

*Mental Causation*

*Edited by*  
John Heil and Alfred Mele

CLARENDON PRESS · OXFORD

1993

## Some Content is Narrow

FRANK JACKSON AND PHILIP PETTIT

ONE way to defend narrow content is to produce a sentence of the form 'S believes that *P*', and show that this sentence is true of *S* if and only if it is true of any duplicate from the skin in, any doppelganger, of *S*. Notoriously, this is hard to do. Twin Earth examples are pervasive.<sup>1</sup> Another way to defend narrow content is to show that only a narrow notion can play the causal explanatory role we require of content in a properly scientific psychology or cognitive science. Notoriously, this is hard to do. The considerations—methodological solipsism, the principle of autonomy, or whatever—invoked to show that a broad notion of content cannot play the required causal explanatory role are open to serious objection.<sup>2</sup> Moreover, this approach is not an argument for the existence of narrow content as such. It is an argument that content had better be narrow.

In this paper we offer a defence of narrow content which makes no (well, almost no) reference to Twin Earth examples or to contentious doctrines about which ways of taxonomizing mental states are right for a scientific psychology. We would like to think that our argument will be found relatively non-controversial. We see it as simply drawing out and making explicit the commitment to narrow content implicit in doctrines about making sense of human

The central idea in this paper was prompted by discussions with Martin Davies, David Braddon Mitchell, David Lewis, John Bigelow, and Robert Pargetter. They should not be held responsible.

<sup>1</sup> Two seminal papers are Putnam 1975*a*, and Burge 1979. For a general introduction and further references see the introduction to Pettit and McDowell 1986.

<sup>2</sup> For the considerations see e.g. Stich 1983, chap. 8, and Fodor 1987, chap. 2; for the responses see e.g. Burge 1986*a*, and Jackson and Pettit 1988. A simple example where the broad causally explains behaviour is: 'The chicken is following the dog because it imprinted on the dog'.

behaviour familiar in the writings of, most particularly, Daniel Dennett and Donald Davidson.<sup>3</sup> However, as will emerge, the narrow content we will be defending is a truth-conditional notion of content which figures in folk psychology. And even (the increasingly beleaguered) defenders of narrow content have of late been conceding that their narrow content is neither a folk notion, nor a truth-conditional notion of content (although they insist that it is closely connected to folk, truth-conditional notions). In particular, the view that all truth-conditional content is (and must be, otherwise it would not 'hook on to the world') broad has become the conventional wisdom.<sup>4</sup> This paper is an attempt to turn back the clock. We will argue that certain points about the way we folk predict human behaviour, about the nature of our solution to the problem of predicting human behaviour, commits us to the existence of narrow, folk, truth-evaluable content. Unlike many extant arguments for narrow content, it will be no part of the argument for narrow content that there is anything 'wrong' from an explanatory point of view with broad content. Also, the sense in which the narrow content we will be defending is folk is that it figures in our everyday understanding of believers and desirers. For all we say, there may, or there may not be, English sentences naturally available to human subjects to express beliefs with these contents. Fred's abilities to discriminate, identify, and re-identify things with a certain shape may show that he believes that something has that shape, and yet he may have no (public-language) word for the shape; and of course animals are in this situation all (or nearly all) of the time.<sup>5</sup>

<sup>3</sup> See e.g. Davidson 1974a, and Dennett 1981. The way we put matters is perhaps closest to Stalnaker 1984. We do not suggest that Davidson, Stalnaker or Dennett would agree with what we find implicit in their doctrines. Indeed, we take it that Stalnaker 1989 is taking a position opposed to the one we defend here.

<sup>4</sup> Thus Dennett's 1982 defence of narrow content is, like Fodor's 1987, chap. 2 defence, explicitly a defence of a non-truth-conditional notion of content. Its link with a truth-conditional notion is that it determines a function from context to truth-conditional content. See also Devitt 1990. For evidence of the extent to which the view that we are opposing has become conventional wisdom, see Davies's 1990 review of McGinn 1989, and Fodor's 1991a, p. 6.

<sup>5</sup> We emphasize the point to make it clear that we leave open the possibility that the contents we discuss are not folk in the sense that they may not be contents which lend themselves to direct expression in the language of the folk. Some will want to express this contrast as one between what are described as personal and sub-personal beliefs. For a related contrast see the distinction between merely intentional contents and thought contents in Pettit 1992.

The first part of the paper is concerned with the folk problem of predicting behaviour. The second part of the paper is concerned with how the solution to the folk-predictive problem commits to narrow content. The paper concludes with some replies to objections.

## 1. The Folk Problem of Predicting Behaviour

### A. Finding Patterns in Behaviour

Suppose I want to predict how someone's body will move on some specified occasion or under some specified conditions: where do I start? One place to start would be with what is going on inside her as described in the language of medical science. If I know enough about her internal neurophysiology, I will be able to predict what will happen in certain muscle fibres, and from knowledge of what will happen in certain muscle fibres along with information about how the fibres link up to various bones, and the like, I will be able to predict how her body will move.

Nevertheless, we do not have to start from what happens inside a person. We do quite a reasonable job of predicting behaviour going by what is observable about people from the outside. Even with this highly restricted database we can do much better than chance. We can call the problem of predicting behaviour going by what is externally observable alone, the folk problem of predicting behaviour. For what the folk in general know about human behaviour is indeed pretty much restricted to what can be known without dissection; and when it does go beyond what can be known without dissection, it is not of great value in predicting what movements a person's body will make. The facts that the blood circulates and that the brain is very important to mental life are nowadays common knowledge, and certainly help predict the movements a body will make in certain circumstances, but they are not of crucial significance in predicting behaviour. (Otherwise Aristotle, who did not know either fact, could not have predicted successfully what movements people's bodies would make.)

We have remarked that we have in fact solved the folk problem of predicting behaviour in circumstances, but how did we do it? We can start by asking what constitutes the externally available evidence we use to solve the folk problem of predicting behaviour. The

obvious answer is: certain observed facts about what is sometimes called raw behaviour, the physical movements our bodies make described as such, rather than, for instance, the movements described in terms of the language of intentionally characterized action. For it is the raw behaviour which we more immediately perceive through the way that it impinges on our sense organs.

But now we face a well-known problem. What we need in order to predict behaviour in circumstances are past patterns; interesting projectable generalizations about what happened when, that are to be found in the historical behavioural data. What else could we reasonably use? But patterns in the raw behavioural data are hard to find. The behaviour of human beings is incredibly diverse. When the wind blows, trees generally do much the same thing; and when the wind blows harder, they do much the same, only more so. It is easy to find projectable patterns going by external data about the behaviour of trees when the wind blows. However, there are enormously many different ways that a human body may move when the wind blows. A person's body may move in such a way that she ends up encased in a jumper, or behind a wind break, or inside a house, or with her back towards the wind, or with her hand holding a string attached to a kite, and so on and so forth. What is more, there are many different bodily movements that put a body inside a jumper, behind a wind break, inside a house, back to the wind, or on the end of a string.

This problem is most familiar perhaps from discussions of functionalism. Michael Devitt, criticizing traditional formulations of functionalism which describe the inputs and outputs in purely physical terms, argues:

What psychological laws explain is not behaviour described as neural impulses, as mere bodily movements, or as any other brute physical event. These descriptions are at the wrong level, the level of psychological *implementation*. The level that yields the interesting generalizations of psychology requires that the behaviour be described as an action. This goes against the demands of old-fashioned reductionism, but so much the worse for that reductionism. Functionalism often seems not to have fully grasped its own message about explanatory levels. . . .the ones [the properties] appropriate for outputs are not brute-physical. (Devitt 1990, p. 393.)

Similarly, Robert Van Gulick (1990, p. 125, our emphasis) remarks that 'such disparate outputs as nodding one's head, raising a hand, or uttering 'yes' get classed together only on the basis of all being

taken as *gestures of assent* by the community of subjects', and Jerry Fodor (1982, p. 102, our emphasis) says that 'we have. . . no notion of behavioural systematicity at all except the one that makes behaviour *systematic under intentional* description'.

Should we follow these leads and retreat from raw behaviour to actions, behaviour described in intentional terms, in our search for the needed generalizations? We think not. Actions are bodily movements caused in the right sort of way, or more precisely, an action occurs if and only if a bodily movement caused in the right sort of way occurs. The fact that philosophers have found it so hard to specify precisely what 'the right sort of way' comes to does not justify holding that actions are some sorts of emergent entities. The fact that the total story about a person's actions supervenes on total information about movements, internal causes, and surroundings, tells us that actions are not emergent. Thus, to know that something is an action is to know about the internal aetiology of certain behaviour (and maybe in addition about social setting, community conventions, and the like, if it is an action like signing a cheque). But the folk problem is how to predict behaviour starting from the outside. We would be begging the question to start by describing the data available to the folk in terms of internal aetiology.

We can put the central point this way: to turn to intentional descriptions of behaviour is in effect to 'go internal'. But we cannot go internal to find the patterns and generalizations we folk need to get started. Rather, we have to find the patterns at the behavioural, external level first, and then maybe we can proceed to go internal to explain the patterns we have discovered, and so describe our data in the language of intentional action. When Robinson Crusoe saw Man Friday's footprint in the sand, the fact that it was Man Friday's footprint was, and had to be, his conclusion; where he finished up, not where he started from. Of course, when we see someone behaving in a certain way, we often find an intentional description like 'signalling a taxi', the immediately natural one to apply; we, that is, see the behaviour as signalling a taxi. Similarly, experts on the fossil record often see a fossil as the fossil of some long extinct insect. They do not say to themselves, 'That fossil is of such and such a shape. We have shown that such and such a shape is best explained as laid down by such and such a long-extinct insect.' Nevertheless, it is the 'raw' fossil and not the fossil under some highly historical description which is their datum.

To avoid possible misunderstanding, we should emphasize that our point is independent of the debate in the philosophy of action about the reference of compound singular terms like 'my intentionally annoying Fred yesterday'. On some views such terms do not refer to a compound of internal state and caused behavioural response, but rather to something entirely internal (a willing, perhaps); and on other views such terms refer to something constituted by, not identical with, an internal state and a behavioural response (in the same way that it is sometimes held that a jug is constituted by but not identical with the sum of its parts). But the epistemological point made above turns on the point that in order to know that I intentionally annoyed Fred yesterday, you need to know about internal aetiology of behaviour, and this point is independent of how the debate about the reference of terms designating actions should be resolved. Here is a simple illustration which brings out the independence of the epistemological issue from the issue about reference. In order to know that that object over there is the President's hat, you need to know, among other things, something about the President as well as something about the hat. But it does not follow from this that the term 'the President's hat' designates a compound entity containing the President as a part. Indeed, the most plausible view is that the term designates the hat alone.

In any case, there is a way to make patterned sense of the apparent chaos of raw bodily movement. It is to attend to what effects bodily movements have on the situations subjects find themselves in. It is to attend to what a subject's bodily movements in a given situation achieve. We remarked earlier that when the wind blows a tree will behave in much the same way on different occasions, but that with people it is far more complicated. When the wind blows their bodies move in a whole host of different ways. Nevertheless, there is a pattern to be discerned in this enormously diverse collection of bodily movements: very many of them have the effect that the person ends up out of the wind. And this is a pattern we can note without first describing things in intentional terms. The pattern is there in the brute physical movements themselves.

When we say that the pattern is there in the bodily movements themselves, we do not mean that the pattern is intrinsic to the movements. The pattern is discerned by noting the effects that the

bodily movements have on situations subjects find themselves in. A student of human behaviour sets us the following problem: find the pattern in Fred's behaviour consequent on Mary's entering the room. We focus on the relatively intrinsic features of Fred's bodily movements and get nowhere. There is no pattern in the movements *per se*. Then we notice that though the various movements on the various occasions are very different, they very often have the same effect on Fred's situation. As a result of them, on many occasions he ends up in the same half of the room as Mary. The penny drops. We then know the projectable generalization governing Fred's behaviour on Mary's entering the room.

We suspect that opposition to the commonsensical view that raw behaviour is our data for projecting behaviour in circumstances from past to future has been engendered by confusing the question of whether our data are facts about raw behaviour with the question of whether our data are those facts about raw behaviour which are central to taxonomizing raw behaviour *qua* raw behaviour. Two bits of elbow-bending may be grouped together as items of behaviour by the student of raw behaviour, by the physiotherapist, by the student of how muscular contraction issues in limb movement, but may consistently be grouped apart by the folk seeking projectable generalizations concerning behaviour in circumstances. The crucial point about one bit of elbow-bending may be that it brings beer near a mouth, so that the generalization that works for the folk in this case is: when a person enters a pub, expect behaviour that brings beer near mouth; whereas the crucial point about the other is that it wards off a fly, so that the generalization to project in this case is: when out in the bush, expect behaviour that leads to the departure of flies.

### B. Improving the Generalizations

What we are saying is that the folk patterns in behaviour, the patterns suitable for projecting going by observation of past behaviour in situations, are to be found by looking at the effects of the behaviour in those situations. Now we can effect a dramatic improvement in our predictions of behaviour by considering not just the effect a piece of behaviour would actually have, but in addition what effect it *would have were things thus and so*. Sometimes we find the pattern in Fred's behaviour not by noting

that his behaviour in fact brings him near Mary, but by noting that his behaviour would have brought him near Mary had things been thus and so. Mary enters a maze. Fred's behaviour brings him nowhere near Mary. We solve the puzzle by noting that had Mary ended up where one might well have expected she would, Fred's behaviour would have brought him near to Mary. A salient pattern in Ivan's behaviour is that it leads to his becoming richer, but there are some notable exceptions. We make sense of the failures by noting that his behaviour on those occasions would have made him richer had things been thus and so, and things being thus and so is how one, or at least Ivan, might well have expected them to be. In sum, to cut a long, familiar story short, we find the projectable patterns by using belief-desire psychology, by using the intentional stance, to put the matter in Dennett's 1979, 1981 terminology. The salient pattern externally available to solve the folk-predictive problem is: people behave in such a way that had their beliefs been true, then their desires would have been satisfied.

We take it that it is not really open to dispute that we are able to capture projectable generalizations concerning behaviour in the terms of belief-desire psychology. We do it every day, and it works. Based on past observations of you and others similar to you, I have opinions about what you believe and desire, and about what you would believe and desire were such and such to happen. I use these opinions to predict what you will do were such and such to happen. And the fact of the matter is that my predictions are right far more often than could possibly be explained by chance.

There is, of course, a great deal that is very much open to question. How exactly do we use behavioural data to arrive at hypotheses concerning a person's beliefs and desires; how determinate are these hypotheses; how important is it that we are language-using creatures; to what extent do we have to presuppose rationality, and so on and so forth (see e.g. Davidson 1974, Lewis 1974, and Dennett 1981). But we mention these highly debatable matters in order to emphasize what we take not to be debatable: belief-desire psychology captures projectable patterns in behaviour. If it did not, our success in using it would be a miracle, and it isn't. (We also take it as obvious that we make a still better fist of finding the patterns if we use not simply belief and desire hypotheses but instead degree of belief and strength of desire hypotheses. We will, however, neglect this complication here. Similarly, we neglect complications arising from belief *de se*; see e.g. Lewis 1979.)

### C. Describing Patterns in Terms of Possibilities

We can describe Fred's Mary-directed behaviour in terms of possibilities. There are many ways Fred's body might have moved. For each of these ways, there is how things would be were Fred's body to move in that way. Fred's behaviour is Mary-directed in the sense that, as a rule, his body moves so as to realize one of the ways where he is near Mary.

Similarly, we can describe the way belief-desire psychology captures the patterns in human behaviour in terms of possibilities. Associated with what a person believes at a time is a set of possibilities, the set compatible with what she believes, the set containing all the ways things might be for all she believes. To say this is not to say that belief is a relation to a set of possibilities. It is to say that belief effects a partition between the possibilities compatible with it and the possibilities incompatible with it. Whether it does this because that is what belief is, or because belief is, say, a relation to an internal sentence in the language of thought, and that sentence effects the partition by virtue of the fact that there are possibilities in which the sentence is true and possibilities in which the sentence is false, is another question. All we are using is the everyday fact that one way of giving pertinent information about what you believe is by specifying the various ways things might be for all you believe (see Lewis 1986a, p. 28). Similarly, associated with what a person desires at a time is a set of possibilities, the possibilities she would prefer to be actual. The way belief-desire psychology captures the projectable patterns in raw behaviour can now be described in terms of possibilities, as follows. Among the various possible ways that a subject's body might move at a time, a subject tends to move in such a way that had any one of the possibilities associated with her beliefs been the way things actually were, then one of the ways associated with her desires would have been actual. Had any one of the believed possibilities been actual, one of the desired possibilities would have been actual. We predict behaviour in circumstances by projecting these patterns.

Incidentally, this approach does not work if you take individual beliefs. You have to think in terms of big, conjunctive beliefs. Suppose someone says that they have found the pattern in Bruce's behaviour in terms of belief-desire psychology thus: Bruce believes

there is beer in the fridge and desires beer, and that is why his body moves towards the fridge. The trouble with this story as it stands is that the possibilities consistent with there being beer in the fridge are vastly too diverse for the rule 'Bruce's body will move in such a way that had any one of the beer-in-the-fridge possibilities been actual, Bruce would have actualized a possibility where there was beer in him' to be of any use at all. The rule won't effect any worthwhile reduction in the number of possible bodily movements. You have, instead, to work with the possibilities associated with a belief like: there is beer in the fridge and the fridge is four feet in front of me and there is no lock on the fridge and the beer is cold and I can traverse the distance from here to the fridge safely and so on. Only by working with a big belief do we cut down the diversity of the possibilities consistent with the belief sufficiently for the rule to give a prediction worth having about the movements Bruce's body will make.<sup>6</sup>

Does this concession mean that what we are putting forward as the way the folk can bring order to the apparent chaos of bodily movement is very different from belief-desire psychology as employed by the folk? We would not normally complain at an explanation of Bruce's behaviour that simply cited his belief that there is beer in the fridge and his desire for beer. However, we cannot take this evidence at face value. We all know that Bruce believes that bodily movement is fatiguing and that Bruce desires not to be fatigued. That combination would seem to lead to the opposite prediction that Bruce will not move towards the fridge, or indeed move at all. The folk know how to solve this problem. They tell us about trade-offs. They tell us that Bruce believes that a certain movement will bring him beer and that it will bring him fatigue, and that Bruce desires beer and fatigue more than no beer and no fatigue. But this is simply the first step away from individual beliefs and desires in their explanations and towards big beliefs and big desires. The folk know implicitly that they have to work with big, conjunctive beliefs and desires.

<sup>6</sup> Why not handle the problem by changing the predictive rule to: Bruce's body will move in such a way that for at least one possibility consistent with his belief, it is true that had it been actual, he would have ended up with beer inside him? Because such a rule is consistent with: Bruce's body will move in such a way that for at least one possibility consistent with his belief, it is true that had it been actual, he would have ended up without beer inside him.

## 2. The Derivation of Narrow, Folk, Truth-Conditional Content

We have rehearsed how to find predictively useful patterns in behaviour in circumstances in terms of the possibilities associated with a person's beliefs and desires. How does what we have rehearsed commit us to narrow, folk, truth-conditional content? The commitment to folk, truth-conditional content is relatively straightforward, and we will deal with this part of the story first.

### A. *The Commitment to Truth-Conditional Content*

We can find patterns in behaviour in terms of sets of possibilities associated with a person's beliefs and desires. But surely sets of possibilities associated with a person's beliefs and desires are among the kinds of thing one might properly mean by the content of what they believe and desire, particularly when these sets enable us to predict behaviour in circumstances. For, as already observed, such sets tell us something very important about what a person believes and desires. Maybe it is not the only thing you might properly mean by the content of what they believe and desire. 'Content' is a recently prominent term of art and may well mean rather different things to different practitioners of the art. However, the sets of possibilities we have been talking about are one thing you might properly mean by the term. But content in terms of sets of possibilities is automatically truth-evaluable. It comes out true precisely if the content set of possibilities contains the way things actually are.

### B. *Is Predictive Content a Folk Notion?*

Let's call the sets of possibilities associated with beliefs and desires 'predictive content'. This highlights the facts that we are not here committing ourselves to there being exactly one notion of content (maybe there are other notions which would need to be marked by differently prefixing the term 'content'), and that we introduced the notion as a way of predicting behaviour in circumstances (but we will see later that the notion is not merely predictive). Predictive content counts as folk on two grounds. First, there is the point that

it is a piece of common sense, as well as a piece of Stalnaker or Lewis theorizing, that sets of possibilities are individually associated with beliefs and desires. It is common currency to explain how you would like things to be in terms of rankings of the various possibilities, and how you take things to be in terms of how likely various possibilities are.

Secondly, the projectable patterns that we have been remarking can be described in terms of sets of possibilities associated with beliefs and desires are patterns in behaviour in circumstances discernible by the folk. The story is in terms of such happenings as Mary's entering the room, Fred's moving near Mary, the wind's blowing, the putting on of jumpers, and so on and so forth. We did not need to use the language of neuroscience, of peripheral stimulations, of retinal activations, or anything of that kind. We were not in the business of identifying a concept foreign to the folk and then arguing that, for whatever reason, the concept was needed in mature cognitive science, properly scientific psychology, or whatever. We were simply spelling out something which is an implicit part of common-sense theory about people's behaviour. The patterns we described in terms of possibilities associated with belief and desire are not news to the folk. They use them implicitly all the time in predicting behaviour. All that is unfamiliar to them is the jargon and the theoretical articulation.

### C. *Is Predictive Content Narrow?*

Opponents of individualism in psychology in general, and of narrow content in particular, frequently argue the case against individualism by urging that our understanding of central psychological concepts is tied to our interactions with what goes on around us described as such—that is, described as what goes on around us. Their claim is that, as a matter of principle, you cannot tell the psychological story in terms of how things are at and inside the skin. Here is how Burge puts this kind of position:

philosophy of psychology must do justice not only to the mechanistic elements in the science. It must also relate these to psychology's attempt to account for tasks that we succeed and fail at, *where these tasks are set by the environment*. . . . The most salient and important of these tasks are those that arise through relations to the natural and social worlds. A theory that insists on describing the states of human beings *purely* in terms that abstract

from their relations to any specific environment cannot hope to provide a completely satisfying explanation of our accomplishments . . . (1986a, pp. 44–5)

This kind of position is sometimes put by saying that a psychological state's 'ecological job description' is essential to it.

A natural thought is that the content we have been talking about, predictive content, is non-individualistic and so broad. The folk-predictive problem we have been talking about was framed in terms of predicting how subjects interact with their surroundings, of predicting how Fred will behave in future when Mary enters the room, of predicting what a person will do when the wind blows. And the kinds of predictions we got were predictions of how the subject's behaviour would change his or her situation described in environmental terms—Fred will end up in Mary's half of the room, the person in the wind will end up out of the wind, or whatever. We were talking very much in terms of 'relations to the natural and social worlds'. Predictive content's job description clearly falls into the ecological category.

It is, however, important to distinguish two anti-individualistic doctrines, doctrines which, it seems to us, sometimes get conflated: one says that some given central psychological property cannot be explained individualistically; the other says that the property does not supervene on how the subject is from the skin in: it is, that is, not necessarily shared by doppelgangers.

Predictive content is certainly non-individualistic in the first sense. It can only be explained by reference to a subject's interactions with his or her surroundings. It is like the concept of water solubility. If you do not understand the connection between being water-soluble and behaving in a certain way on being put in water, you do not understand what water solubility is. Although as a matter of fact (and perhaps as a matter of necessity) there is some internal state of a water-soluble substance which is responsible for the behaviour definitive of being water soluble, if what you know is confined to that internal state, you do not know that the substance is water-soluble. Nevertheless, any internal duplicate, any doppelganger, of a water-soluble substance is itself water-soluble. Hence, water solubility is not a non-individualistic property in the second sense; being water-soluble is a narrow property in the official sense defined in terms of supervenience on internal nature. The explanation of how it is that water solubility is not non-individualistic in



the second sense is, of course, that (a) it is the totality of potential interactions with water which matters for being water soluble, not which of the totality are actual, and (b) if *X* and *Y* are internally identical substances, the totality for *X* is one and the same as the totality for *Y*. Or consider a predicate like 'if *x* were near a tiger, *x* would start to run away from the tiger'. Is it individualistic? You might say no on the ground that what makes the predicate true of some *x* is in part how the world is (it is non-individualistic in the first sense), or you might say yes on the ground that, necessarily, if it is true of *x* then it is true of any doppelganger of *x* (it is not non-individualistic in the second sense).

We can put the distinction in terms of possible worlds as follows. To be individualistic in the first sense is to be inter-world narrow, where a property *P* of *x* is inter-world narrow if and only if in every possible world any doppelganger of *x* has *P*. To be individualistic in the second sense is to be intra-world narrow, where a property *P* of *x* is intra-world narrow if and only if in every possible world where *x* has *P* any doppelganger of *x* has *P*.<sup>7</sup> Being water-soluble is not inter-world narrow. A doppelganger of sugar in a world where the laws are relevantly different need not be water-soluble. But being soluble is intra-world narrow: within a possible world, internal duplicates are duplicates with respect to whether or not they are soluble, and so to find a substance which is internally exactly like sugar but which is not water-soluble, you must move to a different possible world. Or, equivalently, when we said that being water soluble supervenes on internal nature, we should in strictness have said that it supervenes on internal nature plus laws.<sup>8</sup>

Our argument will be that while all predictive content is obviously not individualistic in the first sense, some predictive content is individualistic in the second sense. Some predictive content is intra-world narrow — there are predictive contents such that, necessarily, doppelgangers in the same possible world share them. But predictive content is never inter-world narrow. My doppelganger in a world with very different laws may well have

<sup>7</sup> McGinn 1989, chap.1, distinguishes strong externalism from weak externalism. We think that his distinction is related to the distinction between inter- and intra-world narrowness thus: strong externalism says that no content (or no content of such and such a kind if it is a restricted version of strong externalism) is intra-world narrow; weak externalism says that no content (or no content of such and such a kind if it is a restricted version of weak externalism) is inter-world narrow.

<sup>8</sup> We are indebted here to David Armstrong.

states with totally different predictive contents. Does this mean that we are not really going against the conventional wisdom that all folk, truth-conditional content must be broad? For perhaps all that is meant is that such content cannot be inter-world narrow, not that it cannot be intra-world narrow. However, the arguments that folk, truth-conditional content must be broad rest on claims about how certain (actual) causal links between the head and the world are essential for truth-conditional content, and about how context (Is it water or XYZ that surrounds me? Is it that cup or this cup that is before me?) and one's linguistic community affect truth conditions. And arguments like these are arguments that content can vary without supposing any change in the laws of nature, or in general any change that is 'big' enough to require supposing a change in possible world.<sup>9</sup> They are, therefore, arguments that folk, truth-conditional content cannot be intra-world narrow. In any case, our defence of the narrowness of predictive content is a defence of the intra-world narrowness of predictive content.

Our argument that some predictive content is narrow (understood from here on in the intra-world sense) turns on two facts: one is that we folk are committed to the robustness of solutions to the problem of predicting behaviour for a person *S*; the other is that for any historical solution to the folk problem of predicting behaviour, there is an ahistorical solution.

By a robust solution we mean a solution which would not be exposed as being based on the misleading appearance of a pattern in a biased sample, by the acquisition of any actual or possible new data concerning what a person does or would do in circumstances. We dream of knowing everything of the form: were such and such to happen, then Fred's body would move thus and so; we dream, that is, of a complete solution to the folk-predictive problem in the form of a huge raft of subjunctive conditionals whose antecedents are detailed specifications in folk terms of possible circumstances (were Mary to enter the room when Fred was in the room and the lights were on and Fred's leg was not broken and. . . ), and whose consequents are about bodily movements described in common-sense, environmental terms (Fred's body moves so that he is nearer beer. . . ), a huge raft of subjunctive conditionals of the form '*C<sub>i</sub>* → *B<sub>i</sub>*'. What we actually have in practice is a tiny selection from this

<sup>9</sup> As many have noted, Twin Earth need not be (and typically is not supposed to be) in a different possible world from Earth.

complete solution, a tiny selection which constitutes our observations that  $C_i$  &  $B_i$  for some  $i$ ; a selection which displays patterns, projectable generalizations, capturable as we have seen in terms of certain hypotheses about belief and desire framed in terms of sets of possibilities. (When seeking a pattern for a given individual, we will of course have collateral information of value, most particularly to do with observed truths of the form  $C_i$  &  $B_i$  for other creatures, including ourselves.)

Now, in using our pattern to predict behaviour in circumstances, we commit ourselves to its being a robust pattern. For to hold that the pattern we have discerned is misleading, is a fluke pattern whose deceptiveness would be revealed if only we had the full picture, is precisely to hold that we are not entitled to project the patterns we discern in our database. Our practice of projecting the belief-desire patterns we discern in our fellows from past to future commits us to believing that there is a robust solution in terms of the possibilities associated with belief and desire to the folk-predictive problem—a solution which would survive learning all there is to know about how a person's body would move were such and such to happen. And what is more, we are committed to holding that this survivor solution, though no doubt vastly more detailed than our solution, is an improvement and detailing of our pattern, not a demonstration that the pattern we discerned was a misleading appearance arising from a biased sample. For you cannot properly project a pattern from a sample at the same time as holding that by the lights of the very best possible evidence, the pattern you project is quite the wrong one to project. Perhaps it is—that is the hostage to fortune inductive projection gives—but you must in consistency hold that this unpleasant possibility is unlikely. Here is a simple example to help make the point clear. A machine is printing out numbers on a tape. I observe just the sequence: 1,2,3,4, and I predict that the next number produced by the machine will be 5. I cannot, consistently with this prediction, hold (a) that the most likely sequence to have preceded my observation is: 1,2,3,4,1,2,3,4, 1,2,3,4, and (b) that had I observed that fuller sequence, then the right prediction for me to make would have been that the machine would next print out a 1.

In arguing that we are committed to there being a robust solution, we are not suggesting that we are committed to there being exactly one robust solution. For all we have said, there might be a number

of robust patterns. Davidson and Dennett (most particularly) have made us sensitive to the possibility that not only might there be no single best belief-desire hypothesis covering what we have actually observed of a person's behaviour, there might be no single best belief-desire hypothesis covering the totality of actual and possible observations of a person's behaviour. If this happens, there will be more than one robust solution in terms of the possibilities associated with belief and desire to the folk-predictive problem. But that there is at least one robust pattern is all we will need here.

It might now appear a simple matter to derive our result that some, indeed all, predictive content is narrow. Robustness says that at least one belief-desire pattern is there in the complete raft of subjunctive conditionals saying how a person would behave in such and such a situation. But if  $X$  and  $Y$  are doppelgangers, then the complete raft for  $X$  is one and the same as the complete raft for  $Y$ . The movements that  $X$ 's body would make were  $X$  in such and such a situation are the very same as the movements  $Y$ 's body would make were  $Y$  in such and such a situation. But predictive content captures the belief-desire patterns. So the predictive contents of  $X$ 's beliefs and desires must be the same as those of  $Y$ . (If there is more than one belief-desire pattern that fits the full raft  $X$  and  $Y$  have in common, and there is no non-arbitrary way to choose between the patterns, then  $X$  and  $Y$  share predictive contents in the sense that the range of contents each of which is an admissible, but none of which is a determinately correct, candidate to be a predictive content associated with their beliefs and desires is the same for each.)

You might complain that I and my Twin Earth doppelganger do not behave in the same way in the same situations. I drink water; he drinks XYZ. I reach for that cup; he reaches for this cup. But that is to misunderstand what is meant by 'the same situations' here. It means what Stich (1983, p.165) has in mind when he talks of 'replacement'. If I were replaced by my Twin Earth twin, it would be true that he would drink water, and would be true that he would reach for the very same cup as I in fact reached for. The argument does not turn on the fact that doppelgangers in fact behave in the same way—something which is only true at the level of raw movement so described—it turns on the fact that they would in every situation behave in the same way, and that is true both in the sense that their raw movements are the same and in the sense that the way these raw movements would affect their environmental

orientations is exactly the same. Their movements are, that is, the same in just the sense we saw to be crucial for detecting the belief-desire patterns. In a certain situation, Fred moved towards Mary. Had Fred been replaced by Fred's twin in that very situation, Fred's twin would have moved towards Mary. In this lies the force of the claim that predictive content is narrow. It is this which makes it so hard to see how the best belief-desire pattern (or set of equally good patterns if indeterminacy threatens) could be different for Fred and twin Fred.

But now we need to note a lacuna in the derivation of narrow content from robustness just given. Doppelgangers' behaviour in situations will differ in historical properties. It might be true that if Fred had been in situation *S*, he would have moved for the third time towards Mary, but not true that had twin Fred been in situation *S* he would have moved for the third time towards Mary. Twin Fred has never met Mary, so had he been in *S*, he would have moved towards Mary for the first time. It can be tempting to rule out such differences as irrelevant to psychology. This is Stich's approach. He introduces the notion of an 'autonomous behavioural description', a description which 'if it applies to an organism in a given setting, then it would also apply to any replica of the organism in that setting' (Stich 1983, p. 167), and argues that psychology only needs to concern itself with autonomous descriptions. This would rule out properties like moving towards Mary for the third time or selling one's car (Stich's own example), because instantiating them depends on history — to sell a car (legally) you have to have entered into a certain transaction in the past — as well as on setting (current setting, that is, which is what Stich has in mind). The trouble with this view is that the only reason he gives, and could give, for supposing that autonomous descriptions are all psychology needs is that systematic generalizations of behaviour do not usefully employ non-autonomous descriptions. This seems simply false. Very few people touch a hot stove twice; a cat behaves very differently on the second occasion of meeting a young child; and so on. Historical, and so non-autonomous, properties can play a major role in explaining and systematizing behaviour. Certainly, from our perspective here, you could not possibly rule out historical properties. Noting them can reveal a behavioural pattern apt for projection into the future.

What is true is something different. Whenever there is a belief-desire pattern manifest in behaviour in situations which can be

captured using historical descriptions, there is a belief-desire pattern which can be captured without using historical descriptions; indeed, can be captured using autonomous descriptions alone. This is because history does not act at a distance; causation is local. Consider the behavioural pattern: when children come close to a hot stove for the second time, their bodies slow down. The stove will have the property of having been come across previously, and this very fact will be part of the causal explanation of the observed behaviour. But it will also be true that this historical fact will work via some feature the stove possesses at (or just prior to) the time the avoidance behaviour occurs. There will be something about the stove — an autonomous feature of the stove — which marks it out for the child as having been previously encountered. This means that whenever we can predict the child's behaviour by appeal to the fact of the previous encounter, we can in principle also predict the child's behaviour in terms of an autonomous feature. Or think of a machine programmed to respond when presented with an object of the same shape as it was presented with previously. Any response to the fact that a presented object is the same shape as one previously presented to it, is also a response to the fact that the presented object is some particular shape together with the fact that a previously presented object was that particular shape. The local nature of causation tells us that whenever there is a belief-desire pattern capturable in non-autonomous terms, there is a belief-desire pattern capturable in autonomous terms.

We can now repair the argument for narrow predictive content. The crucial point is that robustness tells us that the best autonomous belief-desire pattern (or patterns) for Fred and twin Fred must be the same, and so that the predictive contents which describe those autonomous patterns, the sets of possibilities associated with Fred and twin Fred's beliefs and desires which bring projectable order to behaviour in situations as described autonomously, must be the same for Fred and for twin Fred. So the conclusion we get is that there exists narrow, predictive, folk, truth-evaluable content, but not that all predictive content is narrow, or, still less, that all content is narrow.

Although the predictive content which brings order to the autonomous belief-desire patterns must be narrow, it does not follow that sentences in English ascribing content using only autonomous descriptions have narrow truth-conditions. If it is true

that Fred moves towards Mary in a certain situation, then it will be true that twin Fred would do so in that situation. But it is not the case that 'Fred believes that — Mary —' is true if and only if 'Twin Fred believes that — Mary —'. Twin Fred will, by virtue of failing to have the appropriate (actual) causal links to Mary, lack any beliefs truly reportable by sentences in our mouths using the name 'Mary'. There will, of course, be an explanation of why it is true that twin Fred would move towards Mary. Fred and twin Fred are reactively sensitive to the very same properties—the properties that underlie Fred's capacity to identify and re-identify Mary—and these 'Mary-distinctive' properties will figure in the predictive contents which make projectable sense of the autonomous patterns Fred and twin Fred have in common. But that is another question. Moreover, it is very much an open question whether or not Fred or twin Fred have words for the 'Mary-distinctive' properties. Think of the familiar problem of verbalizing one's responses to gestalts, or think of how hard it can be to say exactly how you are able to identify a person as someone you have met before.<sup>10</sup>

### 3. On Two Objections

It might be objected that content is explanatory as well as predictive. Perhaps we have shown that there exists something folk, narrow, and truth-evaluable, which is associated with belief and desire and has predictive value. But in order for a notion to count as an interesting notion of content it must, in addition, have explanatory value. Merely predictive content is not content.

We reply that the existence of patterns good for prediction effectively ensures explanatory value in certain kinds of cases, and the case of predicting human behaviour by noting belief-desire patterns is one of these cases. Suppose that you see a robot successfully negotiate a minefield. The robot is not being controlled by a human operator, or by another robot. We have a pattern: getting through a minefield successfully. Suppose that we project it into the future; that is, we predict that the robot will make it

<sup>10</sup> One author, P. P., inclines to the view that narrow, predictive content is generally sub-verbal. The other author, F. J., is more optimistic about our powers to express narrow content in English, while granting that a great many of the sentences we naturally use to ascribe beliefs and desires have broad truth-conditions.

through minefields in future trials, and suppose that the robot does indeed successfully negotiate minefields on a regular basis.

Can we offer anything by way of explaining the robot's success? We can. The probability that the robot makes it through minefields by chance is extremely small. We do not have to look inside the robot in order to be confident that the robot (*a*) registers the location of the various mines, that is, the mines' locations causally impinge on the robot leaving some kind of enduring trace which carries the information as to the whereabouts of the mines, and (*b*) whatever it is that drives the robot accesses this information and employs it to steer the robot away from the mines. And now we can explain the success of the robot: it gets through minefields because it stores information about the whereabouts of the mines and this information controls the movements of the robot. Of course, it would be nice to be able to say a lot more. How does it store the information, how is the systematic co-variance between internal nature and the location of mines secured, and how does the robot's drive-train access the information appropriately? But to have far less than ideal explanatory value is not to have no explanatory value.

The same goes for the belief-desire patterns that solve the folk-predictive problem. The very fact that they hold tells us a great deal about what information from our surroundings is received, stored, and influences our bodily movements. And in projecting the patterns, we take all this for granted. It is obvious that Fred receives, stores, and utilizes information about Mary's whereabouts. And in predicting Fred's future responses to Mary's arrivals and departures, we take all this for granted. But that is to say that the contents we use to describe the belief-desire patterns in solving the predictive problem—predictive contents, as we are calling them—have explanatory as well as predictive value.

When we say that it is obvious that Fred receives, stores, and utilizes information about Mary's whereabouts, we do not mean that it is obvious *no matter what else we might know*. If we discovered that the robot which made its way through the minefield was in fact radio-controlled by a human operator, what would then be obvious would be that the information about the whereabouts of the mines was stored in the human operator. Similarly, if we consider a science-fiction case where Fred turns out to be controlled by radio transmissions from Mars, the obvious hypothesis about the locus of the information reception, storing, and utilization will

become that it takes place on Mars and not in Fred. But as a matter of fact we know more than enough about the human beings who display the belief-desire patterns to know that the receiving, storing, and utilization takes place inside them.

Here is another way in which extraneous information about Fred might undermine what would, in the absence of that information, be the explanatory hypothesis demanded by our success in predicting his behaviour. We present Fred with a number of complex chess positions. Each time he comes up with an excellent suggestion for the next move. Presented with a position *P* he makes excellent response *R*. What feature of *P* prompts his response? In view of the fact that he makes an excellent response each time, the obvious hypothesis is that it is the feature of *P* responsible for the response *R* being excellent. Perhaps *P* contains a certain imbalance in the distribution of white pawns, and it is this imbalance along of course with other matters, like the rules of chess, which (*a*) is responsible for *R* being an excellent response, and (*b*) causes Fred to make *R*. But the hypothesis that Fred's responses are generated by the properties of the positions that make them excellent responses could be defeated. Perhaps we discover that Fred has somehow coded all the positions he could possibly be presented with. He has a unique tag for each, and he has simply learned off by heart, say from a table provided by an expert chess-player, an excellent response for each tag. In that case he is no good at chess. He is good at memorizing chess positions, and that is not the same thing. His defect is that he does not know what makes one of his excellent responses an excellent one. He responds to the tag in each case, but the nature of the tag is not what makes the response an excellent one. That is why forgetting which tag goes with which position would be such a disaster for his performance at chess (or, for that matter, why he would be quite unable to handle even minor changes in the rules of chess).

We mention these two ways in which the obvious explanatory hypotheses arising from predictive success might be undermined in order to highlight the fact that, by and large, we know that these sorts of things are not happening with our fellow human beings.<sup>11</sup> By and large the obvious explanatory hypotheses about the

information we store, retain, and access, though not the various speculations about how we do it, are known to be true. That is how predictive content is also explanatory content.

Secondly, it might be objected that the account of content being defended here is simply a revival of the Lewis-Stalnaker view that the content of belief is a set of possible worlds, and that this view is known to face an overwhelming objection. It entails that when *P* and *Q* are true in just the same worlds, that is, are logically equivalent, *S* believes that *P* if and only if *S* believes that *Q*. And a moment's reflection on belief in mathematical theorems and on failures of logical omniscience shows that this view is absurd (see e.g. Field 1978).

We reply that it is one thing to say that *S* believes that *P* if and only if *S* has the belief attitude to the set of *P*-worlds; quite another to say that an important fact about what I believe can be captured in terms of sets of possibilities. And it is only the latter which this paper employs. It was to emphasize this that we talked of the set of possibilities associated with what I, or Fred, believes. All we need is the fact that if I tell you which are the open possibilities and which are the closed possibilities as far as I am concerned about what will happen tomorrow, I tell you something important about what I believe concerning tomorrow.

#### 4. Summary

We humans can predict behaviour in circumstances going on generally available external evidence. We have, that is, solved the folk-predictive problem. We must have solved it by using facts about raw behaviour, but how can order be brought to the chaos of raw behaviour? The answer is to look on behaviour as determining a function from one set of possibilities, the set associated with what the agent believes, to another set, the set associated with what the agent desires. These sets are properly describable as a species of folk, narrow, truth-evaluable content: they are folk because the story is all in terms that are common currency; they are truth-evaluable because they are sets of possibilities, counting as true precisely if they contain how things actually are; and they are narrow because provided they are the sets we need to solve that part of the folk-predictive problem which can be framed in autonomous

<sup>11</sup> The two ways are variations on cases described in Block 1981a and Peacocke 1983.

terms, the sets which solve the problem for a given agent must be the same as the sets which solve the problem for any doppelgänger of that agent—otherwise the solution is no solution.

## Object-Dependent Thoughts: A Case of Superficial Necessity but Deep Contingency?

H. W. NOONAN

AN *externalist* (or anti-individualist) is one who holds that the psychological states possible for a subject are not independent of the state of the world outside his skin.

One version of externalism is the Evans–McDowell doctrine of ‘object-dependent’ or ‘Russellian’ thoughts: thoughts about a particular concrete object, or objects (other than the thinker and the time of the thought), which are possible for a thinker only if he stands in some relation to that object, or those objects—a relation perhaps appropriately thought of, in Russellian terms, as one of ‘acquaintance’ (see Evans 1982 and McDowell 1984, 1986).

My own view is that externalism is a correct doctrine, but that the Evans–McDowell doctrine of object-dependence is mistaken. In what follows I shall first set out the argument which convinces me of the incorrectness of the object-dependence doctrine and distinguish it from an argument frequently put forward in the literature against externalism in general. (For reasons that will become apparent I shall henceforth refer to the former argument as the *Two List Argument* (TLA) and to the latter argument as *the argument for the causal redundancy of externally individuated contents* or, for short, as the *Causal Redundancy Argument* (CRA).) Granted this distinction, however, it might still seem that the former argument, i.e. the Two List Argument, if cogent at all, must be as effective against externalism generally as against object-dependence in particular. In fact, I believe, the Two List Argument is as effective against some other forms of externalism as it is against object-dependence. In particular, I think that it is as effective against the versions of externalism defended by Tyler Burge (1979 and 1982a)