SOME REMARKS ON BONJOUR ON WARRANT, PROPER FUNCTION, AND DEFEASIBILITY

COLIN P RULOFF The Claremont Graduate School

Abstract

A number of counterexamples have recently been leveled against Alvin Plantinga's Proper Functionalism, counterexamples aimed at showing that Plantinga's theory fails to provide sufficient conditions for warrant — that elusive epistemic property which together with true belief yields knowledge Among these counterexamples, Laurence BonJour's is perhaps the most formidable and, if successful, shows that Proper Functionalism is simply too weak to serve as an acceptable theory of warrant. In this paper, I ar gue that, contrary to initial appearances, BonJour's counterexample is not successful. More exactly, I argue that, once it is recognized that a defeasibility constraint is deeply embedded within Plantinga's proper function condition for warrant — a constraint which says, in effect, that a belief B is warranted for an agent S only if S does not possess any defeaters against B — BonJour's counterexample to Proper Functionalism can be handled quite straightforwardly

A number of counterexamples have recently been leveled against Alvin Plantinga's Proper Functionalism, counterexamples aimed at showing that Plantinga's theory fails to provide sufficient conditions for warrant — that elusive epistemic property which together with true belief yields knowledge Among these counterexamples, Lau rence BonJour's is perhaps the most formidable and, if successful, shows that Proper Functionalism is simply too weak to serve as an acceptable theory of warrant (BonJour 1996) In this paper, I argue that, contrary to initial appearances, BonJour's counterexample is not successful More exactly, I argue that, once it is recognized that a defeasibility constraint is deeply embedded within Plantinga's proper function condition for warrant — a constraint which says, in effect,

[©] Principia, 4(2) (2000), pp 215–28 Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil

that a belief B is warranted for an agent S only if S does not possess any defeaters against B — BonJour's counterexample to Proper Functionalism can be handled quite straightforwardly. The presence of a defeasibility constraint within the proper function condition for warrant draws attention to an important but often neglected feature regarding Proper Functionalism, viz, that, contrary to initial appear ances, there exists an internalist constraint within Plantinga's otherwise broadly externalist theory of warrant

1 Plantinga's Proper Functionalism and BonJour's Counterexample

Let the term "warrant" designate that possibly complex property, whatever it may be, which when conjoined with true belief will yield knowledge In his Warrant and Proper Function (Plantinga 1993, here after "WPF") Plantinga argues that the core idea behind warrant is, appropriately enough, proper function More exactly, according to Plantinga, in order for an agent S's belief B to be warranted it must be the case that B is the product of a belief-forming faculty that is functioning properly There are other conditions as well The beliefforming faculty that produces B must be functioning in an appropriate epistemic environment and it must have as its primary aim the production of true beliefs Furthermore, the belief forming faculty must be reliable. That is to say, it must truth conducive. The heart of Plantinga's theory nonetheless is found in the claim that warrant requires that one's belief forming faculties be functioning properly Plantinga formulates his theory of warrant — call it "Proper Func tionalism" or simply "PF" — roughly as follows 1

PF For any agent S and belief B, B has warrant for S if and only if

- (C1) B is produced in S by a belief forming faculty that is function ing properly
- (C2) B is produced in S by a belief forming faculty that is functioning in an appropriate epistemic environment
- (C3) the segment of the design plan governing the production of *B* is aimed at the production of true beliefs

(C4) the segment of the design plan governing the production of *B* is reliable, that is, it is well-designed or truth-conducive ²

To be sure, this is a bare-bones version of PF and many important details need to be filled in, details which Plantinga does fill in in WPF, but the general idea is clear enough. According to PF, a be lief has warrant for you just in case your belief is produced in you by a belief-forming faculty that is functioning properly in an appropriate epistemic environment according to a design plan that is successfully aimed at the production of true beliefs. And if such a belief is true, it is an instance of knowledge.

Plantinga claims that PF provides the conditions that are each necessary and jointly sufficient for warrant (1993, p. 19). Now, aside from the question as to whether PF provides necessary conditions for warrant, Laurence BonJour has recently argued that PF clearly fails to provide sufficient conditions for warrant. PF fails to provide sufficient conditions, says BonJour, because it's possible to construct counterexamples where an agent S and a belief B satisfy (C1) through (C4), but where B, in the intuitive sense, fails to be warranted for S Let's have a closer look at BonJour's argument

BonJour asks us to imagine the case of Boris, a human being who is designed by God and who is in most respects normal (1996, p. 58). There is one respect in which Boris is not normal, however, and it is this, God has implanted deep within Boris's brain a very narrow and specialized belief-forming faculty designed to guarantee that Boris will have a true belief about some monumentally important future state of affairs. More precisely, the belief-forming faculty is designed in such a way that at some specific future time before the world as we know it comes to an end with the Second Coming, Boris will be caused to have the belief with "maximal firmness and conviction" that this event is about to occur (BonJour 1996, p. 58)

Now, suppose the designated time has come and that Boris finds himself with the relevant belief, viz , the belief that

(p) The world as we know it is about to come to an end

According to BonJour, conditions (C1) through (C4) of PF are all evidently satisfied. Boris's belief-forming faculty is functioning properly

in a suitable epistemic environment, the belief forming faculty governing the production of Boris's belief has as its aim the production of a true belief, and, we may suppose, Boris's belief-forming faculty is reliable. According to BonJour, however, Boris's believing p does not appear to be *warranted* (in the intuitive sense) since Boris would be thoroughly irrational and epistemically irresponsible in accepting p. It would, for example, not be

rational for Boris to act on this belief, e.g., by selling his house to buy full-page ads proclaiming the forthcoming event or canceling his life insurance policies (as of the date of the expected occurrence) (BonJour 1996, p. 59)

Indeed, according to BonJour, Boris has "lots of reasons for being suspicious of beliefs of this kind and none that is apparent for trusting them" (1996, p. 59). Although he doesn't quite state it this way, BonJour's objection to PF seems to amount to something like this p fails to be warranted for Boris because there are other beliefs that Boris already holds which function as evidence or reasons for thinking that p is false. In other words, Boris's believing p fails to be warranted because Boris has access to a number of defeaters for p. On the basis of this counterexample BonJour concludes that Plantinga's PF is too weak and, consequently, a mistaken theory of warrant

Prima facie, BonJour appears to be right, Conditions (C1) through (C4) do appear to be satisfied. Hence, the Boris counterexample does seem to show that PF fails to provide sufficient conditions for warrant. I believe, however, that a careful reading of WPF reveals that Plantinga has the resources within his theory of warrant to show that BonJour's counterexample is merely apparent. As I mentioned at the outset, once it is recognized that a defeasibility constraint is deeply embedded within the proper function condition for warrant or (C1) — a constraint which says, in effect, that a belief B is warranted for an agent S only if S does not possess any defeaters against B — one can see that the offending counterexample can be handled quite straightforwardly

Before I continue, I should point out that Plantinga nowhere in WPF explicitly states that a defeasibility constraint is to be embedded within the proper function condition for warrant. Rather, this con-

straint appears to be entailed by the workings of what Plantinga calls a "defeater system" which is *itself* subsumed under the proper function condition. So in order to derive a defeasibility constraint from this condition it will be necessary to have a somewhat detailed look at Plantinga's notion of a defeater system. To this we shall now turn

2 The Defeater System

According to Plantinga, included within the general doxastic apparatus for humans is the defeater system — that is, a belief forming faculty that manages or regulates the various changes that can take place within an agent's set of beliefs over time. More will be said later on the defeater system — on how exactly it is supposed to function and so on — but for now let us simply highlight the fact that for Plantinga the proper function condition for warrant, (C1), is to be construed broadly so as to apply to the proper functioning of one's defeater system. As Plantinga states

The defeater system works in nearly every area of our cognitive design plan and is a most important part of it. We must therefore explicitly understand the proper function condition of warrant [C1] as applying to the relevant portions of the defeater system (1993, p. 41, italics added)

And again, Plantinga says that a belief B has warrant for you only if

the cognitive faculties involved in the production of B are functioning properly (and this is to *include* the relevant defeater systems) (1993, p. 194, italics added)

For Plantinga, then, the defeater system is an essential component of our general doxastic apparatus. Moreover, and most importantly, Plantinga wants the proper function condition for warrant to *extend* to the proper functioning of one's defeater system. That the proper function condition for warrant is to extend to the proper functioning of one's defeater system is repeatedly overlooked within the literature discussing Plantinga's PF³ Indeed, it is clear from BonJour's article that BonJour *also* fails to recognize that the proper function condition

requires explicit reference to a defeater system. Let us remedy this oversight and make this feature of Plantinga's PF explicit with the following narrow "defeater system" version of (C1)

(C1_{ds}) For any agent S and belief B, B is warranted for S only if S's defeater system is functioning properly ⁴

Now, (Cl_{ds}) is fine as far as it goes but the condition needs some unpacking since it doesn't say what it is for a defeater system to be functioning properly As Plantinga notes, when something is functioning properly, it functions so as to fulfill a particular purpose (1993, p 13) Additionally, when something is functioning properly, it does not function merely to fulfill a particular purpose but also to fulfill that purpose in a particular way For example, a microwave is designed to cook, heat, and defrost food, and it is designed to do so in a particular way There will be a complex set of specifications, or what Plantinga calls a "design plan", which determines how the microwave is to fulfill its purpose Plantinga thinks that broadly the same point can be made with respect to our belief forming faculties. He thinks that our belief-forming faculties are designed, at least in part, to furnish us with true beliefs about ourselves, our environment, abstract objects, etc, But this does not happen in any old way since our belief-forming faculties are designed to function in a particular way according to a particular design plan

To ask what it means to say that a defeater system is functioning properly, then, is really just to inquire about the defeater system's design plan. So we can ask, how is the defeater system designed to function? What is its goal or purpose? From what I can gather from Plantinga's discussion on the matter, the defeater system appears to be a sort of doxastic manager or regulator and it is designed to en sure that the appropriate revisions take place within an agent's set of beliefs given that that agent has acquired any *defeaters* for those be liefs 5 With this in mind, we can formulate a somewhat more precise version of (Cl_{ds}) , (Cl'_{ds})

(C1'_{ds}) For any agent S and belief B, B is warranted for S only if S's defeater system is functioning in such a way that it will carry out the appropriate revisions within S's set of beliefs given that S has acquired any defeaters against B

Plantinga provides us with a number of examples which illustrate in an intuitive way how the defeater system is designed to function A simplified version of one of his examples goes like this ⁶ Suppose lones enters a factory and sees an assembly line on which there rest a number of widgets that appear to be red Being appeared to red widgetly, Jones, naturally enough, forms the belief that he sees a num ber of red widgets This belief has a certain amount of warrant for Jones Suppose, however, that at some later time the factory foreman informs Iones that the widgets are not red — they're actually green - but just appear that way because they are being irradiated by a red light which helps the workers in quality control to detect possible hairline fractures Since Iones now possesses a defeater for his be lief that the widgets he sees are red, according to Plantinga, Jones's defeater system will engage and specify that the appropriate doxastic response is for Jones to revise his set of beliefs and come to believe that it is false that the widgets are red Consequently, Jones no longer believes that he sees red widgets 7

The basic idea here, then, is that our belief forming faculties are designed in such a way that, for example, when you are appeared to in a certain manner you will (ceteris paribus) form a certain belief. When you are appeared to red widgetly, you will (ceteris paribus) form the belief that you see red widgets. But our belief forming faculties, and in particular our defeater system, will also specify circumstances under which you will have to revise your set of beliefs and come to believe it is false that you see red widgets despite the fact that you are appeared to red widgetly. These circumstances would include, for example, your coming to learn that the widgets, despite appear ances, are not red. So ultimately a properly functioning defeater system, along with the rest of our belief-forming faculties, aims at the production and sustenance of true beliefs and the avoidance of false beliefs. In other words, our belief-forming faculties are truth-aimed.

Thus, to say as $(C1_{ds})$ does, that S's belief B is warranted only if S's defeater system is functioning properly is simply to say that S's defeater system is functioning in such a way that it will carry out the appropriate revisions within S's set of beliefs in the *relevant circumstances*, 1 e , in those circumstances in which S has acquired any defeaters against B 10

3 The Defeasibility Constraint

We now have at least a rough idea as to what's packed into (C1) I have not yet, however, formulated the defeasibility constraint that I claim 11 is embedded within that condition. In order to do so we must take an even closer look at the workings of the defeater system Recall how the defeater system works in the case of Jones. Assuming, as Plantinga does, that Jones's truth aimed belief forming faculties are functioning properly, Jones's appearing to red widgetly at time t_1 warrants his believing

(q) The widgets on the assembly line are red

at t_1 That is to say, Jones's believing q at t_1 is warranted for him on the basis of his being appeared to red widgetly at t_1 When Jones acquires a defeater for q at t_2 , however, Jones's defeater system en gages and specifies that Jones come to believe the denial of q at t_2 Consequently, Jones no longer believes q at t_2 Moreover, and most importantly, since the warrant q had for Jones has been defeated by Jones's getting warrant for q's denial, q is no longer warranted for Jones at t_2 12

The basic idea here is that if you come to acquire a defeater for your otherwise warranted belief B at some time and your defeater system responds in the way it was designed to respond by specifying that you come to hold B's denial, then B cannot be warranted for you at that time. Stated a little more smoothly if you come to possess a defeater for your belief B at some time, then, given how your defeater system is designed to function, B cannot be warranted for you at that time. The following, thus seems to follow quite naturally from Plantinga's view of how the defeater system is designed to function.

(α) For any agent S belief B and time t if S possesses a defeater for B at t then B is not warranted for S at t

Now suppose as before that Jones's truth aimed belief forming faculties are functioning properly at t_1 that Jones comes to believe q upon being appeared to red widgetly at t_1 , but that Jones does not

come to acquire a defeater for q at the later time t_2 Since in this counterfactual case Jones fails to possess a defeater for q at t_2 , Jones's defeater system will not engage and specify that he come to believe q's denial at t_2 ¹³ Consequently, q will remain warranted for Jones at t_2 Thus, q is warranted for Jones at t_2 only if Jones does not possess a defeater for q at t_2 More generally, in order for a belief B of yours to be warranted at some time, it must be the case that you not possess any defeaters against B at that time Thus, along with (α) , the following No Defeater Constraint (NDC) seems to follow from Plantinga's view of how the defeater system works

(NDC) For any agent S, belief B, and time t, B is warranted for S at t only if S does not possess any defeaters for B at t ¹⁴

Since (NDC) is entailed by the workings of the defeater system, and since the defeater system is intended by Plantinga to be sub sumed under (C1), it follows that (NDC) is entailed by (C1) as well Hence, as promised, there appears to be a defeasibility constraint deeply embedded within (C1), and more generally within Plantinga's Proper Functionalism

4 BonJour's Counterexample Revisited

Recall BonJour's "Boris" counterexample to PF According to Bon Jour we are to imagine that God has implanted deep within Boris's brain a narrow and specialized belief forming faculty designed to guarantee that Boris will come to have the true belief that

(p) The world as we know it is about to end

Now, suppose the designated time comes and that Boris finds himself believing p with maximal firmness and conviction. According to BonJour, conditions (C1) through (C4) are all satisfied but it doesn't look like Boris's believing p is warranted since Boris has access to a number of defeaters for p—that is, reasons or evidence for thinking that p is false. The upshot according to BonJour is that, since it's possible to come up with cases in which an agent S and a belief B satisfy (C1) through (C4) but where B intuitively falls to be warranted for S, PF is simply too weak to serve as an acceptable theory of warrant

However, BonJour is clearly mistaken in thinking that (C1) through (C4) are satisfied since (NDC) — which, as we have seen, is embedded in (C1) and hence one of the conditions necessary for warrant — is clearly not satisfied Say that (NDC) is satisfied by a belief B just in case the agent holding B does not possess any defeaters against B With this definition in hand, we can see that (NDC) is obviously not satisfied in the case of Boris since, as BonJour himself seems to admit, Boris does possess defeaters against b If Boris believes roughly the same sorts of things that you and I believe, then Boris will presumably believe that it is false, or, at the very least, extremely doubtful that humans can instantiate clairvoyant powers. Boris will believe that beliefs that pop into one's head "out of nowhere", as it were, ought to be treated as doxastically suspect, Boris will believe that acting on such beliefs, say by selling one's house to "buy full page ads proclaiming the forthcoming event" or "canceling [one's] life insurance policies" would be plainly irrational, and so on. In sum. it looks like Boris has lots of reasons for thinking that b is false. If this is right, then, contrary to initial appearances, BonJour's counterexam ple simply does not satisfy the conditions that PF lavs down as each necessary and jointly sufficient for warranted belief Hence, Bonlour's counterexample fails to show that Plantinga's PF is too weak to serve as an acceptable theory of warrant

Contrary to what Bonjour seems to think, then, Plantinga's PF is not committed to holding that p is warranted for Boris. It's not warranted. And this is simply because one necessary condition for warrant, namely (NDC), is not satisfied. 15

One final point At first blush, Plantinga's PF certainly appears to be an externalist theory of warrant ¹⁶ This is how Plantinga's PF is standardly interpreted. The presence of (NDC) within (C1), how ever, suggests that things are not quite so neat and tidy. (NDC) demands that in order for a belief of yours to be warranted, it must be the case that you possess no defeaters against that belief. But this is something to which one has cognitive access. That is to say, one can tell by reflection alone whether or not one has defeaters for one's beliefs. (NDC), thus, is really an internalist constraint on warranted belief. The presence of (NDC) within (C1), thus, highlights an important but often neglected fact regarding Plantinga's PF, viz., that,

garding the internalist elements of Alvin Goldman's externalist theory of warrant — Process Reliabilism — see Thomas Senor's interesting "The Prima/Ultima Facie Justification Distinction in Epistemology" (Senor 1996) I thank an anonymous referee for bringing this article to my attention ¹⁸ My thanks to Alvin Plantinga for e-mail correspondence which led to the writing of this paper, and to Stephen T Davis and Gabrielle Ruloff for helpful comments and suggestions on an earlier draft

- 11 And which BonJour fails to recognize
- ¹² As Plantinga puts it in a slightly different context, "Here the first belief [q] gets defeated, and its warrant disappears by virtue of your getting warrant for another belief [-q] inconsistent with it" (Plantinga 1993, p 42, italics added)
- 13 Jones's defeater system will remain idling, as it were
- ¹⁴ I should add here that there are different versions of (NDC), some stronger than others. A strong version states that B is warranted for S at t only if S does not consciously believe that she possesses a defeater for B at t. A weaker version states that B is warranted for S at t only if S does not believe upon brief reflection that she possesses a defeater for B at t. For more on this issue, see Sudduth (1999, pp. 171, 184, footnote 12), and Michael Bergmann (2000)
- 15 The general line of argument that I have advanced here on Plantinga's behalf has been has been deployed by Plantinga himself in his reply to Keith Lehrer's counterexample, a counterexample that (as was pointed out in note 3) is similar to BonJour's According to Lehrer's counterexample we are to imagine the case of Mr Truetemp, an agent who, unbeknownst to him, has a small temperature-sensing device in his brain that is designed to regularly produce in him the true belief that his bodily temperature is 98 degrees. The problem, as Lehrer sees it, is that, while the example appears to satisfy Plantinga's conditions for warrant, Mr Truetemp cannot at any time claim to know that his bodily temperature is 98 degrees. Hence, PF is too weak to serve as an acceptable theory of warrant. As Plantinga points out in his reply, however, Mr Truetemp has a defeater for his belief regarding his bodily temperature. Says Plantinga

As I see it Truetemp has a defeater for his belief in the fact that (as he no doubt thinks) he is constructed like other human beings and none of them has this [temperature sensing] ability furthermore everyone he meets scoffs or smiles at his claim that he does have it Truetemp's defeater means that his belief does not meet the conditions for warrant hence (contra Lehrer) he doesn't constitute a counterexample to my analysis of warrant (Kvanvig 333)

I thank an anonymous referee for bringing Plantinga's reply to Lehrer to my attention

¹⁶ PF is *externalist* since the satisfaction of conditions (C1) through (C4) are not the sorts of things to which an agent typically has epistemic access

¹⁷ Plantinga is not the only epistemologist who incorporates internalist elements into a broadly externalist theory of warrant. For a discussion re-

A little more will be said in section 3 on Plantinga's notion of a "design plan", however

- ² For a full statement of PF, see Plantinga 1993, p 19 For an even fuller statement see Plantinga 1993, p 194
- ³ So far as I can tell, only two writers namely, Michael Sudduth and Michael Bergmann formally acknowledge that the proper function condition for warrant is to apply to the proper functioning of one's defeater system. See Michael Sudduth (1999), and Michael Bergmann (1997) Strangely, Keith Lehrer *recognizes* that the proper function condition for warrant requires explicit reference to a defeater system, but offers a Bon-Jourian-style counterexample to Proper Functionalism whose very success depends on *ignoring* this feature of Proper Functionalism. See Keith Lehrer (1996, p. 31)
- ⁴ The subscript "ds" stands for "defeater system"
- ⁵ Or, as Sudduth puts it, the defeater system is "a cognitive subsystem that is designed to regulate modifications in a person's noetic structure given new experiences and the acquisition of new beliefs which come with social exposure, mental maturation, and education In short, these are specifications as to the correct or proper ways of changing beliefs in response to experience (doxastic or otherwise)" (Sudduth 1999, p. 169)
- ⁶ I ignore here Plantinga's distinction between "undercutting" and "rebutting" defeaters. For more on this distinction and how it plays itself out in the present example see Plantinga 1993, p. 41
- ⁷ Presumably, a failure of Jones's defeater system to engage (which would result in no revisions in Jones's set of beliefs) would indicate that Jones's defeater system is *mal*functioning, 1 e, not functioning in the way it was designed to function. Generally, then, if an agent S acquires a defeater for a belief B, then *some* sort of revision has got to take place within S's set of beliefs
- ⁸ Or, at the very least, your defeater system will specify circumstances under which you come to believe *less firmly* that you see red widgets despite the fact that you are appeared to red-widgetly. For more on the *range* of responses available to one's defeater system, see Sudduth 1999, p. 169
- ⁹ Plantinga 1993, p 41 For more on defeaters and defeater systems, see Sudduth 1999, pp 169–71
- 10 As Sudduth points out (and as was pointed out in note 7), a *failure* of S's defeater system to carry out the appropriate revisions within S's set of beliefs given that S has acquired a defeater for B would indicate that S's defeater system is *mal*functioning In such a case B would *cease* to be warranted for S (Sudduth 1999, p. 170)

contrary to initial appearances, there exists an internalist constraint within Plantinga's otherwise broadly externalist theory of warrant ¹⁷ And it is only by *neglecting* this internalist constraint that counterex amples such as BonJour's gain any plausibility ¹⁸

References

- Bergmann, Michael (1997) "Internalism, Externalism and the No-Defeater Condition" Synthese 110 399-417
- (2000) "Deontology and Defeat" Philosophy and Phenomenological Research 60(1) 87-102
- BonJour, Laurence (1996) "Plantinga on Knowledge and Proper Function" In Kvanvig 1996, pp 47–71
- Kvanvig, Jonathan L (ed) (1996) Warrant in Contemporary Epistemology Essays in Honor of Plantinga's Theory of Knowledge Lanham, Maryland Rowman and Littlefield Publishers
- Lehrer, Keith (1996) "Proper Functionalism Versus Systematic Coherence" In Kvanvig 1996, pp 25–45
- Plantinga, Alvin (1993) Warrant and Proper Function New York Oxford University Press
- Senor, Thomas (1996) "The Prima/Ultima Facie Justification Distinction in Epistemology" *Philosophy and Phenomenological Research* **56**(3) 551–66
- Sudduth, Michael (1999) "The Internalist Character and Evidentialist Implications of Plantingian Defeaters" International Journal for Philosophy of Religion 45(3) 167–87

Keywords

Plantinga, BonJour, warrant, defeasibility

Colin P Ruloff
Department of Philosophy
The Claremont Graduate School
Claremont, California
colin ruloff@cgu edu

Notes

¹ I shall assume here that the following construal of Plantinga's proper functionalism is sufficiently familiar to the reader to need no special explanation