NOÛS 27:1 (1993) 66–82

Why We Need Proper Function

ALVIN PLANTINGA
University of Notre Dame

First, I wish to express my gratitude to Professors Sosa and Feldman: I have learned much from their searching and carefully crafted papers. I shall begin by responding to the criticisms they make of my books; then I shall turn to broader considerations.

I. Feldman

Feldman's criticisms come in two sorts and display a nice symmetrical structure. First, there is criticism of the criticism I make of others in *Warrant: the Current Debate*; then there is criticism of my account of knowledge in *Warrant and Proper Function*. The criticism of my criticism of others itself comes in two parts. First, Feldman criticizes some of the counterexamples I use to try to show that the views of others are wanting; then second he claims that when my counterexamples are repaired, as he thinks they can be, they don't really damage the views I mean to criticize, because they don't show anything not already conceded by those who put forth the views under attack.

A. Feldman and my Counterexamples

I proposed to use the term 'warrant' for that property or quantity, whatever precisely it is, enough of which distinguishes knowledge from mere true belief. I then argued by counterexample that certain popular accounts of warrant are mistaken. According to Feldman, however, some or many of these counterexamples are defective. How so? Many of them tend towards the unduly outré; but that isn't the problem. The problem is rather that they are not given in sufficient detail; the fact is they can be filled out in such a way that they are *compatible* with the views they are directed against, so that they aren't really counterexamples at all.

This isn't a large point, since (as Feldman concedes) they can also be filled out so that they *are* counterexamples. But by way of brief self-exculpation: there are two kinds of counterexamples. Someone claims that **A** entails **B**; in the first kind of counterexample, you present a state of affairs **C** that is clearly possible, and that entails or includes both **A** and **-B.** In the second kind, you present a state of affairs **C** that clearly entails **A** and is clearly compatible with **-B**, thus showing that ***A and -B*** is possible and hence that **A** does not entail **B.** (So a counterexample of the second kind is successful if it can be filled out to a counterexample of the first kind.) A counterexample of the first kind fails if **C** is compatible with **B**; but that isn't a cogent criticism of a counterexample of the second kind. All **C** has to be, in a counterexample of the second kind, is clearly possible, sufficient for **A,** and clearly compatible with **-B.** I am encouraged to think my counterexamples of this kind were successful, since Feldman had no trouble filling them out to counterexamples of the first kind.

So suppose we turn to the second and central criticism here: that when my counterexamples are properly filled out, they don't really damage the views I mean to criticize. This, says Feldman, is because they turn into *Gettier* examples when filled out; and thus they fail to show anything not already conceded by those who advance the views against which they are directed. Exactly how does this go? In *Warrant: the Current Debate,* I criticize internalist and justificationist views. Among them is *evidentialism,* an important view with a long and distinguished history, and a view endorsed by Feldman himself.[1] On this view, a belief is justified for **S** if and only if (in Feldman's version) it fits **S**'s evidence. Now to evaluate evidentialism, it is crucially important to decide what is to count as evidence. First, there is *propositional* evidence: I have propositional evidence for one of my beliefs if it is evidentially supported by one or more other beliefs I hold.[2] Second, there is *phenomenal* evidence. One kind of phenomenal evidence is *sensuous* evidence, for example, sensuous imagery of the sort involved in visual (and other sorts of) perception. What we have in these cases is 'the evidence of the senses'—the phenomenological sensuous evidence furnished by way of being appeared to in those characteristic ways involved in perception.

Now if we construe 'evidence' in such a way that it is limited to propositional and sensuous evidence, then *fitting the evidence* isn't even necessary for warrant. For in large and important areas of our cognitive establishment, there is very little by way of sensuous imagery, and what there is is not connected with warrant. Consider memory. You remember meeting Paul in California last month; but there need be little by way of sensuous imagery. For some people, apparently, there isn't any sensuous imagery; and for the rest of us, the sensuous imagery there is doesn't play an evidential role.[3] (You don't remember that it is *California* where you saw Paul by virtue of the sensuous imagery that accompanies the memory; the same sensuous imagery might accompany the memory that you saw him in Arizona.) But then one of your beliefs might fail to fit your evidence (restricted as above) even though it has warrant. Perhaps, for example,

you have propositional evidence for the proposition that you were in New York City last Thursday (you are nearly always there on Thursdays and you learn that Paul says he saw you there then); if you remember spending the entire day at home in Los Angeles, then the belief that you weren't in New York last Thursday may have warrant for you, even though it doesn't fit your evidence (again, evidence construed in the narrow way). For there isn't any relevant *sensuous* evidence, and the relevant propositional evidence you have supports your having been there then; yet you may know that you weren't there. So if evidence is limited to propositional and sensuous evidence, then *fitting your evidence* is not necessary for warrant.

But perhaps there is still another kind of phenomenal evidence. The sensuous imagery I mentioned isn't the only kind of phenomenology that can plausibly be thought of as evidence. Consider memory again: in addition to the sensuous imagery there is also something like a certain *felt attractiveness* of the memory belief; it *feels* right, somehow, and other beliefs you might consider in its place feel wrong. You remember that it was *California* where you met Paul, not New York; and that it was *Paul* you met, not Sam. You remember: *that was Paul I met there in California.* That proposition has about it a sense of rightness, or fittingness, or appropriateness—as opposed, say, to the proposition that it was Sam you met on that occasion. Perhaps the thing to say is that there is a sort of felt inclination, or impulse, to accept the one proposition, as opposed to the other; so call this sort of evidence "impulsional evidence".) Perhaps impulsional evidence is no more than the phenomenal reflection of the fact that you do indeed believe the proposition in question; in any event, there is this kind of phenomenology involved in memory belief. Accordingly, in the case of memory beliefs there are really two quite different sorts of experience: the fleeting, indistinct, unstable, sometimes random sensuous imagery, on the one hand, and, on the other, the felt inclination or impulse to believe *that* proposition, as opposed to others that might suggest themselves. Many other kinds of beliefs display both these kinds of phenomenal evidence, and, as far as I can see, all beliefs involve impulsional evidence.

Construe 'evidence' broadly, therefore, in such a way that it includes impulsional evidence as well as propositional evidence and sensuous imagery: then it is indeed plausible to say that you have warrant for a belief only if you have evidence for it and only if the belief *fits* the evidence, whatever exactly this *fitting* relation is. But in criticizing evidentialism I added that your belief's fitting your evidence (in this broad sense) isn't nearly *sufficient* for warrant. The reason is that a person could be such that one of his beliefs fits his evidence, even though, because of cognitive malfunction, that belief has no or almost no warrant for him at all. Here we can obviously appeal to brain in vat and Cartesian demon examples; but if the fancifulness of such examples displeases you, there are also plenty of real life examples. My father has been a victim of manic depressive psychosis (bipolar disorder) for many years; he often winds up spending time in Pine Rest Christian Psychiatric Hospital in Grand Rapids. The last time he was

there, he met a man who complained that he wasn't getting the credit he deserved for inventing a new form of human reproduction, "rotational reproduction" as he called it. This kind of reproduction doesn't involve sex. Instead, you suspend a woman from the ceiling with a rope and get her rotating at a high rate of speed; the result is a large number of children, enough to populate a city the size of Chicago. As a matter of fact, he claimed, this is precisely how Chicago *was* populated. He realized, he said, that there is something churlish about insisting on getting all the credit due one, but he did think he really hadn't got enough recognition for this important discovery. After all, where would Chicago be without it?

Clearly this set of beliefs could fit his evidence (construed thus broadly); they may fit perfectly. Perhaps it seems to him both that he remembers reading about how the method of rotational reproduction populated the whole city of Chicago, and also that he invented it, though he didn't get credit for it. Even if these beliefs do fit his evidence, however, it is clear that they have little or no warrant; they don't have the property or quantity enough of which is sufficient, together with true belief, for warrant. The problem is not with a failure of fit between his beliefs and what is internally available to him: the problem is that his faculties are not functioning properly; he is insane.

Now Feldman is inclined to agree with all this, and here we come to his second criticism of my counterexamples. He claims that when my examples are properly filled out, what they show is nothing that should disturb an evidentialist; they are really Gettier examples. True: they show that fitting the evidence (or justification) isn't sufficient for warrant; but we have known *that* ever since Gettier's 1963 paper. The evidentialist doesn't claim that evidential fit *is* sufficient for warrant; she admits that a fourth (and perhaps externalist) condition is necessary:

> Thus, my second point about Plantinga's criticism of rival theories is that several of his examples, as he describes them, fail to show any defect in these theories. Contrary to what Plantinga says, they are examples in which the believer is not following his evidence or is not believing dutifully. So, the examples as presented don't show that these properties are insufficient for epistemic warrant. There may be variations on the examples which do show that evidence and epistemic duty fulfillment are insufficient for epistemic warrant (In Plantinga's sense). However these examples are simply variations on Gettier style cases, to be dealt with however such examples are generally dealt with (p. 39).

So we don't have here a criticism that should disturb an evidentialist: filling out my counterexamples of the second type to counterexamples of the first type only yields Gettier cases, and we already knew that a fourth condition is needed to deal with Gettier cases.

By way of reply: first the examples I gave aren't strictly speaking Gettier examples, because (as in the rotational reproduction case) the beliefs in question weren't stipulated to be *true;* but that's only a quibble. Second, examples of this kind—examples involving insanity, or brains in vats, or people deceived by

Cartesians evil demons—don't seem much like Gettier's original example of the person who lies to you about having a Ford, thus inducing in you the justified belief that he has a Ford, which belief turns out, unbeknownst to him, to be true, so that your belief is both justified and true. In those cases there is a sort of local hitch or glitch in the epistemic environment: the deceiver is himself deceived in an unexpected way. Here we have a bit of retail lack of fit between epistemic environment and your cognitive faculties; but in the brain in vat, insanity and evil demon examples we have wholesale epistemic failure. If these latter *are* Gettier examples, then I suppose *any* example that shows evidential fit or justification insufficient for warrant will be a Gettier example. If so, then my examples, I guess, are Gettier examples; I don't want to argue about what constitutes a Gettier example, or try to give necessary and sufficient conditions for them. (If I did, Gettier would probably find a counterexample!)

The real point lies in a different direction. I had thought that justificationists think justification necessary and (with true belief) *nearly* sufficient, so to speak, for warrant; the basic idea or basic picture is that knowledge is justified true belief. True: this turns out not to be exactly right; a codicil must be added to take care of annoying little counterexamples of the type Gettier proposed; but the main structure of knowledge is given by justified true belief. Now the examples I proposed were meant to show that this picture is deeply mistaken. It isn't that justification is nearly sufficient for warrant, and needs only the addition of a comparatively minor if hard to state fourth condition; these examples show that justification (construed evidentially) isn't anywhere *nearly* sufficient for warrant. Consider again the man in Pine Rest who thought he had discovered rotational reproduction: he might very well be both deontologically and evidentially justified; but his beliefs don't come anywhere near having warrant.

But suppose the evidentialist retorts that there are no near misses in logic: either a proposed *analysans* is correct or it isn't; there is no such thing as getting it nearly right; all this talk about a basic picture is just picture thinking; we've known for the last 30 years that we need a fourth condition; and fourth conditions can't be said to vary in size or significance. Then it seems to me that the evidentialist faces a dilemma. If we take the notion of evidence *narrowly,* so that it includes propositional evidence together with sensuous phenomenal evidence, but does not include nonsensuous phenomenal evidence (impulsional evidence), then evidentialism doesn't give us even a necessary condition of warrant: for in many cases of warranted belief—many cases of memory and *a priori* belief, for example—there really *isn't* any significant sensuous phenomenal evidence. On the other hand, if you take evidence broadly, so that it includes nonsensuous phenomenal evidence, then *fitting your evidence* is arguably necessary for knowledge; but it adds nothing significant to the belief condition. You have impulsional evidence for **p** just by virtue of *believing* **p.** It isn't even clearly possible that you believe **p** but lack impulsional evidence for it: could it be that you believe **p** although it doesn't seem to you to be true?

If so, however, what the evidentialist really needs is not a *fourth* condition, but a *third*. For evidential justification (construed thus broadly) guarantees only belief and a necessary accompaniment of belief: the nonsensuous phenomenal counterpart of the belief, the sense that the proposition believed is indeed true. So the proposition that **S**'s belief that **p** fits **S**'s evidence can be satisfied by **S**'s merely believing **p**. Take evidence the narrow way, therefore: then evidentialism makes a contribution to the question of warrant by focusing our attention on areas of the whole complex epistemic establishment where indeed warrant and properly functioning faculties do involve evidence of this type; taken this way, however, evidentialism gives us neither a necessary nor a sufficient condition of warrant. On the other hand, take evidence the broader way: then evidentialism states a condition that is necessary for knowledge, all right, but only because this condition is already guaranteed by belief itself. The evidentialist condition, taken this way, adds nothing significant to truth and belief. So it isn't that the evidentialist is on the right track, giving us a third condition for knowledge, and conceding that he still needs a fourth; he *doesn't* give us a third condition, since what he gives us guarantees no more than is guaranteed by truth and belief (the first two conditions) alone.

Similar remarks apply to Feldman's defense of deontologism: the idea that epistemic duty fulfillment is necessary for warrant, and together with true belief, nearly sufficient for it. Cases of massive malfunction—cases of madness, brain in vat cases, Cartesian demon cases, and the like, show that duty fulfillment isn't anywhere nearly sufficient (together with true belief) for warrant; it isn't even in the right neighborhood. These cases show, I think, that justification taken thus deontologically has little to do with warrant. You can be as justified as you please, but still hold beliefs that come nowhere near having warrant. Indeed, in this case things are even worse; it looks as if deontological justification isn't even *necessary* for warrant, as I argue in chapter II of *Warrant: the Current Debate*. What I take my examples to show isn't merely that deontological justification isn't quite sufficient for warrant: what they show is that deontological justification really doesn't help us understand warrant at all. The lesson they teach is that warrant and deontological justification are radically different properties.[4]

B. Feldman and Proper Functionalism

I turn now to Feldman's discussion of 'proper functionalism', my own proposal (I don't like the name, but can't think of a better one.) *His* first criticism has to do with *my* criticism of *Goldman's* early reliabilism, according to which a belief **B** is justified (has warrant) if it is produced by a reliable belief producing mechanism (the latter taken as type, not token); and **B** has a degree of justification proportional to the reliability of the mechanism. But of course any concrete belief producing mechanism is a token of *many* types: which is the one the reliability of which determines the degree of justification enjoyed by the belief in

question? Following Feldman's lead,[5] I said the reliabilist can't find a type of the right generality here. Obviously the type in question must be so narrow that all beliefs produced by it have the same degree of justification (otherwise the degree of justification won't depend solely on the reliability of that type). But if you take the relevant type that narrowly, you have such problems as that of the epistemically serendipitous lesion.[6]

Here Feldman suggests that I am too hasty: there may well be types that will yield the right result. By way of trivial example, he points out that the relevant type could be, for example, *process producing a belief that has warrant.* Clearly, all and only beliefs produced by processes that are tokens of this type will have warrant. This seems right, but how does it help the reliabilist? You might as well propose an analysis of warrant in which the *analysans* is *warranted belief.*

Secondly, he suggests that one can state a version of reliabilism which is "almost equivalent" to proper functionalism by appropriately choosing the types involved: "One can say that the relevant type of a process token includes all and only processes governed by the same modules in the truth-seeking portion of the design plan for the cognitive system." The idea, I gather, is this: the relevant types (call them 'Feldman types') will have as tokens certain concrete processes (not process types): all those that are governed by a given truth seeking module of the design plan (or are part of a given truth seeking faculty or subfaculty of the whole cognitive establishment); and a belief **B** has warrant, on this form of reliabilism, if and only if it is produced by a mechanism falling under a *reliable* Feldman type. Says Feldman: "You thus get a specification of reliabilism almost equivalent to proper functionalism. They swim, or as I see it, sink together."

But this seems to me mistaken; the resulting version of reliabilism isn't anywhere nearly equivalent to proper functionalism, and is (dare I say "consequently"?) subject to devastating counterexample. First, it isn't equivalent to proper functionalism: on this account but not on proper functionalism it is necessary that (a) all the beliefs that are outputs of processes falling under a Feldman type have the same degree of warrant, and (b) all the beliefs in the output of processes falling under a *reliable* Feldman type have warrant. Neither of these is a consequence of my view, and neither is in fact necessary. On my view beliefs produced by the same module may have varying degrees of warrant, and *will* have varying degrees of warrant if they are held with varying degrees of strength. As for (b), note that a Feldman type is a type all of whose tokens are governed by a given module of the design plan. Now suppose, for concreteness, that *vision* is such a module. Suppose furthermore vision is a reliable type: most of the processes it governs (in fact, and perhaps also in appropriately nearby possible worlds) produce true beliefs. Then the conditions for warrant specified by Feldman's version of reliabilism are met: this version entails that if vision is a reliable Feldman type, then (necessarily) every belief produced by vision has warrant. But this is not a consequence of proper functionalism and is false besides. Due to an excess of vodka, my vision malfunctions: the malfunction causes me to believe I see pink rats crawling over my bed. Surely this belief has

little or no warrant, even though it is produced by a faculty or mechanism falling under a reliable Feldman type. So this version of reliabilism is not equivalent to proper functionalism; furthermore it is wholly inadequate just because it fails to include the condition requiring proper function.

One final point in this neighborhood: according to Feldman, there is something like a generality problem for proper functionalism, as well as for classical reliabilism. I'm sorry to say I don't really see what the problem is. (In particular I don't know why Feldman asks (p. 43), "What counts as the same circumstances anyway?") *Part* of the alleged problem seems to be that it isn't clear how to apply proper functionalism in a specific case—the case, say, of Feldman's belief, that he sees a large number of people. But what, exactly, is the problem? This belief of his has warrant, so I say, if and only if it is produced in him by cognitive faculties functioning properly in an appropriate environment according to a design plan successfully aimed at truth. So are the cognitive faculties involved in the production of Feldman's belief functioning properly and aimed at truth? I should think so. Is the cognitive environment appropriate for his kind of cognitive systems? It certainly seems so. And is there a high objective probability that a belief produced by these faculties (the ones involved in the production of Feldman's belief) functioning properly in an appropriate cognitive environment will be true? I see no reason to doubt it. So I don't think I see the problem here. Is it perhaps that in many cases we don't really know what the modules involved are? This is certainly true; but my claim is only that our concept of warrant is such that if a belief has warrant, the module or modules producing it is (are) reliable. In order to evaluate this claim, we do not need to know just what modules our cognitive establishment does indeed contain, or which are operative in a given circumstance. In the same way, we often can't tell whether someone's belief is justified (deontologically, let's say) or whether it fits his evidence; this isn't much of an objection to analyses of warrant in terms of justification or *fitting the evidence*.

I turn finally to what Feldman calls his 'basic and central objection': this too has two parts. According to the first part, as I understand it, proper function doesn't do any work: we can do just as well with reliability alone, leaving proper function out altogether:

> The discussion of the current section has been designed to support the idea that it would have been better (or just as good) to add only the reliability requirement and forget about the proper function requirement. As far as I can tell, there are no examples that demonstrate the need to add this requirement to the reliability requirement (p. 47).

If what Feldman says here is true, then my view really reduces to a form of reliabilism—a rather baroque and unlovely form with an extraneous and wholly pointless curlicue in the form of the proper function requirement. But *is* what he says true? I think not. Reliability isn't anywhere nearly enough to guarantee warrant, and the deficiency, so far as I can see, is made up by adding the proper

function requirement. Suppose once more that vision—the whole complex power whereby one comes to have visual beliefs—is a cognitive faculty and a module of the design plan. Suppose furthermore, my vision is reliable: in this and appropriately nearby possible worlds, the beliefs it produces are for the most part true. Does it follow that all the beliefs produced by my vision have warrant? Of course not; perhaps occasionally this faculty malfunctions. Perhaps I get drunk again (above, p. 72) and see more pink rats. The belief that I see pink rats will then be produced by a reliable faculty; this belief may still be wholly without warrant, because on this occasion of its operation, my vision is malfunctioning.

Of course there will be a thousand similar examples. I ordinarily know what city I am in; if I get the measles and a really high fever, however, the faculty or process that produces such beliefs may malfunction and produce an absurdly false belief—a belief with no warrant; that will be because the faculty that produces it, while reliable, on this occasion malfunctions. A friend of mine was recently hit hard by a bear; he ordinarily remembers what blood pressure medication he takes, but on this occasion, when the medic asked him, he gave the wrong answer. His belief on this occasion was produced by a reliable belief producing faculty or power or mechanism (his memory); due to the stress of the occasion, however, his memory wasn't functioning properly on this occasion and the belief produced had little by way of warrant. Don't we see in these and other examples the need for the addition of the proper function requirement to the reliability requirement?[7]

Now consider the other part of Feldman's central objection, the claim that proper function isn't necessary for warrant: you can have warrant and knowledge, he says, even where you don't have proper function. There are really two quite different kinds of cases here, but since I am running out of time I shall deal with only one. Consider the remarkable twins Feldman (following Oliver Sacks) speaks of: although they were retarded and cognitively damaged in some ways, they could also perform marvellous feats of cognition. Not only could they 'just see' that there were 111 matches on the floor without counting them; they could also just see, so it seems, whether a given 6 and even 8 digit number is a prime. That is no mean feat. (Perhaps we could call them the marvellous prime twins.) And Feldman's claim is this: the twins' cognitive faculties are clearly malfunctioning, in producing the belief, e.g., that 111 matches fell out of the box, but that belief nonetheless has warrant for them: "It seems clear that they know how many matches there are, but they aren't designed to know that. It's a departure from the design plan that enables them to know" (p. 49).

This is an extremely interesting example, and there is much to say about it; I don't have the time to say nearly all that should be said. But let me say just this much. First, it isn't clear that they do know the propositions in question. Consider their being able to tell (so it seems) that **n** is a prime, for some large **n**. According to Sacks, the twins don't seem to have the concept of multiplication; it is therefore not at all clear that they can form the belief ***n** is prime,* for some six

digit **n,** or indeed for any **n** at all. Due in part to the serious damage to their cognitive faculties, one doesn't know enough about what they actually believe, what beliefs they can and cannot form, to be sure precisely what beliefs they do and do not have. There is less uncertainty in the case of their being able to tell that 111 matches fell out of the box; even here, however, there is some residual uncertainty: do they really know what it is for there to be 111 of something if they don't have the concept of multiplication?

It isn't quite clear, therefore, that these twins have the relevant beliefs; but it is much more doubtful, I think, that their faculties really *are* malfunctioning in producing these beliefs. Their faculties obviously seem to malfunction in *some* ways; but how do we know that they are malfunctioning in producing the belief that there are 111 matches on the floor? Mozart (and perhaps a few others) could hear in his mind's ear a complex piece of music in complete detail and then just write it down; the rest of us can't; must we conclude that Mozart's cognitive faculties were malfunctioning (that there was a departure from his design plan) when he did this? There is no need to draw that churlish conclusion; another perfectly possible view is that he had cognitive powers most of the rest of us lack, and hence a slightly (or more than slightly) different design plan. Perhaps these twins and a few others can perform marvellous feats of calculation and cognition: must we conclude that their faculties are malfunctioning in producing these beliefs? Again: there is no need to draw this conclusion. Their faculties certainly malfunction in *some* ways; but why insist that they are malfunctioning in producing these specific beliefs? Indeed, suppose you knew these twins, spoke with them often, tried to understand them and how their cognitive faculties work, witnessed a hundred times the fact that they can apparently just see that there are **n** matches on the floor for large **n,** without counting them as you and I would have to do. Would you take it for granted that their faculties were malfunctioning in producing these beliefs? I doubt it. I certainly wouldn't. Suppose that one of them contracts a serious disease; his temperature goes up to 107 degrees; upon recovering from the fever he thinks he can still see how many matches there are, but most of the time comes up with a totally wrong answer. (Sometimes, when there are 111 matches there, he says there are 15, but other times he gets it right.) Then you would be inclined to think this mysterious power of his has been damaged; some faculty or power he has now malfunctions part of the time. At least that would be my inclination.

How could it be that the twins' faculties (and Mozart's) are working properly when they are working so differently from our own? Well, we aren't just given that all human beings have exactly the same design plan. We aren't just given that the twins don't have a design plan slightly different from ours: perhaps God has given them a design plan a bit different from the one we have. We also aren't just given this if we think there is no God; new design plans obviously arise in some way or other (perhaps by virtue of some of the mechanisms appealed to in current evolutionary theory), and we can account for this case as we do for those.[8]

II. Sosa

I turn now to Ernest Sosa's typically interesting and gracious paper. First, he is quite right when he points out on page 52 that the reason the serendipitous tumor (or lesion) example does not afflict my view is that a person who suffers from the tumor is such that her cognitive faculties are not functioning properly, i.e., in accord with her design plan. She is subject to malfunction, disease, dysfunction.

A. Sosa's Counterexamples

Suppose we consider Sosa's Swampman and Swampbaby examples, the adult version of which he borrows from Donald Davidson. According to this example, Davidson is standing in a swamp; lightning strikes a dead tree near him; his body is vaporized. By the merest chance or coincidence, the tree is turned into a molecule by molecule physical replica of Davidson; the replica swims back to Davidson's house and takes up work on an article on radical interpretation where Davidson left off. (And Feldman thinks *my e*xamples are weird!) Sosa proposes that this replica would have beliefs, and would indeed have knowledge; its beliefs would have warrant. If so, however, proper functionalism must be false: since he has no design plan, the Swampman's faculties would not function either properly or improperly.

Sosa suggests that I would reply to this counterexample as I replied to a similar example suggested by James Taylor: it isn't at all clearly possible, in the broadly logical sense, I said, that a being capable of belief should suddenly pop into existence just by chance or coincidence.[9] Sosa then points out that this isn't clearly impossible either, and makes a very interesting proposal. Suppose **A** is an analysis of some concept or property **P**: **A**, of course, is successful only if it states (broadly logically) necessary and sufficient conditions of **P.** According to Sosa, there can be two kinds of counterexamples to **A**. First, there are *refuting* counterexamples; such a counterexample presents a state of affairs **S** meeting two conditions: (a) **S** is clearly possible, and (b) the *analysans* holds and the *analysandum* doesn't hold (or vice versa) in **S,** so that *analysans* and *analysandum* are not equivalent in the broadly logical sense. But second, there are also *opposing* counterexamples. An opposing counterexample presents a state of affairs **S** which isn't clearly possible but also isn't clearly impossible, and is such that if it is possible, then clearly the *analysans* and the *analysandum* do not both hold in it. And then Sosa claims that

> While a refuting counterexample constitutes a particularly resounding defeat for a philosophical thesis, a merely opposing counterexample also does epistemic damage, the extent of the damage being proportional to the degree to which it seems plausible that the example is really possible (p. 55).

Now *something* like what Sosa says seems to me to be true; as it stands, however, I am inclined to doubt that what he says *is* true. (Perhaps it is a sort of first approximation.) This is because it suggests that every opposing counterexample

(the kind where it isn't clear that the example is impossible) does some epistemic damage to the proposed philosophical thesis. This seems to me mistaken. It might be that the example isn't clearly impossible, but only because it is very complicated: one knows that even if the example *were* impossible, it wouldn't be *clearly* impossible, just because it is too close to the limit of our capacities for detecting the impossible. (Here it is important to distinguish between failing to see that **S** is impossible and seeing that **S** is possible.) In other cases we have the same pattern, but not due to complication. A theist believes that it is impossible that any substance other than God continue to exist without God's supporting and conserving activity; she shouldn't be impressed by the suggestion that it isn't clearly impossible that there be a substance that continues to exist apart from the supporting activity of God. Here the problem isn't complication, but just the fact that this state of affairs' not seeming clearly impossible to us isn't much by way of evidence for the claim that it is indeed possible. A certain sort of materialist thinks pain just is a kind of neural event; should he be impressed by the suggestion that it doesn't seem clearly impossible that there be pain in the absence of that kind of neural event? I don't think so, and for the same reason: here too the fact that the envisaged state of affairs is not clearly impossible is not a good reason for thinking it possible.

And of course the same goes for counterexamples to analyses. You present an analysis; I outline a a state of affairs which is neither clearly possible nor clearly impossible, but is such that if it *is* possible, then it is clearly a counterexample. This doesn't necessarily damage your analysis: everything depends upon the nature of my example. We have a damaging counterexample only if we have reason or it seems plausible to suppose that we *do* see, even if only in an indistinct sort of way, that **S** is possible, rather than merely failing to see that **S** is impossible. My truthfully reporting "I can't see why that couldn't be" doesn't show much. I can't see why there couldn't be gold with an atomic weight of 14; nothing much follows. In some cases our failing to see that something couldn't be true is a good reason for thinking that it could be true (cases where, if the state of affairs were impossible, we would be able to see that it is), but in others it isn't. To make Sosa's principle work, we must distinguish these cases.

Accordingly, suppose we return to Swampman. What we have here is a molecule by molecule replica of Davidson: furthermore this thing behaves just as we should expect Davidson (or his body) to behave. Many would take it for granted that it is *necessary* that a *dopplegänger* of Davidson or his body would have beliefs if and only if Davidson did. But is this really obvious? No. It *could* be, so far as obviousness goes, that Cartesian dualism is true; perhaps what we have here is a duplicate of Davidson's body, but no corresponding mind; so perhaps no beliefs of any sort at all are associated with this duplicate. Does that do epistemic damage to the above assumption? Not as it stands. All depends, here, on whether what we have is a case of seeing, perhaps dimly, that it is possible that Davidson and his dopplegänger differ with respect to belief, or merely a case of failing to see that this is impossible. (End of methodological digression.)

Now Sosa reports that he is "strongly inclined to believe that Swampman is logically and metaphysically possible". He is also inclined to think, apparently, that if there were such a thing as Swampman, it would have beliefs some of which could have warrant for it. But is this incompatible with proper functionalism? Not obviously. The problem is supposed to be that Swampman arises just by chance; perhaps we can imagine God permitting a junior angel of some sort to play an elaborate game of chance which somehow results in the new form taken by the matter from the destroyed tree. So this duplicate of Davidson isn't designed by anything—not by God, but also not by the junior angel or evolution. It just shows up by chance. As I say, I am inclined to doubt that this is possible.

But if it *is* possible, why isn't it equally possible that the being in question pop into existence by chance *complete with a design plan?* The notions of design plan and proper function are correlative: a thing is working properly, in doing **A** in circumstances **C**, if and only if its design plan calls for it to do **A** in **C**. If a being capable of having beliefs can just pop into existence by chance, couldn't the same be said for a being that is capable both of belief and also of functioning properly or improperly? (Couldn't Swampman suffer a heart attack due to overexertion from all that swimming? If he didn't, might not his heart malfunction?) Sosa says (p. 57) that "If a 'design plan' might fall into place independently of any sort of conscious or unconscious design, then I lose my grip on the meaning of the locution 'design plan'." I'm not sure I see the problem. It is at best dubious that a being capable of belief could just pop into existence by chance; but if that *is* possible, couldn't the being function well or ill? Couldn't Swampman get sick or drunk? When Swampbaby grows up, can't *she* get sick, or drunk, or injured? We would certainly think (and say) so.

B. Sosa's Virtue Epistemology

I turn finally to what Sosa sees as successful alternatives to proper functionalism; and of course I don't have nearly enough space to do justice to Sosa's thoughtful suggestions. The first is

> (W) My belief B is warranted only if it is produced in me by a faculty F in a cognitive environment such that F is working properly with respect to the goal of truth acquisition and error avoidance in environment E (p. 58).

Here 'working properly' with respect to a goal', says Sosa, means no more than working in such a way as to promote or produce that goal. This account is therefore simpler and less elaborate than my account, with its more robust sense of 'working properly':

> Why move from simply requiring that a belief must be caused by a reliable faculty if it is to have warrant to requiring more elaborately that the belief must be caused by a faculty that is not only reliable but is functioning properly in some sense that involves design by conscious agent or impersonal process? (p. 58).

I have two comments on (W). First, a comment that applies not only to (W) but to all of Sosa's suggestions; this is my most important point here. Sosa proposes that a warranted belief must be produced by a *faculty*; his reason is the following:

> Against examples like that of the brain lesion, one can now argue that they involve belief-producing processes, but nothing that could properly be called a "faculty." And we can leave for later work the problem of how to define the concept of a faculty (p. 58).

So the idea is that a belief is warranted only if it is produced by a faculty, and a brain lesion isn't a faculty. This seems right; but the problem for Sosa, I think, is that *the notion of a faculty involves the notion of proper function.* A faculty or power—perception, or memory, or reason in the narrow sense, or digestion, or one's ability to walk—is precisely the sort of thing that can function properly or improperly. Indeed, this is just the difference between the brain lesion and a faculty: the concepts of proper and improper function don't apply to the brain lesion. It isn't functioning either properly or improperly in producing the belief it does: it isn't that sort of thing. And of course this notion of proper function can't be explained just in terms of producing mainly true beliefs: the brain lesion, after all, may do that. (Maybe it causes only one belief: that you have a brain lesion.) What is involved here is precisely that more robust notion of proper function that (W) is designed to avoid. So we don't really avoid that notion in (W); we smuggle it in in the very notion of a faculty.

Second comment: If a belief is true, and is produced by a faculty, then isn't that faculty working properly with respect to the goal of truth acquisition? If so, however, it looks as if every true belief produced by a faculty has warrant, even if the faculty is malfunctioning in the old sense, the ordinary sense, in producing that belief; and that seems wrong. You get drunk; some of your belief producing faculties malfunction; one of the beliefs produced by these temporarily dysfunctional faculties happens to be true: they misfire in such a way as to produce in you, at random, the true belief that the Pope is at present in Barcelona. That belief, I would say, doesn't have warrant for you, even if as a matter of fact the Pope does happen to be in Barcelona. (W) as it stands, therefore won't do the job.

We might try to mend matters by denying that

> Belief **B** is true

is sufficient for

> the faculty that produced **B** was functioning properly with respect to the goal of truth acquisition in producing **B**;

we say instead that a faculty is working properly with respect to that goal if and only if *most* of the beliefs it produces, or some sufficiently high percentage, are

true. But no hope. For then a faculty could be working properly with respect to truth acquisition even if in one smallish area of its operation it malfunctioned (in the old sense) and produced absurdly false beliefs. I get drunk and hallucinate again: if my perceptual faculties produce truth in the vast majority of cases (I hallucinate only twice in my whole life) then on this revised account my hallucinatory belief has warrant for me. But clearly it doesn't. So the first simple account doesn't seem to me to work.

It isn't clear how serious Sosa is about (W); I believe, however that he is completely serious about

> V1 For any world w, a belief B is justified $_1$ in w if and only if B derives in w from the exercise of one or more intellectual virtues that in w virtuously produce a high ratio of true beliefs (p. 61).

V1 speaks not of warrant but of justification; but according to Sosa it gives us "at least a partial account of 'justification' (or of something close to warrant, perhaps, or of aptness), one which offers an alternative to proper functionalism..." (p. 62). But what is an intellectual virtue? Sosa doesn't explicitly say; so far as I can tell, however, a virtue is a reliable belief-producing *faculty*[10]—a faculty that produces a high ratio of true beliefs in the appropriate environment. V1 therefore has the crucial problem of (W); it presupposes the notion of proper function in the robust sense and therefore does not offer us an alternative to proper functionalism.

There are further problems with V1. For example (as Sosa recognizes) if a belief is to be justified or warranted, then the faculty that produces it must be working properly *on the occasion of its production.* Of course Sosa hopes to avoid appeal to the notion of proper function; he therefore suggests that "hence it must be that in the circumstances one would (most likely) believe P iff P were the case—i.e., one (at least probabilistically) tracks the truth... (p. 62). So this tracking requirement is to do the work of the proper functioning requirement. But the tracking requirement, I think, is too strong. Consider my belief that

> (1) I am not a brain in a vat on Alpha Centauri, serving as a subject in an experiment in which the experimenters give me the very experiences and beliefs I do in fact have just now.

This doesn't satisfy Sosa's tracking requirement: it is not the case that probably, if it were false I would not believe it.[11] (If it were false, I *would* believe it, since it is one of the beliefs I do in fact have just now.) But couldn't this belief have warrant for me? Note also that while (1) doesn't meet the tracking requirement,

> (2) I am at home in Indiana and I am not a brain in a vat on Alpha Centauri, serving as a subject in an experiment in which the experimenters give me the very experiences and beliefs I do in fact have just now

does. (If (2) were false, it would be because of the falsehood of its first conjunct, in which case I would not believe that conjunct or the whole proposition.) Can a view be right if it implies that I can know (2) but cannot know (1)?[12]

On the other hand, the tracking requirement is also too weak. Suppose we accept the usual semantics for counterfactuals: then any belief in a necessary proposition will automatically satisfy the tracking requirement.[13] Hence any belief of mine in what is in fact a necessary truth will meet this condition—even if I acquire the belief by virtue of getting drunk, so that my reason temporarily malfunctions. So any such belief (provided it meets the other conditions Sosa sets for warrant) would have warrant for me. But that can't be right.

Sosa adds a couple of further conditions:

> If a faculty operates to give one a belief, and thereby a piece of direct knowledge, one must have some awareness of one's belief and its source, and of the virtue of that source both in general and in the specific instance... . And, finally, one must grasp that one's belief non-accidentally reflects the truth of P through the exercise of such a virtue. This account therefore combines requirements of *tracking* and *nonaccidentality,* of *reliable virtues or faculties,* and of *epistemic perspective.*

These further conditions, however, also seem too strong, if taken as conditions of warrant generally. This is another indication of the complexity and articulation of our design plan; the conditions Sosa mentions may well be required for certain kinds of knowledge—perhaps for wholly mature and self-conscious knowledge—but they don't seem to be generally necessary. A 10-year-old boy, for example, may know what his name is, even though he does not have awareness of the virtue of the source of this belief, both in general and in the specific instance in question. A person could know a proposition, I should think, even if she didn't have any idea at all as to the source of the belief: perhaps all she can say is that it just popped into her head. And one can certainly know a proposition without grasping (believing ?) that the belief in question non-accidentally reflects the truth of that proposition; many people, I should think, know much without even having the concept of non-accidentally reflecting the truth.

There are problems, therefore, for Sosa's virtue epistemology. I am gratified to see, however, that insofar as an epistemic virtue is an epistemic *faculty,* Sosa's virtue epistemology is really a variety of proper functionalism. So I suggest in conclusion: we need the notion of proper function in order to give an account of warrant; there seems to be no way to do without it.[14]

Notes

[1]Richard Feldman and Earl Conee, "Evidentialism", *Philosophical Studies* 48, pp. 15 ff.

[2]Perhaps other conditions should be added; I don't have the space here to go into the matter.

[3]I don't have the space to explain this in proper detail: see pp. 57ff. of *Warrant and Proper Function.*

[4]See my "Justification in the Twentieth Century", *Philosophy and Phenomenological Research,* vol. L, Supplement, Fall, 1990.

[5]See his "Reliability and Justification", *The Monist* 68, pp. 159 ff.

[6]"There is a rare but specific sort of brain lesion (we may suppose) which is always associated with 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 associated processes, however, causes the victim to believe that he has a brain lesion. Suppose, then, that **S** suffers from this sort of disorder and accordingly believes that he suffers from a brain lesion. Add that he has no evidence at all for this belief: no symptoms of which he is aware, no testimony on the part of doctors or other expert witnesses, nothing. (Add, if you like, that he has much evidence *against* it; but then add also that the malfunction induced by the lesion makes it impossible for him to take appropriate account of this evidence.) Then the relevant type (while it may be hard to specify in detail) will certainly be highly reliable; but the resulting belief—that he has a brain lesion—will have little by way of warrant for **S**." *Warrant: the Current Debate*, P. 199.

[7]We might suggest that what the reliability condition requires is that the faculty in question be functioning reliably *on this very occasion.* This raises a host of problems, however, problems I don't have the space to investigate properly here. But just to point to their general location: first, we must say that a faculty can be functioning *reliably* even when it produces *false* beliefs; else we shall have to hold that no false beliefs have warrant. But *can* a faculty be functioning reliably on an occasion when it produces a false belief? And secondly, to avoid supposing that all true beliefs have warrant, we shall have to say that a faculty or power can be functioning *un*reliably on an occasion when it produces a *true* belief. How will this go? Perhaps we shall say that a faculty is functioning reliably, on a given occasion, only if it produces true beliefs in the appropriately nearby possible worlds. But then we face once more the problem of the epistemically serendipitous lesion.

[8]In footnote 5, Feldman suggests that "if the abnormality is the result of a lesion or radiation, it is difficult to understand the sense in which it is part of their design." But of course we aren't given that the unusual powers of these twins *are* the result of lesion or radiation. He also suggests that if we think Mozart and the twins have a slightly different design plan, then we must say the same for cases of cognitive defect: "Moreover if these cognitive abnormalities which are (in some respects) advantageous are held to be part of the design, then comparably caused defects should also be held to be part of the design. In that case, many of the examples Plantinga uses against other theories will apply equally to his own." But I fail to feel the force of this consideration. Can't it be that some deviations from the usual are by way of powers most of us don't have, while others are by way of malfunction?

There is much more to be said about examples of this kind. In some cases of this kind, a certain max plan gets adopted as a design plan (see "Warrant and Designing Agents: Reply to Taylor" *Philosophical Studies* 64 (1991) p. 209); In other cases what we have is an analogical extension of the central notion of proper function and thus of the central notion of warrant *(Warrant and Proper Function* (Hasker)).

[9]See "Warrant and Designing Agents: Reply to Taylor" p. 206.

[10]See p. 61. If an intellectual virtue were *not* a faculty but instead any belief producing process, then of course the epistemically serendipitous lesion would once more rear its ugly head.

[11]See Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981), p. 203.

[12]See my "Positive Epistemic Status and Proper Function in *Philosophical Perspectives 2, Epistemology, 1988,* ed. James Tomberlin (Atascadero: Ridgeview Publishing Co., 1988).

[13]According to the usual semantics, a counterfactual with true antecedent and consequent is true; hence if **p** is necessary and I believe it, the first half of the tracking requirement is satisfied. Further, according to the usual semantics, if **p** is necessarily true, then (trivially) any counterfactual with its denial as antecedent is true; so the second half of the tracking requirement is also satisfied.

[14]One final point: a belief has warrant, on my account, only if the module of the design plan governing its production is aimed at truth and only if that design plan is a *good* design plan. I went on to explain the goodness of a design plan in terms of the objective probability of a beliefs being true, given that it was produced by faculties functioning according to that design plan in the right sort of environment and given that the modules of the design plan governing its production are aimed at truth. Some of the things Feldman and Sosa say suggest that this reliability condition is really only a first approximation; perhaps there is more to goodness of design plan, in this context, than reliability. Perhaps they are right; if so, that would be a project for further work.