

FRANK JACKSON & PHILIP PETTIT

IN DEFENCE OF FOLK PSYCHOLOGY

(Received 14 October, 1988)

It turned out that there was no phlogiston, no caloric fluid, and no luminiferous ether. Might it turn out that there are no beliefs and desires? Patricia and Paul Churchland say yes.¹ We say no. In part one we give our positive argument for the existence of beliefs and desires, and in part two we offer a diagnosis of what has misled the Churchlands into holding that it might very well turn out that there are no beliefs and desires.

1. THE EXISTENCE OF BELIEFS AND DESIRES

1.1. *Our Strategy*

Eliminativists do not insist that it is certain as of now that there are no beliefs and desires. They insist that it might very well turn out that there are no beliefs and desires. Thus, in order to engage with their position, we need to provide a case for beliefs and desires which, in addition to being a strong one given what we now know, is one which is peculiarly unlikely to be undermined by future progress in neuroscience. Our first step towards providing such a case is to observe that the question of the existence of beliefs and desires as conceived in folk psychology can be divided into two questions.

There exist beliefs and desires if there exist creatures with states truly describable as states of believing that such-and-such or desiring that so-and-so. Our question, then, can be divided into two questions. First, what is it for a state to be truly describable as a belief or as a desire; what, that is, needs to be the case according to our folk conception of belief and desire for a state to be a belief or a desire? And, second, is what needs to be the case in fact the case? Accordingly, if we accepted a certain, simple behaviourist account of, say, our folk

conception of belief, we could dispose of eliminativism about belief by advancing the following argument:

PREMISE 1: It is sufficient for having the belief that snow is white that one be in a state which disposes one to utter the sentence 'Snow is white'. (Folk conception premise)

PREMISE 2: We are sometimes in a state which disposes one to utter the sentence 'Snow is white'. (Factual premise)

CONCLUSION: We have beliefs, including the belief that snow is white.

If the first premise were true, this argument would dispose of eliminativism about beliefs, for we know here and now that the factual premise is true. No doubt neuroscience is full of surprises, nevertheless we can be sure that it is not going to tell us that utterances are uncaused or that they are caused by the evil demon rather than by states inside us. The trouble with this reply to eliminativism is that the first, folk conception premise is false. It is not enough, according to our folk conception, in order for it to be the case that one believes that snow is white, that one be disposed to utter the sentence 'Snow is white'. A moment's reflection on liars and actors makes this clear. It is part of our folk conception that they are on occasion disposed to utter sentences while lacking the relevant beliefs.

The argument is though a suggestive failure. It suggests that we should look for a replacement for the first premise which plausibly does follow from our folk conception of belief and desire, provided that the corresponding replacement for the second, factual premise retains the needed property of being peculiarly unlikely to be refuted by progress in neuroscience. The failure suggests that we look for an argument of the following structure:

PREMISE 1: It is sufficient for having beliefs and desires that one be in states which satisfy (Folk conception premise)

PREMISE 2: We are sometimes in states which satisfy (Factual premise)

CONCLUSION: We have beliefs and desires.

The trick, or the tricky bit, is to find a uniform filling for the dots in the two premises which both makes the first one plausible as a reflection of our conception of belief and desire, and makes the second one sufficiently plausible given what we now know to render refutation by progress in neuroscience unlikely.

As one would expect, the two requirements tend to work against each other. We want something strong enough to give a condition sufficient for having beliefs and desires when inserted in the first premise, but not so strong that when inserted in the second premise we get something open to serious doubt either now or in the future. Nevertheless, it is, we will argue, possible to satisfy both requirements together. We will start by considering how to make premise one true.

1.2. *The Folk Conception of Belief and Desire*

Eliminativists emphasize that folk psychology is a *theory*. We agree entirely. Indeed it seems to us that Paul Churchland has provided an excellent, succinct account of the theory. He observes that

... the average person is able to explain, and even predict, the behaviour of other persons with a facility and success that is remarkable. Such explanations standardly make reference to the desires, beliefs, fears . . . to which agents are presumed subject.

... Each of us understands others, as well as we do, because we share a tacit command of an integrated body of lore concerning the law-like relations holding among external circumstances, internal states, and overt behavior . . . This approach entails that the semantics of the terms in our familiar mentalistic vocabulary is to be understood in the same manner as theoretical terms generally . . . [as] fixed by the network of laws in which [they] figure.²

This is a commonsense version of functionalism as applied most particularly to beliefs and desires. (And it is the application to beliefs and desires which we have in mind when we speak of folk psychology in this paper.) It is a version of functionalism because it approaches the semantics of belief and desire in terms of “the law-like relations holding among external circumstances, internal states, and overt behavior” — in short, to believe (or to desire) that so-and-so is to have a state in one playing the role definitive of that belief between inputs, outputs, and other functionally specified states. It is a commonsense version because the functional roles are given in terms of “a tacit command of an integrated body of lore . . .” The roles are, that is, ones we know about

here and now, rather than (as in psycho-functionalism) roles that future research in neuroscience or empirical psychology will reveal. They are folk roles, as we will later put it, rather than psycho-functional roles.³

A familiar complaint about the commonsense functionalist account of belief and desire (or of any psychological state, if it comes to that) is the amount of ‘handwaving’ involved. We are told that to have beliefs and desires is to have internal states playing certain commonsense functional roles without being told exactly which commonsense functional roles are definitive of which beliefs and desires. Stephen Schiffer has recently suggested that this failure to be more specific has a principled source.⁴ We simply do not know enough about the functional roles occupied by psychological states to provide a common lore adequate for commonsense functionalism. He suggests indeed that “perhaps no-one today is a commonsense functionalist”⁵ on the ground that in order for folk theory to yield a commonsense functionalism it needs “to contain at least three kinds of generalizations: those determinative of . . . internal functional roles; . . . perceptual input conditions . . . ; . . . and output conditions”,⁶ and he argues that in none of the three cases is the required generalization sufficiently part of common knowledge to be suitable for incorporation as part of our *folk* conception of beliefs and desires. Schiffer argues, for example, that the person in the street confronted with a candidate input condition like ‘if there is a red box directly in front of *x* and . . . , then *x* will believe that there is a red box in front of *x*’ will have very little idea of how to fill in the dots.

We think that Schiffer somewhat exaggerates the inability of the common person to come up with appropriate fillings, which may of course include *ceteris paribus* clauses, to complete the generalizations of commonsense functionalism. But in any case, what is important here is that there is *implicit* knowledge in addition to explicit knowledge about functional roles.⁷ Each of us knows *how* to move back and forth from behaviour in situations to beliefs and desires. It is a commonplace of belief-desire psychology that we interact successfully with our fellows by virtue of our ability to move from observations of someone behaving in a certain way in a certain situation to a hypothesis about her beliefs and desires, and then from that hypothesis to expectations about future behaviour in future situations. Hypotheses about beliefs and desires

play an essential middle role in these inferences. We handle tables and chairs well enough without belief-desire hypotheses, but we need belief-desire hypotheses in order to handle our fellow human beings. Now the ability to move back and forth from behaviour in situations to belief-desire hypotheses in successfully explaining and predicting behaviour shows that we have implicitly mastered, whether or not we have always explicitly noted, the needed generalizations between the inputs in the situations, the behavioural outputs, and the beliefs and desires. The alternative is to suppose that we have arrived at our predictions by chance. But then the success of our predictions is also chance — and that is incredible. We go wrong often enough, but nevertheless our successes are more than sufficient to make the chance hypothesis incredible.

A skilled racing-car driver has an enormous stock of implicit knowledge about how inputs of the turning-the-steering-wheel-clockwise-halfway-through-a-skid kind connect to outputs of the car-moves-to-inside-line-in-the-corner kind which he displays through his ability to get around race-tracks quickly; but unless he is a theoretician of the art of racing-car driving, he may know explicitly only the most rough-and-ready generalizations. In the same way, we are all highly skilled users of the generalizations of belief-desire psychology in the sense that we are able to move back and forth with great facility from behaviour in situations to beliefs and desires in our day-to-day traffic with people — think of what is involved in playing a game of tennis, or negotiating a roundabout — but only the theoreticians among us have much idea of how to formulate explicitly the generalizations our displayed abilities show that we have mastered implicitly.

We should emphasise that we are not basing our defence of commonsense functionalism on any highly controversial version of the doctrine of implicit or tacit knowledge: to what extent can an *ability* demonstrate knowledge *that* rather than knowledge *how*? Our point is simply that what people know is not limited to what they can write down on paper off the bat. The empirical fact is that we are able to predict each other's behaviour to a remarkable extent (remarkable indeed, when you compare it to what guessing would achieve). We do this by, in many cases, working via belief-desire hypotheses. What is essential to the success of this activity is the functional roles we

associate with belief and desire. These roles cannot be ones only the experts are privy to — because the ability to predict correctly is not an ability restricted to the experts — despite the fact that only the experts can write down the sentences which correctly and precisely describe the roles.

We are now in a position to fill in the dots in premise one. We agree with eliminativists that folk psychology is a theory, and that in particular our conception of belief and desire is given by the role they play in that theory. Some of the folk think that beliefs and desires are located in ‘ghost stuff’, some think that they are brain states, but most have no opinion one way or the other on the metaphysics of belief and desire. It is the role in the theory which is common and peculiar to the folk attitude towards beliefs and desires. The theory is a version of commonsense functionalism. The functional roles definitive of belief and desire in this version are given by those roles we (in part implicitly) presuppose in our predictive and explanatory practice in terms of belief-desire psychology. Thus, premise one becomes ‘It is sufficient for having beliefs and desires that one be in states which satisfy the functional roles embodied in our everyday practice of predicting and explaining human behaviour (for short, the folk roles)’. And our overall argument for the existence of beliefs and desires becomes:

PREMISE 1: It is sufficient for having beliefs and desires that one be in states which satisfy the folk roles. (Folk conception premise)

PREMISE 2: We are sometimes in states which satisfy the folk roles. (Factual premise)

CONCLUSION: We have beliefs and desires.

1.3. *Why Premise One Should Be Non-controversial*

We seek to make the existence of beliefs and desires as non-controversial as possible (to be realistic, as non-controversial as possible by philosophical standards). Hence we seek non-controversial premises. How can premise one be non-controversial given that it rests on a theory as controversial as commonsense functionalism? The answer is

that it rests on the relatively non-controversial parts of commonsense functionalism. It is notorious, for instance, that functionalism has trouble accounting for qualia, the sensuous side of psychology. However beliefs and desires do not have qualia — or if they do, they are not essential to the states being beliefs and desires.⁸

A second hard question for commonsense functionalism is what to say about creatures with brains very like ours but which are located in very different environments — in the extreme are brains in vats. According to many, they have beliefs and desires like ours, and yet their states do not occupy the same *commonsense* functional roles in any obvious sense, though they do occupy the same psycho-functional roles in some accounts of what a psycho-functional role is.⁹ Their internal states play the same roles between inputs and outputs at the surface of the brain, but do not play the same roles between distal surroundings and motor behaviour. And it is of course the latter roles which count as the folk roles in our account of commonsense functionalism, for it is our mastery of the latter roles which underlies our folk predictive practice in using belief-desire psychology. However, the issue brains in vats raise is whether filling the folk roles is strictly *necessary* for having beliefs and desires — does the brain in a vat have beliefs and desires despite not having states playing the kind of roles that we all know about and use in our folk practice? And premise one is a claim about what is *sufficient* for having beliefs and desires. Provided that having states playing the folk roles is enough for having beliefs and desires, premise one is safe.

A third controversy is over whether commonsense functionalism should be rejected on the ground that it is in the grip of the bad, old-fashioned description theory of meaning. Commonsense functionalism as standardly developed is indeed a description theory of the meaning of mental state terms. This is explicit in the Ramsey sentence formulation of it.¹⁰ However, the theory can instead be developed as a theory about reference fixing, at least as far as its role in our argument is concerned.¹¹ The specification of the functional roles would then be viewed, not as giving the meanings of the mental state terms, but as giving the way the reference of these terms is fixed. Roughly, the story would go as follows. A believer and desirer has in him or her a certain highly complex set of neurophysiological states which both occupy

certain folk roles and certain psycho-functional roles. 'Belief' and 'desire' have their reference fixed in our world by the fact that particular folk roles, FR, are filled. The references so fixed are not to the neurophysiological occupants of the folk roles (that would wrongly make it automatic that creatures with different neurophysiologies could not be alike in being believers and desirers), but are to the psycho-functional roles, PF, associated with FR in our world. In other possible worlds a believer and desirer is a creature with states occupying PF, regardless of whether the states occupying those roles are the same as the states which occupy them in our world, and regardless of whether they occupy FR in that other possible world.

The important point for us is that this change in the status of commonsense functionalism would not affect the role of premise one in our argument. We need premise one to be a truth in consequence of our folk conception, in some broad sense, of belief and desire. It does not matter whether the details of the story are told in terms of the commonsense functional roles giving the descriptive meanings, or in terms of those roles somehow fixing the references, of 'belief' and 'desire'. Either way premise one comes out true a priori, and that is enough for us.

Finally there is the controversy over content. It is part of our folk conception that beliefs are true or false, and that desires are satisfied or unsatisfied. But a belief that such-and-such is true precisely if such-and-such is the case, and a desire that so-and-so is satisfied precisely if so-and-so is the case. It is thus part of our folk conception that beliefs are beliefs that such-and-such, and desires are desires that so-and-so; that, as we say, beliefs and desires have contents. Hence if we are right that, according to the folk conception, satisfying the folk roles is sufficient for having beliefs and desires, then satisfying the folk roles is sufficient for having contents. But is it the case that satisfying a certain commonsense functional role is sufficient for having content?

It is important to distinguish three possible positions on the connection between functional role and content. The position that ties content most closely to functional role holds that content supervenes on functional role.¹² If two subjects behave in exactly the same way in every possible situation then their beliefs and desires have exactly the same contents. That is to say, if they would never behave differently in any

situation whatsoever, then they believe and desire exactly alike. This is consistent with allowing a degree of indeterminacy to content provided it is the same degree for each behaviourally identical subject. It is clear that if this is the right position on content then content has not been left out of our story. Once God settled the functional roles, He settled the contents (up to whatever indeterminacy content ascriptions have). There was no more to be done. If Jerry Fodor is right, God will have found it very hard to get the functional roles right without using linguistically structured entities as the occupants of the roles.¹³ Nevertheless, it is the functional roles per se which are sufficient. Of course we have said nothing about how to pair off functional roles with contents, nothing about which functional roles determine which contents. We could hardly afford to and still keep premise one non-controversial. But all that is required of us is that content be included in our common-sense functional story in the sense that it be certain that there are contents that our beliefs and desires have; not that there are particular contents which it is certain that our beliefs and desires have. We are defending the claim that it is virtually certain that we have beliefs and desires, not that it is virtually certain that we have the belief that *P* or the desire that *Q*, for some particular *P* and *Q*.

The second possible position holds that which particular contents a being's beliefs and desires have depends on a *combination* of functional roles and environmental matters (such matters as whether it was water or swater that played a certain causal role in the development of language, the particular linguistic practices of one's speech community or the details of the evolutionary history of one's species).¹⁴ Nevertheless, according to this position, one's internal states filling appropriate functional roles is sufficient for one's beliefs and desires having contents, despite not being sufficient for their having exactly the contents they do have. Keep the functional roles of my beliefs and desires fixed while varying the relevant environmental matters, and you may change the contents of my beliefs and desires, or at least you may change the right sentential characterizations of their contents, but you cannot remove content altogether; you cannot make my beliefs and desires contentless. Again, this position is no threat to premise one. Functional role ensures content possession in and of itself, although it does not ensure the possession of any particular content.

The final possible position holds that more than functional role is essential for even the possession of content. One view might be that having an evolutionary history of a certain kind is necessary not just for my beliefs having the particular contents that they in fact have, but for their having any content at all.¹⁵ If Fred was the product of an immensely improbable coincidental coming together of molecules, then even if his states play the same functional roles as mine, they have no contents whatsoever. We find this more extreme position very implausible — it seems to us to be too like the mistaken view in Aesthetics that *Macbeth* would have been a bad play had it been written by a monkey playing on a typewriter. But the position could in any case be accommodated within our defence of folk psychology. Replace premise one by, ‘It is sufficient for having beliefs and desires that one be in states which satisfy the folk roles and have such-and-such an evolutionary history’, and make the corresponding change in premise two. Progress in neuroscience is not going to refute the theory of evolution; hence, even if the theory of evolution does enter our folk conception of what it is to have beliefs and desires indirectly via what is needed to possess content, this does not threaten the existence of beliefs and desires.

Premise one then is relatively non-controversial, because, first, commonsense or folk functionalism — the theory we took from the eliminativists themselves — is a plausible gloss on the folk theory of beliefs and desires, and, second, the role it plays in premise one does not (as we have just seen) require a stance on the controversies that surround commonsense functionalism.

1.4. *Defence of Premise Two*

Premise two says that we are sometimes in states which satisfy the commonsense roles. Given our explanation of what this claim amounts to, is premise two plausible? More to the point, is it sufficiently plausible for us to be confident *ahead of what neuroscience may turn up* that we are sometimes in states which satisfy the folk roles? The answer may seem to be obviously no. For instance, is it not part of the commonsense view about belief’s role, that a belief typically causes both verbal and motor behaviours, and so that the same state causes

both?¹⁶ And that view is surely highly vulnerable to possible refutation by progress in neuroscience. But this view is not part of our understanding of the folk roles of belief in the sense just explained. It is not part of the knowledge, explicit or implicit, that we display when we move back and forth between situations, behaviour, beliefs, and desires. What we display is knowledge to the effect, say, that often when someone believes that coffee is available, they will both speak in a certain way and act in a certain way. It does not matter for the success of our passage back and forth between situations, behaviour, beliefs, and desires how many states inside the agent are required to work the trick. Perhaps we folk suppose that there is only one state, but this supposition plays no essential role in how we predict behaviour from information about beliefs, desires, and situations. The unitary assumption may be a bit of folk supposition, one which will perhaps turn out to be a bit of folk mythology, but it is not part and parcel of the folk theory, described so well by Paul Churchland, that forms the core of the folk conception of belief and desire. That is to say, our question about premise two can be put this way. Our folk practice determines a hypothesis about how we are functionally organized — the hypothesis we follow, whether explicitly or implicitly, in moving between behaviour, situations, beliefs, and desires; the hypothesis that we have states filling the folk roles. Do we have good reason to accept this hypothesis as capturing, near enough, the relevant part of how we are in fact functionally organized, and to be confident that progress in neuroscience will not destroy this good reason?

It is clear that we have as of now good reason to accept the hypothesis about how we are functionally organized; following it yields a great many successful predictions — we bet our lives on it every time we drive a motor car — and successful predictions constitute good evidence. But what justification is there for taking the hypothesis — call it the folk hypothesis, for it is the hypothesis which underlies the folk practice, the hypothesis that the folk roles are filled — to be *peculiarly* epistemically secure? The answer is that it is entirely in terms of the functional roles played by internal states. If the folk hypothesis is wrong, it is wrong by having the functional roles wrong. There is no other way for it to be wrong. But having the functional roles wrong is a very hard matter to conceal precisely because they are functional roles.

The hypothesis that metals are good conductors of heat and electricity was established much earlier and much more securely than the explanation of the fact in terms of the atomic theory, and this was precisely because being a good conductor is functionally defined in terms of playing a certain role between observable inputs and outputs. Provided, but only provided, that we insist that the folk hypothesis be understood in purely functional terms, we can regard it as peculiarly well-confirmed. By being non-committal about the nature of what realizes the functional roles, it minimizes the extent to which it goes beyond its observational base, and so is made peculiarly secure by that observational base.¹⁷

It might be objected that although the folk hypothesis minimizes the extent to which it goes beyond its observational base by being purely functional, nevertheless functional hypotheses do go beyond their observational bases. They hypothesize *internal* causal links and concern *hypothetical* as well as actual inputs and outputs. How then can it be that there is so little room for doubt? Well, suppose that I punch 4 followed by 7 on my pocket calculator, and that 11 then appears on the screen. I discard the chance hypothesis and adopt one which includes the clause: my calculator stores for subsequent utilization the number n when n is punched. This is a functional hypothesis that goes beyond its observational base both in respect to internal causal links and in respect to hypothetical inputs and outputs. Nevertheless, after a relatively few trials, the hypothesis would become overwhelmingly credible, although there might still be much doubt about *how* the calculator does the storing. (And note that we would still be entitled to great confidence in certain functional hypotheses even if the calculator malfunctioned every so often. We could be confident that numbers were being stored on many occasions even if it became obvious on certain other occasions that either they were not being stored or they were but not to any good subsequent effect.) Of course the folk hypothesis is much more complex than this simple one about the storing of numbers for subsequent utilization; but it is also true that it has been put to the test very much more often.

We are all familiar with the idea that postulating entities of kind K may be extremely fruitful for predictive purposes, and yet there are, or well might be, no K s. Ptolemy's epicycles were, and a system of particles' centre of gravity is, predictively useful postulates, and yet

neither epicycles nor centres of gravity exist. However, in both cases the postulates are *more* than 'that which' fills the appropriate functional roles. They are that which fill the roles and are of *such-and-such a nature*, and the doubt about existence turns on this latter fact. It is beyond question that the functional roles are filled, but as we now know not by epicycles, and as we always knew not by centres of gravity. Had Ptolemy been content to say merely that the planets' orbits were of a nature which explained the astronomical observations without saying what the nature was, he would not have been refuted by Copernicus and Kepler. (Also, of course, he would have been saying something glaringly obvious — like, if we are right, the existence of beliefs and desires.) This point explains why an instrumentalism about belief and desire suggested in some of Dan Dennett's writings seems to us not to be a live option.¹⁸ To behave *as if* we had beliefs and desires is to provide overwhelming evidence that we *in fact* have beliefs and desires, precisely because to have beliefs and desires is to be understood purely functionally. Of course it is logically possible for an organism to behave as if it had beliefs and desires without actually having beliefs and desires. Consider an organism which is nothing more than a receiver for behavioural instructions from Mars, and which is controlled by a Martian to behave exactly as you do in any given situation. It behaves as if it had beliefs and desires without having them (though no doubt the Martian does). But how likely is it that a hypothesis anything like this is true of one of us?¹⁹

This completes the case for premise two. The folk hypothesis is a well-confirmed purely functional theory, and well-confirmed purely functional theories are essentially competing against only the incredible chance hypothesis. We have earlier seen the case for premise one, so our positive case for the existence of beliefs and desires is now before you. It is time to discuss the main eliminativist argument — the argument from predictive imperfection.

2. THE ARGUMENT FROM PREDICTIVE IMPERFECTION FOR ELIMINATIVISM

2.1. *Preliminary Statement of the Argument and Preliminary Reply*

The history of science is full of theories which had for a time consider-

able currency because of their predictive value, but which were subsequently abandoned when new theories were developed which had a superior predictive value. Sometimes the old theory was shown to be incomplete. It yielded no prediction about a matter that the new theory got right. Sometimes the old theory yielded a prediction, but one that turned out to be mistaken. Most times the old theory exhibited both kinds of predictive failure.

Folk psychology exhibits both kinds of failure: it makes wrong predictions, and is conspicuously silent about a whole host of facts about behaviour. Completed neuroscience by contrast will, when suitably supplemented by purely physical information about an agent's environment, never make a mistake and never be stumped for an answer about what an agent will do.²⁰ The eliminativist conclusion is then the apparently inevitable one that completed neuroscience refutes, or will refute, folk psychology by superseding it in the same general way that Copernicus and Kepler refuted Ptolemy, Einstein refuted Newton, the oxidation theory refuted the phlogiston theory, and so on and so forth. Folk psychology may be left with instrumental value, or perhaps with approximate truth, but not with truth itself.

We think that it is a mistake to reply to this argument by pointing out that we have not yet attained, and perhaps never will attain, the completed neuroscience the argument speaks of. We are far enough along the road to make it plausible that it is there to be attained if only we are clever enough, and that is all the argument really needs. Ptolemy was refuted once it became clear that there was a superior theory to be had. (This is in effect what Copernicus did. As has often been observed, a theory superior in predictive power was not actually achieved until Kepler corrected Copernicus' mistake about the nature of the planets' orbits.)

Our reply to the argument from predictive imperfection starts from the point that not all cases of scientific progress involve radically falsifying what has gone before. Our tribal forebears no doubt noticed that it got colder the higher up one climbed a mountain, and we may suppose they developed a very simple theory of climate: *A*'s being higher than *B* explains *A*'s having a colder climate than *B*. This theory exhibits all kinds of predictive failures and has now been superseded by one that mentions, in addition to altitude, distance from the equator,

prevailing winds and sea currents, and so on and so forth. But it would of course be quite wrong to infer from this that it is never literally true that one city is at a greater altitude than another, or to infer that this fact is never part of the explanation of why one city is colder than another.

The inference would be mistaken because to say that the climate of a city involves all of altitude, distance from the equator, prevailing air and sea currents, and so on is to say *inter alia* that it involves altitude. That is, the new theory contains the relevant part of the old one. The same will be true of completed neuroscience. Completed neuroscience will contain folk psychology, or at least this is by far the most reasonable stance to take given our present state of knowledge. For completed neuroscience, along with enough environmental information, will tell us all about the functional roles of our internal states. It will tell us all about how the various states interact with each other, what does and would cause them, and what they do and would cause, and enough information of this kind gives us which states play which functional roles. And it is virtually certain that some of these functional roles will be those definitive of beliefs and desires, because as we have recently seen, premise two is virtually certain. As we saw, we *know* that the folk roles are occupied, and so, as completed neuroscience will tell us all about which roles are occupied by which states, it will tell us that the folk roles are occupied.

When we say that completed neuroscience will contain the functional information, we do not of course mean that it will contain the information in any particularly transparent or easily usable form. If I tell you that the corners of a figure are located at coordinates (1, 1.4), (1.4, 2), (2, 1.6), and (1.6, 1), relative to some orthogonal axes, I tell you in one sense that the figure is a square, for what I have told you entails that it is a square. Its being a square supervenes on what I have said. In another sense I do not, for I left the matter opaque. The sense in which completed neuroscience will tell us about functional roles is the second, only more so. Extracting the relevant functional information may be a Herculean task, but all that our reply to the argument from predictive imperfection needs is that the information is in there, however buried and however hard to extract. For, provided the information is in there, completed neuroscience supersedes folk psychology in the sense of

containing it, along with a great deal else besides, and not in the sense of refuting it.

We expect three replies to our reply to the argument from predictive imperfection: first, that we are ignoring a principal teaching from the history of science; second, that our reply leaves untouched the fundamental point that completed neuroscience, or rather its possibility, shows that we do not ultimately need beliefs and desires in order to predict behaviour, and what we do not need we should not believe in; and third, that wheeling in functionalism in order to save old theories from refutation by new theories is thoroughly retrograde and typical of degenerating research programs.

2.2. *Objection From the History of Science*

We used a rather bald statement of the argument from predictive imperfection. Eliminativists do allow that in one kind of case a new theory with superior predictive power can leave intact and in place much of the theoretical apparatus of the old theory.²¹ The case is where a reductive identification of the old in terms of the new is possible. The classic example is the reduction of the thermodynamic theory of gases to the kinetic theory of gases. The kinetic theory's superior explanations in terms of molecular kinetic energy and molecular momentum were not taken to show that the thermodynamic theory's inferior explanations in terms of temperature and pressure were explanations in terms of the non-existent. Instead, temperature and pressure were identified with molecular kinetic energy and molecular momentum properties, and by being so identified were retained rather than discarded.

Their claim is that the history of science teaches us that this exception to the displacement of the old by the new is the *only* exception. Hence, completed neuroscience will leave folk psychology's explanatory posits — beliefs and desires, for particular instance — intact only if a reductive identification of belief and desire in terms of the fundamental explanatory notions of completed neuroscience is possible; and, they go on to observe, it is surely at least open to question whether this will be possible. We agree that it is far from certain that reductive identifications of the fundamental properties of folk psychology — believing that so-and-so, and desiring that such-and-such — with the fundamental

explanatory properties in completed neuroscience will be possible. Surely it is very much an open possibility that the two theories will organize and taxonomise our internal states quite differently in framing the principles which generate their behavioural predictions.²² Our disagreement is with the claim that the history of science teaches us that reductive identifications of basic properties are the only way out of elimination of the old by the superior new. The kinetic theory itself shows us a counterexample to this principle.

The kinetic theory of gases can be viewed as a staging post on the road to the *super* kinetic theory of gases. The kinetic theory is framed in statistical mechanical terms and in terms of macro-properties of collections of gas molecules; the super kinetic theory is framed in terms of the exact position, mass, velocity, and size of every gas molecule, and its laws are those deterministic ones (let's suppose) governing molecular interactions. Statistical considerations do not enter into the calculations. They do not need to. We have the full, fine-grained information in the super kinetic theory. And for this reason the super kinetic theory will be an even better predictor of gases' behaviour than the kinetic theory. For instance, the super kinetic theory will get the rate of diffusion of a gas from an open flask *exactly* right, by comparison with the kinetic theory's statistical prediction about the rate.

The super kinetic theory will never actually be at our disposal. The calculations are far, far too complex. The important point for us, though, is the following: were the super kinetic theory by some miracle to be at our disposal, that would be no reason at all to deny that gases have temperature and pressure. (If you disagree, you should here and now deny that gases have temperature, pressure, and volume properties, because we all know here and now that the super kinetic theory is there to be had in principle.) And yet no reductive identifications of temperature and pressure with the fundamental properties of the super kinetic theory are possible. Temperature, for instance, is not the mass, velocity, or position of any individual molecule. There will be no isomorphism even by the most relaxed standards between the gas laws framed in terms of temperature, pressure, and volume and the laws framed in terms of mass, position, and velocity of molecules of the super kinetic theory. The basic taxonomic principles of the two theories are very different.

Why doesn't the super kinetic theory show that there is no such thing

as the temperature and the pressure of a gas? The reason is that though temperature, for instance, is not identified with any fundamental category of the super kinetic theory, the information contained in the super kinetic theory enables us to work out the temperature. Knowledge of the mass and velocity of each and every particle along with the relevant laws of particle interactions enables (in principle) the calculation of a gas's temperature.²³ The story is essentially similar to the one we told about the relationship between completed neuroscience and folk psychology. What matters most is not the existence of reductive identifications of the fundamental properties in the old theory's laws with the fundamental properties in the new theory; not the possibility, as it is sometimes put, of smooth reduction from old to new; but, instead, whether the new theory contains information that supports the relevant part of the old theory.

Discussions of eliminativism have been dominated by the question as to whether neuroscience (or cognitive science, or . . .) will come up with a theory which is comprehensively successful in explaining and predicting human behaviour by means of laws framed in quite different terms from those of folk psychology. The idea is that this would show that there were no beliefs and desires (though pretending that there were could still have a certain predictive value). The example of the super kinetic theory shows that this does not follow. The super kinetic theory's laws are framed in terms quite different from those of the thermodynamic theory; nevertheless, gases really do have temperature and pressure, just as we really do have beliefs and desires.

2.3. *The Objection From Need*

There is a line of thought about ontology which can be paraphrased roughly as follows: to decide what one ought to believe in, look to what is needed for science. Don't waste time on conceptual analysis. Philosophy's role in ontology is simply the discerning of what science requires for the prediction and explanation of phenomena.²⁴

This position is sometimes referred to as philosophical naturalism. We are sure that it lies behind much of the doubt naturalistically oriented philosophers have about beliefs and desires. But why should philosophical naturalism be thought to threaten the existence of beliefs

and desires? The whole thrust of our discussion of premise two was that we have excellent reason to believe that the neurophysiological states we need to explain behaviour occupy the functional roles definitive of beliefs and desires. Hence, beliefs and desires *are* required to explain behaviour. It would be a mistake to object here that all this shows is the need for the states which are beliefs and desires, rather than the need to suppose that they have the *properties* of being beliefs and desires. The states having the properties of being beliefs and desires is nothing more than their filling the appropriate roles, and we certainly need to suppose that they do that. For if they did not have these properties, subjects would behave *very* differently.²⁵

The concern of the naturalists is that in order to *see* the need for beliefs and desires, we had to indulge in some conceptual analysis. The science per se will tell us which roles are occupied by which neurophysiological states. The fact that this constitutes having beliefs and desires required the defence of commonsense functionalism, and so some conceptual analysis. We had to say something about how we conceive of beliefs and desires. It was, however, a very ontologically modest bit of conceptual analysis. We did not use our account of the folk conception of beliefs and desires as part of an argument that completed neuroscience is incomplete, but instead as part of an argument that beliefs and desires will very likely be found *within* what completed neuroscience tells us. The role of our account of the folk conception of beliefs and desires was simply to show which part of any likely complete neuroscience story is the part which says (though not in so many words) that there are beliefs and desires.

Whatever role you think conceptual analysis has in general in ontology, it is hard not to allow it at least this modest role — the role of helping us with what to say about some posit or other of an old theory when the new theory comes along. Even Patricia Churchland implicitly allows this much to conceptual analysis in her discussion of the puzzle for Aristotle's mechanics arising from the fact that a projectile does not drop straight down when released from a catapult. She observes that medieval physicists explained this fact in terms of a special force called *impetus*, and that Newton's theory shows that there is no such thing as *impetus*. Uniform motion, as opposed to acceleration, does not need a force to sustain it. As she says, "There is simply nothing in classical

mechanics with which to identify impetus. ‘Momentum’ may suggest itself, but momentum is not a *force*, whereas *being a force was the whole point of impetus*”.²⁶

We think that this is exactly the right way to look at the matter. Saying that being a force was the whole point of impetus is another way of saying that impetus as conceived in the medieval theory was a force. Her methodology here is essentially the same as ours in the folk psychology case. Take the new theory to have the truth of the matter — or at least more of it than the old theory. The question of the fate of a particular posit of the old theory is then decided by considering how the posit was conceived in the old theory and seeing whether something satisfying that conception nearly enough is to be found in the new theory. If the answer is yes, as we argued it was for beliefs and desires, the old posit survives; if the answer is no, as it is for impetus, phlogiston, caloric fluid, et al., the old posit is eliminated. (And, of course, there may be no determinate answer one way or the other in some cases.)

2.4. *Objection From the Retrograde Nature of Functionalism*

Paul Churchland sees what he calls “the functionalist stratagem” as potentially “reactionary, obfuscatory, retrograde, and wrong”. His objection is that functionalism can always be invoked to save old theories from more than deserved elimination. He argues, for example, that a “cracking good defense of the phlogiston theory of combustion can . . . be constructed along [functionalist] lines” by construing “being highly phlogisticated and being dephlogisticated as functional states defined by certain syndromes of causal dispositions . . .”²⁷

Paul Churchland is right that it would have been wrong to defend the phlogiston theory of combustion in this way. But the reason it would have been wrong is that being highly phlogisticated and being dephlogisticated were *not* merely functional states in the theory. It was an essential feature of the phlogiston theory that ‘phlogiston’ in part meant — perhaps in the traditional sense, perhaps in the reference-fixing sense — ‘stuff given *off* by the object during combustion’. That is why the discovery that combustion involved stuff coming *in* from outside the object refuted the theory. The moral to be drawn from the

example is not that it is bad to give functionalist definitions per se, but that it is wrong to try to save a theory by *redefining* its terms after the event. This however is not a charge that can be made against our defence of folk psychology. Our defence was in terms of commonsense functionalism, and that is a theory about what we folk do in fact mean by 'belief' and 'desire', as displayed in the use we make of beliefs and desires to explain and predict behaviour. It is not a theory about what it would have been nice to have meant if we wished to avoid refutation by progress in neuroscience.

What we do learn from Paul Churchland's example is that it is a mistake to try to save folk psychology by invoking a purely posterior form of psycho-functionalism. It is a mistake to say *nothing* about the functional roles definitive of beliefs and desires according to the folk conception in advance of what science tells us about the functional roles which are in fact occupied by the states inside us. For that amounts to saying that the roles which need to be filled in order for there to be beliefs and desires are those which will turn out to be filled. But then no amount of talk about demise of the analytic-synthetic distinction, open texture, the vagueness of the boundary between matters of meaning and matters of fact, and so on and so forth can conceal the fact that the hypothesis that there exist beliefs and desires is being made irrefutable — a matter Paul Churchland would indeed be entitled to complain about.

3. CONCLUSION

Our argument has been that a commonsense functionalist approach to our folk conception of beliefs and desires shows that it is *very* likely that they exist, where commonsense functionalism is understood as implicitly defined by our folk practice in moving back and forth between behaviour, situations, and beliefs and desires. Completed neuroscience will indeed provide a complete story about when and why we do what we do, but will incorporate rather than eliminate beliefs and desires in this complete story. The irony is that our defence uses an account of folk psychology fully in accord with that provided by eliminativism's sympathizers when they insist that folk psychology is a theory. They see this insistence as opening the way for serious con-

sideration of the possibility that folk psychology is radically mistaken. Any *theory* can be radically mistaken. But, of course, folk psychology is radically mistaken for a great many objects — the Taj Mahal, for instance. The Taj Mahal does not have beliefs and desires precisely because it does not satisfy the theory. Our point is that because the theory is a purely functional theory, the evidence that we satisfy it (and for that matter that the Taj Mahal does not) is peculiarly strong evidence.²⁸

NOTES

¹ In a number of places. We will be principally referring to Paul M. Churchland, "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy*, 78, 2, 1981, pp. 67–90; "Folk Psychology and the Explanation of Behaviour", *Proceedings of the Aristotelian Society*, supp. vol. 62, 1988, pp. 209–21; and Patricia Churchland, *Neurophilosophy*, Cambridge, MIT Press, 1986. See also Stephen Stich, *From Folk Psychology to Cognitive Science*, Cambridge, MIT Press, 1983.

² "Eliminative Materialism and the Propositional Attitudes", *op. cit.*, pp. 68–9. For similar sentiments about the theoretical nature of folk psychology see Patricia Churchland, *op. cit.*; Stich, *op. cit.*; and Wilfrid Sellars, "Empiricism and the Philosophy of Mind", in *Science, Perception and Reality*, London, Routledge & Kegan Paul, 1963.

³ For more on psycho-functionalism see, e.g., Ned Block, "Troubles with Functionalism", in *Readings in Philosophy of Psychology*, vol. 1, Block, ed. London, Methuen, 1980. For more on commonsense functionalism, see below; Block, *ibid.*; and David Lewis, "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy*, 50, 3, 1972, pp. 249–58. The distinction between commonsense functionalism and psycho-functionalism cuts across that between narrow and broad functionalism, though, given the importance of what goes on in the brain, psycho-functionalists often hold that as a matter of fact the psycho-functional roles of interest will concern what goes on inside the brain. We set aside for the moment the complications arising from the problem of specifying content, and we set aside entirely the question of whether it is best for theory to talk of a person's system of belief and system of preference rather than of their beliefs and desires.

⁴ In *Remnants of Meaning*, Cambridge, MIT Press, 1987, ch. 2.

⁵ *Op. cit.*, p. 28.

⁶ *Op. cit.*, p. 30.

⁷ Or tacit, as Paul Churchland put it above.

⁸ This view is widely, though not of course universally, accepted, and we simply take it as read here.

⁹ See Block, *op. cit.*

¹⁰ See Lewis, "Psychophysical and Theoretical Identifications", *op. cit.*

¹¹ In the sense of Saul Kripke, *Naming and Necessity*, Oxford, Blackwell, 1980; see, e.g., p. 55.

¹² This is the position called individualistic functionalism in the postscript to "Radical Interpretation" in David Lewis, *Philosophical Papers*, vol. 1, Oxford, Oxford University Press, 1983. It is close to Lewis's own position in "Radical Interpretation". See also

Robert Stalnaker's treatment of content in "Propositions", in *Issues in the Philosophy of Language*, Alfred Mackay and Daniel Merrill, eds., New Haven, Yale University Press, 1976, and *Inquiry*, Cambridge, MIT Press, 1984, where causal origin is brought into the picture to help separate out belief from desire.

¹³ See, e.g., Jerry Fodor, "Propositional Attitudes", in *Representations*, Harvester Press, 1981. At times though Fodor seems to be pressing for the stronger conclusion that it is part of the folk *conception* that beliefs and desires are inner sentences in some significant sense.

¹⁴ The first is discussed in the literature spawned by Hilary Putnam, "The Meaning of 'Meaning'", in *Mind, Language and Reality*, Cambridge, Cambridge University Press, 1975; see, e.g., the papers in *Subject, Thought, and Context*, Philip Pettit and John McDowell, eds., Oxford, Oxford University Press, 1986; the second is particularly emphasised in Tyler Burge, "Individualism and the Mental", in *Midwest Studies in Philosophy*, vol. IV, P. A. French et al., eds., Minneapolis, University of Minnesota Press, 1979; and the third in, e.g., Ruth Millikan, "Thoughts without Laws; Cognitive Science with Content", *Philosophical Review*, 95, 1986, pp. 47–79. We here presuppose a taxonomy of functional roles according to which the states of me and of my *doppelgänger ipso facto* fill the same functional roles. Thus, for example, the fact that what actually causes one state is, say, water and not swater does not figure in the typing of the role filled by that state. Other taxonomies are possible.

¹⁵ This seems to be Millikan's position, loc. cit.

¹⁶ See Stich, op. cit., p. 231.

¹⁷ Contrariwise, if you think that it really is part of the folk *conception*, and not merely an empirically plausible hypothesis, that the functional roles are occupied by entities of a certain kind — linguistically structured ones perhaps — anyone more modest than Hegel must concede to eliminativism at least the possibility that neuroscience may show that there are no beliefs and desires. At times Paul Churchland takes it to be part of the folk conception that the roles are occupied by linguistically structured entities, but see "Folk Psychology and the Explanation of Behaviour", op. cit., esp. p. 210, for his most recent thoughts on this. In our view the relation between the language of thought hypothesis and our folk practice of moving between behaviour, situations, and beliefs and desires can be at most that the success of the latter makes the former a plausible empirical hypothesis.

¹⁸ See, e.g., "Intentional Systems" and the introduction in Dan Dennett, *Brainstorms*, Sussex, Harvester, 1978. Note that we say 'suggested', not 'definitely endorsed'.

¹⁹ We borrow the idea for this example from the final chapter of Christopher Peacocke, *Sense and Content*, Oxford, Oxford University Press, 1983. Note that the receiver organism's states do not play the functional roles definitive of beliefs and desires. A different but equally compelling kind of example is described in Ned Block, "Psychologism and Behaviorism", *Philosophical Review*, XC, 1981, pp. 5–43. But, as Block himself notes, the probability that his kind of example is in fact realised is vanishingly small because it would involve a combinatorial explosion.

²⁰ We presuppose this without further ado on the grounds that it is plausible and is common ground between us and the eliminativists.

²¹ See particularly Patricia Churchland, *Neurophilosophy*, op. cit., § 7.2, and the references therein.

²² See Paul Churchland, "Eliminative Materialism and the Propositional Attitudes", op. cit., § V.

²³ Exactly how to *describe* the calculation depends on whether temperature's functional role gives the meaning or fixes the reference of 'temperature'. But as our point goes through on either view, we need not enter this debate.

²⁴ We are indebted to Kim Sterelny and Michael Devitt for convincing us of the need

to address this position. It informs Devitt's approach to ontological questions in *Realism and Truth*, Oxford, Blackwell, 1984. He is not, though, an eliminativist about beliefs and desires.

²⁵ What is true, in our view, is that these properties are not causally efficacious and yet are causally explanatory. In this respect they are like dispositional properties, while being of greater explanatory significance. For details see Frank Jackson and Philip Pettit, "Functionalism and Broad Content", *Mind*, XCVII, 1988, pp. 381–400.

²⁶ *Neurophilosophy*, op. cit., p. 289, last emphasis ours. Similarly, she argues, op. cit., pp. 287–8, that mass was eliminated by the theory of relativity by arguing, first, that our conception of mass is that of a property which plays a certain role in mechanics, and, second, that the theory of relativity shows that it is instead a relation which plays this role. This is precisely the methodology we espouse for beliefs and desires. (Of course, you may doubt the conclusion that mass was eliminated, while accepting the methodology, because you may suspect that our folk conception of a property playing the mass role employs the wide notion of property which includes relations.)

²⁷ Paul Churchland, "Eliminative Materialism and the Propositional Attitudes", op. cit., p. 81.

²⁸ We are much indebted to comments from Akeel Bilgrami, David Lewis, Michael Smith, and Kim Sterelny, and to very helpful discussions from a number of audiences.

*Research School of Social Sciences,
Australian National University,
Canberra ACT 2601,
Australia.*