Facta Philosophica 5, 2003: 183–201
© Peter Lang, Switzerland

Belief Ascription and the Illusion of Depth

Lynne Rudder Baker

Philosophy, like other forms of inquiry, is guided by metaphors. One particularly tenacious metaphor in philosophy is the metaphor of depth. So entrenched is the metaphor of depth that the fact that it is a metaphor at all passes unnoticed. Here, for example, is a well-known passage by Jerry Fodor: "I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of spin, charm, and charge will perhaps appear upon their list. But aboutness surely won't; intentionality simply doesn't go that deep." (Fodor, 1987, p. 97)

Through the lens of the "depth" metaphor, reality is arranged hierarchically in levels. The most fundamental ("deepest") level is that of microphysics; ascending through the other physical and special sciences, we arrive at the "shallowest" level—the intentional level. The intentional level is described in an intentional vocabulary of psychological states (like belief, desire, and intention), of character traits (like generosity, honesty, and irritability) as well as the intentional vocabulary of social, legal, economic and political phenomena (like heir, wife, stock-holder, mayor). All of these—heir, wife, stock-holder, mayor—are intentional in the broad sense that there could be no exemplification of the properties of being an heir, or a wife, or a stock-holder, or a mayor in a world without individuals with propositional attitudes.

With this picture of levels in place, it is natural to think that a description or explanation that is closer to microphysics is more ontologically significant than a description or explanation at a "shallower" level. This is what I call 'the illusion of depth'—the unqualified assumption that the closer to microphysics an explanation is, the more profound and revealing it is. What is pernicious about the "depth" metaphor is not the notion that there are levels, but rather the idea that an explanation or description at level n-1 is in any way more ontologically significant, or more worthy of respect or any other honorific, than an explanation or description at level n. The "depth" metaphor works spectacularly well in some areas: Few would deny that the discovery of DNA significantly

deepened what was learned from Mendel and his peas. And the genome project promises still more profound knowledge to come.

Now if we consider belief against this background picture governed by the metaphor of depth, we can see why the idea that beliefs are brain states has such a grip. Subpersonal levels are deeper than the personal level; the states of the brain are at a subpersonal level. So, if we can understand beliefs (attributed at the personal level) in terms of (subpersonal) brain states, we have made progress—according to the captivating metaphor of depth. But as I have argued elsewhere, it is a mistake to try to understand belief in terms of brain states.(Baker, 1995; Baker, 1999) There is a fundamental disanalogy between intentional states like belief and, say, genes. With genes, we get deeper and deeper explanations as we go from genes, to the double-helix of DNA, to the nitrogenous bases A,C,G and T that make up DNA molecules. But other applications of this metaphor of depth are less fruitful. To replace intentional explanations or descriptions in terms of belief by "deeper" physical explanations or descriptions in terms of brain states is just to change the subject, as Donald Davidson rightly pointed out. (Davidson, 1970)

Many philosophers who are attached to the metaphor of depth would respond that if beliefs are not brain states, then believing that p is not a genuine property of people; these philosophers simply require there to be "deeper" descriptions and explanations in order to validate ascriptions of belief in the first place. If such deeper descriptions are not forthcoming, then, they conclude, belief ascriptions are merely convenient devices that we use in the absence of knowledge of what is really going on. These philosophers endorse a kind of fictionalism about belief. Thus, the widespread attachment to the metaphor of depth by philosophers of mind generates a dichotomy: Either beliefs are brain states or they are fictions. I hope to persuade you that this is a false dichotomy.

To do so, I'll briefly set out my own account of belief and compare it with Dennett's. Then, I want to contrast belief-ascriptions with physical disposition-ascriptions. Next, I shall argue that my construal of beliefs allows belief-explanations to be causal explanations. I'll conclude by indicating what I think is wrong with the "depth" metaphor.

A Practical-Realist Account of Belief

My aim is to provide an account—I call it 'Practical Realism'—of what it is to have a belief. (I am using 'belief' as a stand-in for desire, intention, expectation, hope and fear—psychological states with propositional content (i.e., propositional attitudes) generally.) This account of belief

is not intended to be an analysis or a definition. An analysis (or a definition), gives the meaning of a concept (or an expression) without presupposing the concept (or expression) under analysis. I do not think that 'belief' can be reduced to other terms that do not presuppose 'belief.' So, I am not proposing to give an analysis or definition of 'belief' or of 'S believes that p.' (In this respect, my project is like Kripke's in *Naming and Necessity*, which sets out a "picture"—not even a full-fledged theory—of naming, and "takes the notion of intending to use the same reference as a given." (Kripke, 1972, p. 97) There is no thought in Kripke of analyzing or defining the idea of reference. I am no more offering a reductive analysis of belief than Kripke did of reference.)

The best that we can do, I think, is to say what is required for someone to have the property of believing that p. So, I intend to give a schema for necessary and sufficient conditions for S's believing that p. Whereas an analysis or definition reductively shows how to replace one concept or expression with another, provision of necessary and sufficient conditions for the exemplification of a property may be altogether nonreductive. For example, we may know necessary and sufficient conditions for someone's being in pain (in terms of brain states, for example) without having any analysis or definition of 'pain.' Pain, like belief, may be irreducible to anything else. We also have nonreductive necessary and sufficient conditions for someone to be President-elect of the U. S.—viz., to be certified as having received 270 or more electoral votes.

Let me begin with a few matters that I am going to take for granted: First, the primary bearers of beliefs are people, not sub-personal parts of people like brains. Even if it were true that one has a belief in virtue of being in a certain brain state, the locus of belief would still be the whole person. Second, a belief is not an entity. The root idea of belief is of believing. Believing that snow is white is a property; the term 'belief' is just a nominalization of 'believing.' Third, belief ascriptions are true or false in as robust a sense as any other property attribution. (It is as true that you believe that England has a monarchy as it is that you attend philosophy conferences, or that you have been asleep within the past 48 hours.)

Believing that, say, Rome is in Italy is a property that someone has in virtue of there being a relevant set of true counterfactuals about the person. (I count subjunctive conditionals with antecedents true in the actual worlds as counterfactuals in a somewhat extended sense.) The counterfactuals concern what the believer would do, say or think in various circumstances and this set is both necessary and sufficient for that person's believing that p. To be sufficient for someone's believing

that p, the set of relevant counterfactuals for a belief often includes not only those in which one would do, say or think something but also those in which one would *not* do, say, or think something. For example, suppose that the following counterfactual is relevant to Jones's belief that she left her car keys in her office: "If she wanted her car keys, then she would return to her office." But suppose that Jones also believed that she had a spare set of keys in her pocket, and she returned to her office to pick up her briefcase.[1] In that case, Jones may lack the belief that she left her car keys in the office even though the counterfactual is true of her. This shows that the counterfactual just mentioned—"If she wanted her car keys, then she would return to her office."—is not sufficient for Jones to believe that she left her car keys in her office. But now add to the set of counterfactuals following "negative" counterfactual: "If she did not want her car keys, then she would not return to her office." The corresponding "negative" counterfactual is false: she still would have returned to her office to pick up her briefcase. Since the "negative" counterfactual is not true of Jones, the pair of counterfactuals—"If she wanted her car keys, then she would return to her office" and "If she did not want her car keys, then she would not return to her office"—may be sufficient for Jones's belief that she left her car keys in her office. So, the set of relevant counterfactuals may include counterfactuals about what one would *not* do, say or think in certain circumstances as well as counterfactuals about what one do, say, or think in certain circumstances.

I am not making an epistemological claim: The account does not entail that anybody knows exactly which counterfactuals are sufficient for a given belief. Nevertheless, we often do know what counterfactuals are relevant to the beliefs that we ascribe to someone. We know in part by knowing the meaning of the 'that' clause in the belief ascription, in part by knowing the person's other attitudes, in part by knowing general facts about the world and how people typically behave.

Since counterfactuals are very tricky to evaluate, let me make some sketchy remarks about the counterfactuals associated with a belief. Generally, we evaluate counterfactuals by considering the nearest possible world in which the antecedent is true (sometimes this is the actual world). If the consequent is true in that world, then the counterfactual is true; otherwise, the counterfactual is false. In any context of belief-ascription, many background assumptions are in play, and the evaluation of the counterfactuals is confined to those worlds where the background

1 Don Gustafson showed me the need to discuss this problem.

assumptions are true. Call such worlds "belief-relevant" worlds. Belief-relevant worlds have the same laws of nature as ours. In addition, belief-relevant worlds for the counterfactuals in the associated set for 'S believes that p' are subject to the following constraints:

(1) A belief-relevant world is one in which S is physically the same as (or as similar as possible to) the actual S at the time that S is alleged to have the belief. This prevents having to conclude that a newborn baby believes that $E = mc^2$ on the grounds that if the baby were asked if $E = mc^2$, and understood the question and wanted to cooperate, then the baby would say, 'Yes.' That counterfactual is true of a newborn baby and of most anyone else: Almost anyone who understood the question and wanted to cooperate would say 'Yes, $E = mc^2$'. But the truth of that counterfactual is not relevant to any belief had by a baby, who did not understand 'Does $E = mc^2$?' Confining belief-relevant worlds to those in which the believer is physically similar to the believer in the actual world avoids considering worlds in which a newborn is asked 'Does $E = mc^2$?' when evaluating counterfactuals.

(2) A belief-relevant world is one in which the conventions and institutions are the same as (or as similar as possible to) those in our world. (This rules out evaluating the following counterfactual for your belief that putting on the brakes will stop the car in a world in which one stops on green lights and goes on red: "If the traffic light were red and you intended to obey the law, you would put on the brakes." This counterfactual would be false in a world in which you believed that putting on the brakes will stop the car but the convention was to stop on green lights.) Confining belief-relevant worlds to those with conventions like ours avoids evaluating intuitively true counterfactuals as false.

(3) A belief-relevant world is one in which there are no interfering factors (like evil demons) that prevent S from acting when the circumstances mentioned in the antecedent obtain. (Martin, 1994) (This constraint rules out cases in which you believe that foxes have tails, but every time you are asked whether foxes have tails, for example, an evil demon makes you say 'no.' This constraint also rules out evaluating the counterfactual in a world where there are natural interfering factors, such as your having a heart attack between the time that you were asked and the time that you would have answered.) Excluding worlds with interfering factors (like evil demons), again, allows intuitively true counterfactuals to turn out to be true.

This list does not exhaust the constraints, but it does give a flavor of how the account of belief can meet challenges to the counterfactuals.[2] The constraints on belief-relevant worlds are intended to be a way to take background assumptions into consideration without trivializing the account.

In sum: According to Practical Realism, whether or not someone has a certain belief depends on what she would do, say, or think under various circumstances. The relevant counterfactuals are not counterfactuals about brain states, but about action, broadly construed as what one does, says or thinks. This is not behaviorism; belief is not identified with behavior. Nor is it mentalism; belief is not an internal brain state. The conditions for S's believing that p are exhausted by the set of associated counterfactuals. This is the sense in which belief lacks 'depth.' However—and this is my major point—to lack depth in this sense is in no way to be ontologically inferior. Believing that p is as respectable a property as having two legs.

Before considering dispositions, let me compare the Practical-Realist view of belief with Daniel Dennett's view.[3] There are two similarities between Dennett's and my views: The first similarity is that we both deny that beliefs should be identified with brain states. The second similarity is that we agree that "there is a real pattern being described by the terms of folk psychology" (Dennett, 1991, p. 119)—prominent among which is belief. But here the similarities end. In the context of Dennett's physicalism, I think that it is fair to say that Dennett regards physical properties as ontologically superior to what belief ascriptions ascribe. Ascriptions of belief and other attitudes, on Dennett's view, are interpretations in a way that ascriptions of physical properties are not.

My disagreement with Dennett can be illustrated by an example that he has given of two people, call them Brown and Jones, who both know someone well, Ella, and make different predictions about whether

2 Here is another one: S does not acquire additional information between the time that she satisfies the circumstances mentioned in the antecedent and the time that she satisfies the consequent. (For example, every time you are asked for a certain person's phone number, you may be able to produce it by looking it up in the directory. You do not have the belief that the number is 123-4567, but the following counterfactual is true of you: "If you are asked what the phone number is and you understand the question and you want to cooperate, you say, '123-4567.'" Jay Garfield and Joe O'Rourke pressed this example on me.

3 Dennett, 1991. I have commented on Dennett's program elsewhere: Baker, 1987; Baker, 1994.

Ella will commit suicide. Brown and Jones have different overall interpretations of Ella; although their predictions of Ella's behavior through the years have been generally accurate, sometimes one or the other, or both, have been wrong. Now Brown predicts that Ella will commit suicide, and Jones predicts that she will not. On Dennett's view, there may be no fact of the matter from the intentional stance about which one had the better overall interpretation of Ella. Suppose that Ella did commit suicide; her suicide would not thereby show that Brown had the better overall interpretation. Jones would chalk up his mistaken prediction to "noise" in the pattern—just as Brown has when he has been wrong in the past. Dennett's point—which he takes to be the point of Quine's indeterminacy thesis—is this: "There could be two interpretation schemes that were reliable and compact predictors over the long run, but which nevertheless disagreed on crucial cases." (Dennett, 1991, p. 117) Perhaps if Brown and Jones had gone into enough detail about Ella, they would have found sufficient evidence that favored one interpretation over the other. But, Dennett suggests, "to have gone into those details earlier would have been to drop down from the intentional stance to the design or physical stances." (Dennett, 1991, p. 117) The intentional stance is useful as far as it goes; but it plays out, and the only way to improve one's prediction of Ella's suicide, says Dennett, is to "drop down" to the design or physical stance.

But note: Ella's death would not have been a suicide if Ella had not intended to kill herself. But intending to kill oneself is not even *visible* from the physical stance. One can hardly be in a better position to make a prediction from a vantage point that renders what one wants to predict invisible. My objection is that the intentional phenomenon whose occurrence is the object of prediction just disappears altogether if we drop down to the design or physical stance. Dennett's two-fold point is that there may be two conflicting interpretations for which no further evidence from the intentional stance level could decide between them, and that the disagreement in prediction could be settled from a lower stance. I take issue with this point: A physical-level explanation or prediction just aborts the intentional-level *explanandum* altogether; it does not decide between two seemingly-adequate but nonequivalent intentional explanations of an intentional phenomenon.

As friendly as Dennett is to intentional explanations, he is fully committed to the "depth" metaphor. He says that "if one wants to predict and explain 'actual, empirical' behavior of believers, one must [...] cease talking about belief, and descend to the design stance or the physical stance for one's account." (Dennett, 1978, p. 22) More pointed-

ly, "physical stance predictions trump design stance predictions which trump intentional stance predictions—but one pays for the power with a loss of portability and a (usually unbearable) computational cost." (Dennett, 1991, p. 119) But it is not just that physical-stance explanations are more fine-grained than intentional-stance explanations; rather physical-stance explanations have different kinds of *explananda* from intentional-stance explanations.

It is as if Dennett were saying: "The intentional stance is fine if you don't need the more accurate understanding that you can get from the design or physical stances." That's like saying that looking at the Mona Lisa from 5 ft away is fine if you don't need the more accurate understanding that you can get from a microscope; but, of course, you see it better with a microscope. This would be a joke. You can see the paint better, but you do not see the Mona Lisa better through a microscope. You lose the Mona Lisa altogether. What you see by ordinary observation of The Mona Lisa is not just a summary of brush strokes. You can't just drop down a level for a better view. The patterns seen from the intentional stance are not just a kind of rough sketch of what is really going on the level of the physical stance. So, despite certain affinities between my account of belief and Dennett's, the two views part ways over the "depth" metaphor.

Now let us broaden the scope and compare belief, on the Practical-Realist view, to paradigmatic physical dispositions.

Beliefs and Dispositions

It is a commonplace to say that beliefs are dispositions to act. Right on the surface, there are similarities between beliefs and familiar dispositions like solubility and fragility. Beliefs, like solubility, are manifested in various some circumstances. Second, someone may have a belief even when the person is not manifesting it; a fragile thing is fragile even when it is not breaking. Despite these pretheoretical similarities between belief and solubility, my account of belief departs from others' accounts of physical dispositions like solubility or fragility. I shall suggest that the construal of beliefs as dispositions, with dispositions construed on the model of fragility is an instance of the "depth" metaphor at work. I want to point out some disanalogies between belief and fragility, on standard physicalistic views. I shall use David Lewis's account of dispositions as an example. Consider an ordinary fragile glass; it is struck and it breaks. Lewis suggests that the breaking was caused jointly by the striking and by some property F of the glass. F is an intrinsic property

of the glass which joins with the striking to cause the breaking. Call this property F, a causal basis for the fragility of the glass.[4]

Here is Lewis's analysis, which he characterizes as "an unlovely mouthful!": "Something x is disposed at time t to give a response r to stimulus s if and if, for some intrinsic property B that x has at t, for some time t' after t, if x were to undergo stimulus s at time t and retain property B until t', s and x's having of B would jointly be an x-complete cause of x's giving response r." (Lewis, 1997, p. 157) An x-complete cause is "complete insofar as havings of properties intrinsic to x are concerned, though perhaps omitting some events extrinsic to x." (Lewis, 1997, p. 156)

I want to show that this account—whether or not it is the correct account of physical dispositions like fragility—is not adequate for psychological dispositions generally. Not only belief, desire, and expectation, but also psychological dispositions like generosity or irritability fail to fit this account of dispositions. One reason that this kind of account cannot capture psychological dispositions is that it requires that for each disposition, there be an intrinsic causal basis.[5] If believing is a disposition and if this analysis of dispositions applies to believing, then S's believing that, say, Rome is in Italy must have an intrinsic causal basis. The obvious, perhaps the only, candidate for a causal basis for S's believing that Rome is in Italy is that S is in a particular brain state which would be part of a joint cause of each of S's disparate actions that is a manifestation of that belief.

Consider a case of Jane's applying for a job because she believed that she was well-qualified for it. Suppose that the stimulus s—analogous to the glass's being struck—is Jane's reading a description of the job. (I am assuming that Jane's reading a job description is suitable as a stimulus. If a stimulus like the one denoted by 'Jane's reading a job description' is disallowed on the grounds that it presupposes intentionality or that it is intensional, then the account is a nonstarter for psychological dispositions). Omitting qualifications, the account that applies to fragility would entail that Jane has an intrinsic property—call it 'J'—such that if Jane were to read a certain job description without losing the property J, then her reading the description and her having of J would jointly cause her applying; it would be "Jane-complete." Is it

4 For example, see Lewis (1997), p. 149. Lewis is by no means alone in his commitment to the thesis that all dispositions have causal bases. See Prior et al (1982). See also Quine,1960) and Mackie,1973.

5 For an intriguing argument that "dispositions do not need to be propped up by non-dispositional properties," and that dispositions are not causally impotent, see Ellis and Lierse, 1994.

plausible that Jane has such an intrinsic property J such that J, together with the reading of the job description causes Jane's applying for the job? Could J, together with the reading of the job description, be a Jane-complete cause of Jane's applying for the job?

Notice that if Jane's reading the job description and her having of J jointly caused her applying, then the having of J must be sufficient for Jane's *wanting* the job. If she had not wanted the job, then she would not have applied; the joining together of her reading the description and her having an internal property J would not jointly cause her to apply unless J were sufficient for her wanting the job. A Jane-complete cause of Jane's applying for the job must include Jane's desire for the job as well as her belief that she is well-qualified. But if J is sufficient for her wanting the job, then J cannot supervene on Jane's intrinsic properties. In order to want the job, a person must have a number of social and economic concepts (e.g., the concept of employment, of money, etc.). Having such concepts requires that Jane be in a certain kind of social and economic environment. Having the intrinsic properties that she has does not require that Jane be in a certain kind of social and economic environment. Therefore, her having these concepts does not supervene on her intrinsic properties. In that case, there is no intrinsic property (like J) sufficient for her wanting the job, and hence, there is no intrinsic property that, when joined to her reading the job description, causes her to apply.[6]

There is no such event as applying for the job in the absence of a certain kind of environment—social, economic, linguistic. It's not just that Jane *would* not apply for the job unless she were in a certain kind of environment. Rather, Jane *could* not apply for the job unless she were in a certain kind of environment: There is no such phenomenon as applying for the job apart from the right kind of environment. It's not just that chimpanzees never apply for jobs because they don't need to work. Rather, in a chimpanzee environment, nothing would count as applying for the job.

Contrast the case of the cause of Jane's applying for the job with the case of the cause of the glass's breaking. In the case of the glass, there are extrinsic conditions for its breaking—e.g., the glass must not be wrapped in thick velvet. But such extrinsic conditions in the case of the glass's breaking are familiar *ceteris paribus* conditions. Even with *ceteris*

6 McKitrick, 2003, persuasively defends extrinsic dispositions—e.g., vulnerability, visibility, recognizability. If she is correct, then Lewis's analysis is generally inadequate—in nonpsychological as well as psychological cases. Extrinsic dispositions do not have intrinsic causal bases.

paribus conditions, we still have an informative description of the breaking in terms of intrinsic properties. By contrast, in the case of Jane's applying for the job, the social and economic conditions that must obtain for Jane to apply for the job are not mere *ceteris paribus* conditions. They are constitutive of *any* event of applying for the job. They must obtain in order for it to be possible that *anything* is an application for the job. These social and economic conditions are extrinsic to Jane. Nothing about Jane's intrinsic properties makes it the case that what Jane does is to apply for the job. Therefore, the cause of Jane's applying for the job cannot be confined to Jane's intrinsic properties. The role of environmental conditions (determined by social and economic properties) in causing Jane's applying for the job is quite different from the role of environmental conditions (determined by the immediate surroundings of the glass) is in causing the glass to break. In the case of Jane's applying for the job, the environmental conditions determine the identity of the event in question (applying for a job); in the case of the glass's breaking, the environmental conditions do not determine the identity of the event in question (a breaking of a glass).

Someone may object that all Lewis is claiming is that the stimulus together with the intrinsic property are only internal causes of the behavior. Of course, it may be claimed, the behavior could not occur without the appropriate environment. That is, Lewis only claimed that the stimulus and the intrinsic property are "a Jane-complete cause" of the behavior, where an x-complete cause is "complete insofar as havings of properties intrinsic to x are concerned, though perhaps omitting some events extrinsic to x." (Lewis, 1997, p. 156) This observation does not blunt the point: Properties extrinsic to Jane are the "nature" of applying for a job. In this case, Jane's intrinsic properties, together with a stimulus, do not come near being the cause of Jane's applying for the job. What is missing is not just "some events extrinsic" to Jane; what is missing is any part of the cause that would make it the case that what Jane did was to apply for the job.

In my opinion, philosophers have generally misplaced confidence in intrinsic properties. For example, Lewis is so committed to the idea of a disposition's having an intrinsic causal basis that if it turned out that the causal basis of electrons' repelling each other (i.e., having negative charge) was an extrinsic property involving the aether, say, he would say that electrons are not disposed to repel each other after all.[7]

7 Lewis, 1997, p. 155. Lewis offers thought experiments—which I found unconvincing—for the need for intrinsic causal bases of dispositions.

Something is repellent if there is some kind of thing that it repels. Since 'repellent' is clearly a dispositional term pretheoretically, we would need a very strong reason (not just the preservation of intrinsicness) to accept an account that has the consequence that 'repellent' might not be a disposition when applied to electrons. Presumably, 'repellent' would be a disposition in other cases with intrinsic causal bases (like insect repellents). But to say—as Lewis's account would force us in the example envisaged—that 'repellent' is dispositional in the case of insect repellents but not in the case of electrons is theoretically unsatisfactory. It seems to me that attachment to intrinsic properties at such a high price is another manifestation of the power of the "depth" metaphor.

If we think of a disposition in terms of a stimulus/response pair, then S's belief that p is not a disposition; at best, it is an huge number of dispositions—one for each counterfactual in the associated set of counterfactuals for S's belief. And if we think of dispositions in terms of stimulus/response pairs requiring a single intrinsic causal basis, S's belief has little in common with dispositions. It is thus important not to assimilate belief to paradigmatic physical dispositions like solubility, brittleness, and fragility. To do so leads straight to the "depth" chamber that I am at pains to avoid.

Nevertheless, it is plausible to consider beliefs to be dispositions, provided that we notice the difference between beliefs and physical dispositions like fragility. In addition to the absence of an intrinsic causal basis for each belief, the case of belief has complications that fragility lacks. The counterfactuals sufficient for someone to have a belief are relativized both to the person and to the proposition. There is no one set of associated counterfactuals for a belief that p true of all believers that p. There is no single response (like breaking) or even a single set of responses that manifest a given belief. In the case of belief, unlike the case of fragility, there are no normal conditions in which a fairly simple stimulus produces a fairly simple response.

There are two further dissimilarities between physical dispositions and psychological dispositions. In the first place, as we saw in the case of Jane's applying for the job, the counterfactuals for belief (but not counterfactuals for physical dispositions) invoke conditions that are intentional—that is, conditions that presuppose that there exist beings with propositional attitudes. I doubt that we will ever be able to replace 'reading the job description' and 'applying for the job' with nonintentional descriptions, nor that we will ever have nonintentional specification of Jane's belief that she is well-qualified for a particular job. In the second place, beliefs and other attitudes are holistic: Whether a stimulus

produces a response depends on the believer's other attitudes and on what the believer believes about the stimulus. (See Chisholm, 1957) (E.g., a prickly believer who believes that rudeness should be met with rudeness would respond to an insult with a sharp remark—but only if she believed that she had been insulted. An insult that the prickly believer did not regard as an insult would not elicit the response of a sharp remark. Whether or not a glass breaks when struck does not require the glass to "regard" the striking one way or another.) So, there are significant differences between beliefs and physical dispositions. However, if we understand dispositions in a bare, schematic way—the meaning of the attribution of a disposition is exhausted by certain counterfactuals[8]—then beliefs in this broad sense are dispositions.

Beliefs and Causal Explanations

I anticipate the following response to this discussion of psychological dispositions as not requiring intrinsic causal bases. The respondent may say: So much the worse for psychological explanations. If beliefs lack a causal basis in a person's brain, then perhaps we should conclude that beliefs do not figure in correct causal explanations of her actions. In this section, I want to defend belief ascriptions as figuring in causal explanations of action.

There is widespread agreement now that beliefs do not supervene (or at least not usually) on the believer's intrinsic properties. (This just follows from externalism about belief, which has been ably defended by many.) Let us say that a property is extrinsic if it is possible that there be microphysical duplicates, x and y, such that x has it and y does not. Extrinsic properties include not only those properties expressible as two-place predicates: 'is two feet from' or 'is married to'. They include also those properties expressible as one-place predicates: 'is a felon' or 'is in debt' or 'is a suspect.' Whether or not one is a felon or is in debt or is a suspect is completely independent of one's intrinsic properties. Almost all social, political, legal, economic, etc. properties are extrinsic.

Some extrinsic properties are causally-explanatory properties: They make a difference in what happens. Your being arrested for speeding is an obvious causal explanation of the increase in your car insurance rate. Your giving a bribe causally explains your getting the only rent-controlled

8 See Mellor, 2000. While I agree with Mellor's account of the semantics of disposition predicates, I depart from his account of the ontology of dispositions. In my opinion, the concept of belief reveals the nature of belief.

apartment available in town. Some of these causally-explanatory extrinsic properties are psychological properties, like believing that p. You wrote a check to your landlord for $1000 because you believed that you owed rent of $1000. Needless to say, other things' being equal, you wouldn't have written him the check if you hadn't believed that you owed him the money. So, the causal-explanatoriness of S's believing that p is not impugned just because believing that p is an extrinsic property of a believer.

Nevertheless, many charge that without a brain state to appeal to, beliefs cannot be causally explanatory. Given that *all* of our social, political, economic, etc. explanations lack backing of "deeper" explanations, it seems just arbitrary to suppose that belief-explanations must be backed by "deeper" explanations. Here I have a partial ally in Dennett, who defends his account against the charge that on it, beliefs are just epiphenomena. The charge, he counters, stems from a simplistic notion of causation. When one finds a certain kind of predictive pattern, "one has ipso facto discovered a causal power—a difference in the world that makes a subsequent difference testable by standard empirical methods of variable manipulation." (Dennett, 1991, p. 112) For example, we can test the causal powers of a sign saying "Free lunch" by comparing its power to attract people in the window of a New York restaurant and its power to attract people in the window of a Tokyo restaurant. It seems undeniable that the variations in its crowd-drawing power is predictable and explainable at the intentional level. No deeper explanation is needed or available.[9]

Let me conclude the discussion of the causal-explanatoriness of belief by showing how a Practical-Realist account of belief can handle two problems:

I. Consider how a Practical-Realist can handle a situation in which an agent has two beliefs that are both good candidates for explaining a bit of behavior. What makes it the case that the agent acts on one of the beliefs and not the other? For example, suppose that Jones is a rich landlord who wants to cut taxes on the rich. Suppose also that we explain Jones's vote against the school budget by his belief that the school budget would raise taxes on the wealthy. Also, suppose that Jones believed that the school budget had misplaced priorities; it was bloated with support for frivolous courses like Latin and Greek and provided too little for the football team. Now suppose that the associated

9 One of the things that makes Dennett unique is that his metaphysical commitment to the "depth" metaphor does not does not interfere with a plausible construal of causal explanation.

set of counterfactuals for each belief—that the school budget would raise taxes on the wealthy and that the school budget had misplaced priorities—both contained this counterfactual: "If Jones were to vote on the school budget, he would vote against it." Now we have two equal candidates for explaining Jones' vote against the school budget. Does the Practical-Realist, with sole reliance on counterfactuals, have resources for holding one of the beliefs but not the other to explain Jones's vote?[10]

Yes. Further use of counterfactuals will distinguish which belief was the actual cause, or whether Jones's vote was overdetermined. Without considering the counterfactual—"If Jones were to vote on the school budget, he would vote against it"—ask: Would Jones have voted against the school budget if he had not believed that it would raise taxes on the wealthy, but still had believed that the school budget had misplaced priorities? Would Jones have voted against the school budget if he had not believed that it had misplaced priorities, but still had believed that it would raise taxes on the wealthy?

If the answer to the first question is yes and the second is no, then the explanation of Jones's vote was that he believed that the school budget had misplaced priorities. If the answer to the first was no and the second was yes, then the explanation of his vote was that Jones believed that the school budget would raise taxes on the wealthy. If the answer to both questions was yes, then his vote was overdetermined; and if the answer to both was no, then neither belief explained his vote. So, we do not need to appeal to brain states to distinguish between equally good candidates for an explanation of Jones' vote against the school budget. Counterfactuals alone will distinguish between the two beliefs without descending to brain states. Of course, we may be unsure about whether Jones would have voted against the school budget if, for example, he had not believed that it would raise taxes on the wealthy; but then we are often unsure about motivation—our own and others'.

II. The second problem stems from the fact that one of the counterfactuals in the set associated with Jones's belief that the school budget would raise taxes on the wealthy is this one: If Jones were to vote on the school budget, he would vote against it. If this counterfactual is in the associated set for 'Jones believes that the school budget would raise taxes on the wealthy,' then it seems that Jones' belief is "conceptually" connected to his vote against the school budget and hence cannot explain the vote.

10 Sehon, 2000, also advanced this kind of case against my view in "An Argument Against the Causal Theory of Action."

To prevent this appearance of causally explaining an action by a belief conceptually connected to the action that it is to explain, we can remove the counterfactual that mentions the action to be explained from the set. As long as that counterfactual—If Jones were to vote on the school budget, he would vote against it—was not necessary for Jones's belief that the school budget would raise taxes on the wealthy, then removal of that counterfactual would free up ascription of Jones's belief in a causal explanation. In general, if S's believing that p is to figure in a causal explanation of S's doing A, then there must be at least one conjunction of counterfactuals sufficient for S's believing that p that makes no mention of S's doing A. This restriction insures that S has the belief that p independently of doing A.

In sum: Belief-explanations are causal explanations. But belief-explanations are not stand-ins for deeper explanations in terms of brain states or internal mechanisms generally. Indeed, given the role of the environment in making possible the *explananda* of belief-explanations, it is difficult to see how belief explanations could give way to explanations in terms of brain states or internal mechanisms. Belief-explanations are justified by their systematic and ineliminable role in our successful explanatory and descriptive practices. The apparent need for further support from below stems, I believe, from the illusion of depth.

Depth and the Illusion of Depth

In the physical sciences, the metaphor of depth is a prod to further discovery. In explanation of rational behavior, the "depth" metaphor is a dead-end. The reasons are not far to find. First, beliefs and other attitudes are individuated by content, and many contents are social. That is, many beliefs are such that it is logically impossible that they be entertained in a world without other people (who also have beliefs). Beliefs about employment, finances, appointments that we have made, politics—you name it—are social in this way. And so are the actions that they explain. Although I cannot argue for it here, I think that it is unlikely that social phenomena can be reduced to properties of individuals that do not themselves presuppose social phenomena. (See, for example, Ruben, 1984/85). Human mentality is ineliminably social—look at almost any work in developmental psychologist—and there is no "deeper" level underlying social phenomena.

The second reason that the metaphor of depth is inappropriate when trying to understand belief is that beliefs rationally, as well as causally,

explain behavior. And, as Davidson has argued, the norms of rationality are not reducible to physical constraints. (Davidson, 1980) No physical property makes an action or belief rational. The third reason that the metaphor of depth is inappropriate for belief ascription is that the "depth" metaphor finds its home in application to things that are understandable in terms of their parts. But S's believing that p is not reducible to, nor deducible from, any information about S's proper parts and their relations to each other. Belief-explanations are autonomous because people are not just the sums of their parts. Persons are embedded in social and physical environments, that are not just causally, but are logically, implicated in persons' having the psychological properties that they have. (Baker, 2000) Therefore, to understand belief, the metaphor of depth is out of place.

Conclusion

Theoretical inquiry—philosophical or scientific—takes place against a backdrop of what might be called 'everyday phenomena.' Everyday phenomena are thoroughly and ineliminably intentional. In descriptions and explanations of everyday phenomena, ascriptions of belief and presuppositions of belief are ubiquitous. More specifically, everyday phenomena include things like believing that the pointer is at 14 kg., like believing that if you push such-and-such button, you run a simulation program. Such beliefs are presupposed by scientific inquiry. If science reveals what is "deep," then there is a sense in which the deep (or rather its discovery) presupposes the shallow. So, to deny *depth* to belief is not to derogate it.

The upshot is that the reality of belief does not require us to find intentionality among the likes of spin, charm and charge. The "depth" metaphor, so fruitful in the natural sciences, leads us astray elsewhere. There is no deeper reality to belief than the patterns of life that we are all familiar with, but that's reality enough.[11]

Lynne Rudder Baker
Department of Philosophy, 352 Bartlett Hall
University of Massachusetts
130 Hicks Way
Amherst, MA 01003-9269
USA
lrbaker@philos.umass.edu

11 Thanks to Jay Garfield and to Samuel Guttenplan for helpful comments.

References

Baker, Lynne Rudder (1987) *Saving Belief: A Critique of Physicalism* (Princeton: Princeton University Press)

Baker, Lynne Rudder (1994) "Content Meets Consciousness," *Philosophical Topics 22*: 1–22

Baker, Lynne Rudder (1995) *Explaining Attitudes: A Practical Approach to the Mind* (Cambridge: Cambridge University Press)

Baker, Lynne Rudder (1999) "What is This Thing Called 'Commonsense Psychology'?" *Philosophical Explorations* 2: 3–19

Baker, Lynne Rudder (2000) *Persons and Bodies: A Constitution View* (Cambridge: Cambridge University Press)

Chisholm, Roderick (1957) *Perceiving: A Philosophical Study* (Ithaca: Cornell University Press)

Davidson, Donald (1970) "Mental Events" in *Experience and Theory*, Lawrence Foster and J.W. Swanson, eds. (Amherst: University of Massachusetts Press): 79–101

Davidson, Donald (1980) *Essays on Actions and Events* (Oxford: The Clarendon Press)

Dennett, Daniel C. (1978) "Intentional Systems" in *Brainstorms: Philosophical Essays on Mind and Psychology*: 3–22

Dennett, Daniel C. (1991) "Real Patterns," *Journal of Philosophy* 88: 27–51

Ellis, Brian and Lierse, Caroline (1994) "Dispositional Essentialism," *Australasian Journal of Philosophy* 72: 27–45

Fodor, Jerry A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge MA: MIT/Bradford)

Kripke, Saul A. (1972) *Naming and Necessity* (Cambridge MA: Harvard University Press)

Lewis, David (1997) "Finkish Dispositions" *The Philosophical Quarterly* 47: 143–158

Mackie, J.L. (1973) *Truth, Probability, and Paradox* (Oxford: Oxford University Press)

Martin, C.B. (1994) "Dispositions and Conditionals," *Philosophical Quarterly* 44: 1–8

McKitrick, Jennifer (2003) "A Case for Extrinsic Dispositions," *Australasian Journal of Philosophy* 81 (2003): 155–174

Mellor, D.H. (2000) "The Semantics and Ontology of Dispositions," *Mind* 109 (2000): 757–780

Prior, E.W., Pargetter, R. and Jackson, F. (1982) "Three Theses about Dispositions," *American Philosophical Quarterly* 19: 251–257

Quine, W.V.O. (1960) *Word and Object* (Cambridge MA: MIT Press)
Ruben, David-Hillel (1984/85) "Social Properties and Their Basis," *Proceedings of the Aristotelian Society*: 23–45
Sehon, Scott R. (2000) "An Argument Against the Causal Theory of Action Explanation," *Philosophy and Phenomenological Research* 60: 67–85