Construing it as a mere prediction, on the other hand, it would be the *prediction* that is 'at fault', not the *action*. To use Seale's distinction (derived from Anscombe), declarations of intention have a world-to-word 'direction of fit', whereas predictions and other 'assertive' speech acts have a word-to-world direction of fit (Searle, 1983). This distinction does not affect the essential point being made here.
- 2 A further possibility is that a degree of normative commitment is added by the *declaration* of an intention, even if it is announced only to oneself: one is then motivated to *model* one's behavior to the declared intention. This was suggested to me by Brian McLaughlin.

- 3 More precisely, *what is expressed* by these declarations are often products of practical reasoning.
- 4 Imagery is not always needed in such simulations. For example, I need no imagery to simulate having a million dollars in the bank. Mere *supposition* would be enough.
- 5 Contrast, 'I would (if such a situation were now to arise) reach for the telephone and dial an emergency number' uttered as a declaration of *conditional intention*. The difference can be partially explicated in terms of 'direction of fit'.
- 6 Granted, if one were to do some of the pretending out loud, one might say, e.g. 'I believe someone has broken into the house.' But such a verbalization has a role in practical not nomological reasoning: one is articulating a possible basis for action, not giving a state description that is to be plugged into laws that bridge between internal states and behavior.
- 7 Needless to say, like any attempt to explain or predict one's own behavior, this may be corrupted by prejudice or self-deception.
- 8 Closer to my own view is Morton (1980, ch. 3) on the uses of imagination in understanding another's behavior.
- 9 As Quine has noted: 'Casting our real selves thus in unreal roles, we do not generally know how much reality to hold constant' (1960, p. 92).
- 10 Nozick seems to miss this point in his account of *Verstehen* as 'a special form of inference by analogy, in that I am the thing to which he is analogous'. He argues that the inferences depends on two empirical correlations: 'that he acts as I would, and that I would as I (on the basis of imaginative projection) think I would' (1981, p. 636). Nozick's mistake is to think it relevant to ask, and indeed essential as the inferential link, how I would *in fact* behave in the other's shoes.
- 11 To do *this* job properly, it would have to meet certain standards of objectivity. And Stich argues with considerable force that it cannot. For it never frees itself fully from the subjectivity it necessarily begins with, the speaker-relativity that is built into the ascription of content.
- 12 The psychologists who conducted this study credit three philosophers (J. Bennett, D. Dennett, and G. Harman) with suggesting the experimental paradigm, each independently, in commentaries published with Premack and Woodruff, 1978).
- 13 My account simplifies the experiment and the results; but not, I think, unconscionably.
- 14 But a functionalist might wish to say that, whereas the correct *explication* of the concept requires that one cite such laws, *mastery* of the concept, i.e. capacity to *use* it, does not require that one have internalized such laws. Thus some functionalists might even be prepared to embrace something like my account of belief attributions. This possibility (or something close to it) was pointed out to me independently by Larry Davis and Sydney Shoemaker. I am inclined to think that this would be an uneasy alliance, but I confess I don't (as yet) have the arguments to persuade anybody who might think otherwise.

- 15 My account is close in many respects to the theory the investigators were themselves testing in the autism experiment. This is presented in Leslie, 1987.
- 16 I have benefited enormously from the advice and criticism of Stephen Stich, in correspondence and conversation. I am indebted to Fred Adams and Larry Davis for much help in seeing through the murk, and have benefited further from comments by Robert Audi, John Barker, Hartry Field, Brian McLaughlin, Sydney Shoemaker, Raimo Tuomela, and (no doubt) others.

References

- Baron-Cohen, S., Leslie, A. M., and Frith, U. 1985: Does the autistic child have a 'theory of mind'? *Cognition*, 21, 37–46.
- Churchland, P. M. 1981: Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78, 67–90.
- Collingwood, R. G. 1946: *The Idea of History*. New York: Oxford University Press.
- Doyle, A. Conan 1894: The Musgrave Ritual. In *The Memoirs of Sherlock Holmes*. New York: Harper Bros.
- Grandy, R. 1973: Reference, meaning, and belief. *Journal of Philosophy*, 70, 439–52.
- Leslie, A. M. 1987: Pretense and representation: The origins of 'theory of mind'. *Psychological Review*, 94, 412–26.
- Morton, A. 1980: *Frames of Mind*. Oxford: Oxford University Press.
- Nozick, R. 1981: *Philosophical Explanations*. Cambridge, Mass.: Harvard University Press.
- Premack, D. and Woodruff, G. 1978: Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–26.
- Putnam, H. 1978: *Meaning and The Moral Sciences*. Boston: Routledge & Kegan Paul.
- Quine, W. V. O. 1960: *Word and Object*. Cambridge, Mass.: MIT Press.
- Schutz, A. 1962: *Collected Papers*. The Hague: Nijhoff.
- Schutz, A. 1967: *Phenomenology and the Social World*. Evanston, Ill.: Northwestern University Press.
- Searle, J. R. 1983: *Intentionality*. Cambridge: Cambridge University Press.
- Stich, S. 1983: *From Folk Psychology to Cognitive Science: The Case Against Belief*. 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.
- Von Wright, G. H. 1971: *Explanation and Understanding*. Ithaca: Cornell University Press.

READINGS IN MIND AND LANGUAGE

- 1 Understanding Vision: An Interdisciplinary Perspective
Edited by Glyn W. Humphreys
- 2 Consciousness: Psychological and Philosophical Essays
Edited by Martin Davies and Glyn W. Humphreys
- 3 Folk Psychology: The Theory of Mind Debate
Edited by Martin Davies and Tony Stone
- 4 Mental Simulation: Evaluations and Applications
Edited by Martin Davies and Tony Stone

Folk Psychology

The Theory of Mind Debate

Edited by

Martin Davies and Tony Stone


BLACKWELL
Oxford UK & Cambridge USA

First published 1995

Blackwell Publishers Ltd
108 Cowley Road
Oxford OX4 1JF
UK

Blackwell Publishers Inc.
238 Main Street
Cambridge, Massachusetts 02142
USA

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

British Library Cataloguing in Publication Data

A CIP catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data has been applied for.

ISBN 0-631-19514-9; ISBN 0-631-19515-7 (pbk.)

Typeset in 9¹/₂ on 11 pt Palatino
by Best-set Typesetter Ltd., Hong Kong
Printed in Great Britain by Hartmolls Ltd., Bodmin, Cornwall

This book is printed on acid-free paper.

Contents

List of Contributors

Acknowledgements

Introduction

MARTIN DAVIES AND TONY STONE

1 Replication and Functionalism
JANE HEAL

2 Folk Psychology as Simulation
ROBERT M. GORDON

3 Interpretation Psychologized
ALVIN I. GOLDMAN

4 The Simulation Theory: Objections and Misconceptions
ROBERT M. GORDON

5 Folk Psychology: Simulation or Tacit Theory?
STEPHEN STICH AND SHAUN NICHOLS

6 'He Thinks He Knows': And More Developmental Evidence
Against the Simulation (Role-taking) Theory
JOSEF PERNER AND DEBORRAH HOWES

7 Reply to Stich and Nichols
ROBERT M. GORDON

vii

ix

1

45

60

74

100

123

159

174