

WARRANT AND DESIGNING AGENTS:
A REPLY TO JAMES TAYLOR

(Received 1 July, 1991)

James Taylor argues that my account of warrant — that quantity enough of which, together with true belief, is sufficient for knowledge — is faulty. I'd like first to thank him for his searching questions; they have certainly given me something to think about. However, I also think those questions have answers.

Now roughly speaking, and subject to a lot of qualifications, my account of warrant goes as follows:

(Warrant) *S*'s belief that *p* has warrant for *S* if and only if *p* is formed by *S*'s cognitive faculties functioning properly in an appropriate epistemic environment according to a design plan that is successfully aimed at truth.¹

By way of objection, Taylor offers a dilemma:

I hope to have established that if the concept of proper functioning is analyzed in terms of an actual designing agent or process, then accounts of warrant employing this concept are, although interesting and novel, subject to at least one counterexample. In addition, I have argued that if the concept of proper functioning is not analyzed in terms of an actual designing process or agent, then either Plantinga's account of warrant does not clearly constitute a superior alternative to reliabilism, or it reduces to a version of reliabilism.

Accordingly, the two essential premisses of Taylor's dilemma:

(1) if the concept of proper function is analyzable in terms of an actually designing agent, (whether a conscious designer or impersonal process, such as evolution²) then my account is subject to counterexample,

and

(2) If the concept of proper function is not analyzable in terms of an actually designing agent, then either my account reduces to a version of reliabilism or at any rate it is not a superior alternative to reliabilism.

Two initial problems loom. First, of course, there are a thousand different ways in which the concept of proper function might be thought analyzable in terms of an actually designing agent. Does the actually designing agent have to know what he is doing? Does he have to intend to bring about the result he does bring about? Could he design something by accident? Couldn't something be designed by a committee, or by a series of designers (like some of the medieval cathedrals), and would that count? Does the designer have to know much of anything about how the object that results from his design works? There are plenty of other questions, but perhaps we can simply note their existence and try to conduct the discussion in such a way that we need not answer them.

The second initial problem is more interesting. To appreciate it, we must first understand why someone might proceed in Taylor's dilemmatic fashion. Perhaps a reason is as follows. Suppose you analyze some property P in terms of a property Q . Now suppose further that it is fairly plausible to think that Q itself is analyzable in terms of some third property R — but also fairly plausible to think that it *isn't* analyzable in terms of R . I assume that Q is analyzable in terms of R , and show that under this assumption, your analysis of P is at best doubtful. You then point out that my assumption that Q is analyzable in terms of R is doubtful, and do not consider yourself refuted. I then assume that Q is *not* analyzable in terms of R and show that under *that* assumption your account is also doubtful. You now point out that *this* assumption is doubtful too, and once more refuse to consider yourself refuted. In exasperation, I say: "Look: either Q is analyzable in terms of R or it isn't. Either way your account of P is doubtful. Therefore your account of P is doubtful."

Am I right? No. Suppose your account is absolutely impeccable: you analyze P in terms of *itself*, or some other property Q whose equivalence to P is wholly and utterly obvious. It could still be that both the conjunction of your analysis of P in terms of Q with the proposition that Q is analyzable in terms of R , and the conjunction of your analysis of P in terms of Q with the proposition that Q is *not* analyzable in terms of R — it could be that both of these are doubtful. The doubt in each case could emanate from the doubtfulness of the second conjunct. Suppose it is doubtful that whales evolved from small land creatures —

but also doubtful that they *didn't*. Then both *All men are mortal and whales evolved from small land creatures* and *All men are mortal and whales did not evolve from small land creatures* are doubtful; it doesn't follow that there is any doubt about *All men are mortal*.

The problem here really arises as follows. The conjunction of your analysis of *P* in terms of *Q* with the proposition that *Q* is analyzable in terms of *R* is indeed doubtful: but it doesn't follow that if *Q* in fact *is* analyzable in terms of *R*, then your analysis of *P* in terms of *Q* is doubtful. Similarly, then, for my account of warrant in terms of proper function. Perhaps the conjunction of that account with the proposition that proper function is analyzable in terms of an actually designing agent is problematic; it doesn't follow that if proper function in fact is analyzable in terms of an actually designing agent, then my account of warrant in terms of proper function is problematic. But then it could be that both

warrant is analyzable in terms of proper function and proper function is analyzable in terms of an actually designing agent

and

warrant is analyzable in terms of proper function and proper function is *not* analyzable in terms of an actually designing agent

are dubious and problematic; it doesn't follow that my account of warrant in terms of proper function is dubious or problematic.

Bearing this possibility in mind, suppose we consider Taylor's (1): how does the argument for it go and what is the alleged counterexample? As a matter of fact, there are *two* counterexamples. According to the first, it is possible that a knower (a being that has beliefs some of which have a high degree of warrant) come into existence, not as a result of evolution, nor of the intentional creative activity of God, but as a result of unusually careless or incompetent creative activity on the part of someone else. (You set out to make a refrigerator, but through really monumental incompetence make something that works just like a toaster.)

Suppose God appointed a powerful but clumsy angel to create a few things (things incapable of warranted belief). In the process, the angel botches one creative attempt and Theodore is the unintended result. Though Theodore has not been designed, given

that his cognitive functioning is like that of a normal human, it seems possible that he have warranted belief.

Here there is an initial puzzle: the clumsy angel of the example *is* an actually designing agent (even if an exceedingly clumsy one); hence this case is not one in which there is a knower not designed by an actually designing agent. For present purposes, however, I think we can ignore this point. The structure of this part of Taylor's argument is as follows: according to my account, warrant necessarily involves proper function; but if proper function is to be analyzed in terms of an actually designing agent (evolution or God, for example), then the Theodore case is a case where there is warrant but no proper function; so if proper function is to be analyzed in that way, the Theodore case is a counterexample to my account of warrant.

There are many interesting issues worth pursuing here, but I want to pursue only one. Is it really clear that the Theodore case is indeed possible? My account of warrant is an account of what it is for a person's beliefs to have warrant; if the Theodore case is to be a counterexample, Theodore must be a *person*; but is it really possible for someone (even an angel) to create a person unintentionally and by virtue of titanic incompetence, intending to create something not capable of warranted belief, but winding up, oddly enough, having created a person? I haven't any reason to think this possible. I'm not positive that is *impossible*; but I am inclined to think it is, and hence don't consider it a counterexample to my account. To put it another way, the Theodore case isn't at all clearly possible; so I don't mind if my account implies that it isn't possible. It is true, I think, that one has *some* inclination to think it possible that a being capable of warranted belief should pop into existence "by chance" or by virtue of an incompetent agent's trying to create something quite different. This needn't be thought of as casting doubt on my account of warrant, however; maybe instead it casts doubt on the idea that proper function requires a reasonably successful and not wholly incompetent designer.

I think Taylor himself should come to the same conclusion. He recounts some reasons ((a)–(e)) for thinking that the Theodore case isn't possible, says "it is not clear that each of (a)–(e) should be granted. Most if not all of them are controversial" (p. 189), and adds that

it is therefore “reasonable to believe, at least, that the Theodore case has not been shown to be impossible” (p. 189). So far, so good; but he goes on to say

Since I am not confident about each premise in this argument for the impossibility of unintended persons, I am inclined to believe that, whether persons are material things or immaterial things, it is possible for them to come to exist without having been intended, and hence, without having been designed (189).

(We must note parenthetically that what is at issue is not the possibility of unintended persons, but the possibility of persons who are both unintended (created by virtue of whopping incompetence) and also not a product of evolution.) But suppose someone offers you reasons (a)—(n) for the proposition that it is impossible to square the circle; suppose (a)—(n) are controversial and you are not confident of them; should you conclude that in fact it *is* possible to square the circle? Surely not. Even if you are convinced that the argument in question is the *best* argument for the impossibility of squaring the circle, the most you can properly conclude is that it has not so far been shown to be impossible. Suppose someone offers inconclusive and controversial reasons for thinking it impossible that there be a necessarily existent being than which none greater is possible; should you conclude that in fact this *is* possible? If you do, you will find yourself (granting S5 like modal assumptions) committed to the conclusion of the ontological argument.

Taylor goes on to say that “If I am right about this, then it is reasonable to affirm that the Theodore counterexample shows that the nature of our origin, whatever it might be, is not conceptually linked to our capacity to have warranted beliefs.” But this seems much too strong. What we have is that it hasn’t been shown that the Theodore case is *not* a counterexample to the claim that the nature of our origin is conceptually linked to our capacity to have warranted beliefs; but that doesn’t make it reasonable to affirm that the alleged counterexample shows that the nature of our origins is not conceptually linked to our capacity to have warranted beliefs. At most, it makes it reasonable *not* to affirm that the nature of our origins *is* conceptually linked to our capacity to have warranted beliefs. You propose an analysis of some concept; I propose a counterexample; you think that the counterexample isn’t in fact possible, and offer reasons for thinking it isn’t; I don’t accept the reasons. I can then properly continue to hold that your analysis is

mistaken; but I can't properly claim that my counterexample *shows* your analysis mistaken. To show that an account is mistaken, an alleged counterexample must be *clearly* possible; and the fact that someone gives controversial reasons for supposing it impossible is insufficient for its being clearly possible.

So far, therefore, the most that Taylor should allege against my account is that there is a state of affairs *S* such that if *proper function* is analyzable in the above way, then my account implies that *S* is impossible, when it is not independently clear that *S* is impossible. But I don't see this as much by way of an objection to an analysis or account. Indeed, it can be a positive advantage: if you are convinced that the notion of proper function is analyzable in the way Taylor suggests, and furthermore you accept my account of warrant, then you have a way of coming to see something you couldn't see otherwise.

Taylor has a second counterexample, this one involving a person's acquiring a new cognitive power "by accident". Clarence goes into a coma; when he wakes, he "has acquired the capacity for precognition and clairvoyance"; this is not as a result of any intentional activity on the part of anyone, but strictly a matter of chance. It just happens.

Now from the theistic point of view that Taylor and I share, the idea of a thing's coming to be strictly by chance is at best suspicious: how are we to think of it? According to theism, God not only creates the world, but constantly upholds and sustains it, by way of a conserving activity apart from which creation would vanish like a dream upon awakening. He therefore takes an active hand in whatever happens in his creation; and, of course, for any time *t* he knows what will happen at *t*. But is there then room for Clarence to acquire these new cognitive powers just by chance? Would God confer these powers on him, but do so in a way that essentially involves some kind of randomizing element? How is that to be understood? Is this really possible? It isn't clear to me that it is. But it also isn't clear that it *isn't*; so let's suppose for the moment that it is possible. If it is, is it the case that "if the concept of proper function is to be analyzed in terms of an actually designing agent . . .", then we have here a counterexample to my account?

I'm inclined to doubt it. Suppose I have a portable radio; I take it up on the roof with me to keep me company while I clean the eaves troughs. I clumsily knock it off the roof; it falls some 15 feet to the

ground. I expect it to be broken; oddly enough, however, it does everything it did before, and in addition now receives stations from much farther away — eastern Europe, say. It “works better” as we would probably say, than it did before. Naturally I’m pleased (if a bit puzzled) by this new way of working. After a couple of weeks, however, it stops working this new way, backsliding to its old. I am both disappointed and perplexed; I claim that it is no longer working properly; I want to fix it. Of course I don’t know why it started working in the better way in the first place, so I don’t have a clue as to how to fix it. For want of a better idea, I take it back up to the roof and knock it off again — whereupon it once more receives those eastern European stations. I then tell my wife that the radio is fixed and works properly again.

Now in the paradigmatic and central cases of proper function, a thing works properly when it works the way it was designed to work: that is, a conscious agent sets out to make something that will perform a given function, and this thing works properly when it performs that function and does it the way in which it was designed to do it. In the case of the radio that falls off the roof, we don’t have this structure: here (after the fall) the radio doesn’t work the way it was designed to, but does perform the same function, and in fact does it better. Still, I’m right to say that the radio works properly again after I drop it the second time, and wasn’t working properly after it stopped working the new way. In this case, its new way of working gets *adopted*, somehow, as its design plan. It thus acquires a new design plan: a new way in which it is supposed to work; a new way in which it works when it is working properly. And can’t we think of Taylor’s case (if it is indeed possible) along the same lines? Clarence’s cognitive system now (by chance) works differently from the way it did before: this new way of working gets adopted by us (and no doubt by Clarence and possibly by God) as a new design plan; his cognitive faculties are now functioning properly when they function in accord with this new design plan, and hence his precognition and clairvoyance can constitute knowledge. As I see it, therefore, two things are especially interesting about this example of Taylor’s: first, it brings out the relativity of proper function to design plan; and second, it illustrates some of the subtlety of the possible relations between designer and design plan.

Accordingly, (1) of Taylor’s original dilemma is false or doubtful —

at least if we understand its antecedent in such a way that it is plausible. To defend my account of warrant, therefore, I don't strictly speaking need to go on to consider the second premiss of his dilemma, i.e.,

(2) If the concept of proper function is not analyzable in terms of some agent, then my account is not a superior alternative to reliabilism.

Interesting issues arise here, however; so I want to say something about (2). The antecedent, says Taylor, gets divided into three cases: (a) the notion of proper function isn't analyzable at all, (b) it is analyzable, but only in terms of 'internal' conditions and (c) it is analyzable, not in terms of an actual designing agent, but instead by way of a metaphorical extension of the concept. Thought of this third way, the notion of proper function is perhaps something like a useful fiction.

Taylor discounts (a); he thinks the concept of proper function is indeed analyzable. And (b) figures into his argument only in the following way: if we think (b) is true, then we may find it plausible to suppose that a being *with a design plan* could pop into existence just by chance; in the same way, if we think (b) is true, we may find it plausible to think that a being could acquire a *new* design plan by chance. If so, then we will not be moved by the alleged counterexamples of the first part of his paper.

A new problem, however, so he says, would emerge for my view. This is that a counterexample I use to reliabilism will no longer *be* a counterexample. Why so? I employ the "Case of the Epistemically Serendipitous Tumor" as a counterexample to Fred Dretske's account of warrant.³ I use the same case in explaining "the problem of generality" that arises for the early Alvin Goldman's account. According to the early Goldman, the degree of warrant enjoyed by a belief depends upon the degree of reliability enjoyed by the belief producing process that produces it. Here we are to think of process *types*, not process tokens, says Goldman. But of course there will be *many* process types involved in the production of a given belief: vision, night vision on the part of a 45-year-old male human being, and so on. Which is the relevant type? Well, at any rate the relevant type will have to be narrow: narrow enough so that all of its outputs have the same degree of warrant. But if we take the relevant types thus narrowly enough, then clearly there will be reliable process types that do not confer warrant:

Consider, for example, the person whose belief that he has a brain tumor is caused by his brain tumor. There is a rare but specific sort of brain tumor, we may suppose, such that associated with it are a number of cognitive processes of the relevant degree of specificity, most of which cause its victim to hold absurdly false beliefs. One of the processes associated with the tumor, however, causes the victim to believe that he has a brain tumor. Suppose, then, that *S* suffers from this sort of tumor and accordingly believes that he suffers from a brain tumor. . . . Then the relevant type, while it may be hard to specify in detail, will certainly be highly reliable: but surely it is not the case that this belief — the belief that he has a brain tumor — has much by way of positive epistemic status for *S*.⁴

Now *apropos* of this objection to the Goldmanian and Dretsikian versions of reliabilism, Taylor says

But here is the crucial question: what is to prevent it from being the case that, in acquiring the brain tumor, Jane has also acquired a new cognitive design plan? If Jane did acquire a new design plan, or if her existing design plan changed, then her tumor-beliefs could have been produced in a way that was in accordance with her (new) design plan. . . . If there is no way to discount the possibility that Jane's design plan changed so as to sanction the tumor mechanism, as there appears not to be, then Plantinga's counterexample against reliabilism and the motivation he derives from it for his view of warrant lose their force. If this is the case, then it is unclear what advantages Plantinga's view has over reliabilism.

The suggestion seems to be that if in fact it is possible to acquire new design plan by chance, then this counterexample to reliabilism fails, in which case, so the suggestion apparently goes, my view is not a superior alternative to reliabilism. But here I think there is misunderstanding. The versions of reliabilism to which the above are relevant are such that it *follows* from them that in the envisaged conditions, the person in question (call her 'Jane') would have knowledge. By way of the counterexample in question, I argue that it is *possible* that under those conditions Jane *not* have knowledge and the beliefs in question not have warrant for her. Now what Taylor really points out is that under my own account (and assuming that it is possible to acquire a design plan or a new design plan by chance) it is *possible* in those circumstances, that Jane *have* knowledge, for it is possible that she has acquired a new design plan here just by chance. Perhaps that is possible (and perhaps it isn't), but of course even if it is, that has no bearing on the success of my counterexample. On the views of Goldman and Dretske, it *follows from the description of the case* that Jane has knowledge; that is the problem for their views, since clearly it is *possible*, under those conditions, that she *not* have knowledge. Taylor points out that on my view,

together with the assumption that it is possible that a person acquire a new design plan by chance, it is also possible that Jane *have* knowledge under the circumstances described. That may or may not be true, but it is neither here nor there.

Finally, Taylor considers the possibility that one might analyze proper function in counterfactual terms, perhaps thinking of the analyzans as a sort of useful fiction:

Given this counterfactual analysis of the concept of proper function, the resulting partial analysis of the concept of warrant is that a belief is warranted only if it is produced by a cognitive process which is working the way it would work were it to have been designed and were it to be working according to design.⁵

Here Taylor points out first that there are *many* ways in which a thing might work if it were designed; he notes next that the multiplicity can be reduced by the suggestion that our cognitive faculties are working properly when they are working the way in which they *would* work if the theistic story were true. He then goes on to note that if the theistic story *is* true, then there seems to be a certain determinate way in which our cognitive faculties work when they work properly. This would be their *design plan*; and on the theistic story, this design plan would specify or encapsulate the way God designed those faculties to work. Let 'C' refer to this way of working — the way in which our faculties work when they work properly, the way they work when they work in accord with our design plan. He then makes the following remark:

Given this, we can now employ a line of reasoning . . . to show that one construal of Plantinga's view reduces to reliabilism. If the members of C are required for warrant, then the partial view of warrant which incorporates the concept of proper functioning analyzed counterfactually boils down to this: A belief is warranted only if it is produced by properly functioning cognitive equipment. Cognitive faculties function properly if and only if they are functioning as they would were they to have been designed by God and were they to be functioning as God would have designed them to function. Cognitive mechanisms are functioning in this latter way if and only if they possess the properties which are members of C. Hence, a belief is warranted only if it is produced by cognitive faculties which instantiate the properties which are members of C. . . . the appeal to proper function in the analysis of warrant is unnecessary. Reliability plus the other properties in C and not proper functioning are the concepts doing all the explanatory work.

Now first, there seems to be a logical problem here. We start with a *necessary* condition for warrant: "a belief is warranted only if it is produced by properly functioning cognitive equipment". In the next

sentence Taylor suggests that the condition in question is equivalent to another condition in which there is no reference to proper function. He concludes the appeal to the notion of proper function in the analysis of warrant is unnecessary. But all that we have learned, strictly speaking, is that there is a *necessary condition* of warrant that makes no reference to proper function; and that will be so whether or not an *analysis* of warrant requires a reference to proper function. (Thus a necessary condition of a belief's having warrant is that it be *held* by someone or other.)

Of course it might be that an analysis of warrant contained several conjuncts, only one of which involved the notion of proper function; then, barring a logical nicety or two, it might be that if *that* conjunct is equivalent to one that makes no reference to proper function, then the proposed analysis is equivalent to one that makes no reference to proper function. But what is required is that the two conditions be equivalent in the *broadly logical* sense; *material* equivalence is insufficient; but it is only the latter that we have in the present case. We can see this as follows. Taylor's suggestion seems to be this. Take the way our faculties function when they *do* function properly (the specification of *that* way of functioning is the design plan) and call it S. Now say simply that a belief has warrant if and only if (a) it is produced by faculties functioning according to S, and (b) S is reliable (i.e., beliefs formed by faculties functioning according to S will be for the most part true). Thus you altogether bypass the reference to proper function.

But there is a very serious problem with this suggestion. Functioning the way *our* cognitive faculties function (when they function properly) is not a broadly logically necessary condition of cognitive faculties' functioning properly; functioning according to the *human* design plan is not necessary for warrant. It is possible that there be angels or Alpha-Centaurians with cognitive faculties that function *properly*, but function in a way very different from the way *our* cognitive faculties function; hence functioning according to *our* cognitive design plan isn't necessary for cognitive proper function. (If in fact there *are* such creatures, functioning according to our cognitive design plan isn't even materially equivalent to cognitive proper function.) In the same way cardiac health on the part of a hummingbird, say, does not require having a heart that meets the specifications for a properly functioning *human* heart. If it

did, hummingbirds would be in trouble: a properly functioning human heart weighs about 100 times as much as a hummingbird is supposed to weigh. (Flying might present a problem.)

Well, could we avoid the notion of proper function by saying that a belief has warrant if it is produced by cognitive faculties functioning according to *some* reliable way of working? Here there are two problems. First, something must be said about the cognitive *environment* (and perhaps we could try saying that *B* has warrant for *S* if *B* is produced by *S*'s faculties functioning according to a way of working that is reliable in the cognitive environment in which *B* is produced.) But second, even if the set of cognitive powers is reliable *überhaupt*, it doesn't follow that every belief produced by them has warrant. Maybe some modules of the design plan are aimed, not at the production of true belief, but at something else. (For example, perhaps there is an optimistic overrider that cuts in when you suffer from serious disease, so that you think your chances of recovery to be greater than they really are; this optimism increases your chances of survival, but doesn't necessarily lead to warranted belief.) Shall we say that the belief has warrant if and only if the cognitive process or module is functioning reliably in *this very instance* — i.e., things are such that in most of the appropriate nearby possible worlds, a belief produced by this module functioning in this way in this environment is true? But this entails once more that beliefs produced by the serendipitous tumor have warrant. Shall we say that *B* has warrant if and only if it is produced by a cognitive faculty that *usually* functions reliably? Not so: my vision usually functions reliably, nevertheless some of the visual beliefs I acquire when drunk or drugged do not have warrant. Here we seem to need, once more, the notion of proper function: the problem is that in those circumstances my visual apparatus does not function properly. So far as I can see, there is no way of explaining warrant in terms of reliability; there is no way at all of bypassing the notion of proper function.⁶

NOTES

¹ In "Positive Epistemic Status and Proper Function" in *Philosophical Perspectives 2, Epistemology*, 1988, ed. James Tomberlin (Atascadero: Ridgeview Publishing Co., 1988), "Chisholmian Internalism", in *Philosophical Analysis: a Defense by Example*, ed. David Austin (Dordrecht: D. Reidel, 1988), "Justification and Theism", *Faith and*

Philosophy (special issue edited by Alvin Plantinga) Oct. 1987, "Justification in the Twentieth Century" *Philosophy and Phenomenological Research*, vol. L, Supplement, Fall, 1990, and *Warrant: the Current Debate and Warrant and Proper Function*, forthcoming, I hope, in 1992.

² Here we can ignore doubts as to whether evolution can properly be thought of as 'an actually designing agent'; that issue doesn't essentially enter into the discussion.

³ "Positive Epistemic Status and Proper Function" p. 22.

⁴ See pages 24–31 of "Positive Epistemic Status and Proper Function"; this quotation is from page 30.

⁵ Kant considers the same suggestion:

There can be, then, purposiveness without purpose, so far as we do not place the causes of this form in a will, but yet can only make the explanation of its possibility intelligible to ourselves by deriving it from a will. Again, we are not always forced to regard what we observe (in respect of its possibility) from the point of view of reason. Thus we can at least observe a purposiveness according to form, without basing it on a purpose (as the material of the *nexus finalis*), and remark it in objects, although only by reflection.

And:

... an object, or state of mind, or even an action is called purposive, although its possibility does not necessarily presuppose the representation of a purpose, merely because its possibility can be explained and conceived by us only so far as we assume for its ground a causality according to purposes, i.e. in accordance with a will which has regulated it according to the representation of a certain rule. *Critique of Judgment*, 55/6 Translated with an Introduction by J. H. Bernard (New York: Hafner Press, 1951), pp. 54–55. These quotations are both to be found in Christel Fricke's "Explaining the Inexplicable. The Hypothesis of the Faculty of Reflective Judgement in Kant's Third Critique", *Nous* March, 1990, pp. 52–53.

⁶ For an argument, chapter IX of *Warrant: the Current Debate*.

Department of Philosophy
University of Notre Dame
Notre Dame, IN 46556
USA