Swampman: A Dilemma for Proper Functionalism

(~Penultimate Draft~, Forthcoming in *Synthese,* please *only* cite published version)

Jeffrey Tolly

University of Notre Dame

word count (including references): 10412

Proper functionalism claims that a belief has epistemic warrant only if it's formed according to the subject's truth-aimed cognitive design plan. The most popular putative counter-examples to proper functionalism all involve agents who form beliefs in seemingly warrant-enabling ways that don't appear to proceed according to any sort of design. The Swampman case is arguably the most famous scenario of this sort. However, some proper functionalists accept that subjects like Swampman have warrant, opting instead to adopt a non-standard account of design. But critics of proper functionalism hold that this strategy comes at a high cost: the design-plan condition now seems explanatorily superfluous. James Taylor construes cases like Swampman as posing a dilemma for the proper functionalist: either deny warrant in these cases and concede that proper functionalism doesn't capture our intuitions, or affirm warrant and undermine the explanatory power of the design-plan condition. Proper functionalists have replied to both horns of this dilemma. Recently, Kenny Boyce and Andrew Moon have argued that warrant-affirming intuitions on cases like Swampman are motivated by a principle that has a clear counter-example. Also, Alvin Plantinga presents a set of cases that supposedly cause problems for any analysis of warrant that lacks a design-plan condition. In this essay, I present a counter-argument to Boyce and Moon's argument, and show that a more robust reliability condition can accommodate Plantinga's problem cases. I conclude that we're left with no good reason to doubt that cases like Swampman raise a troubling dilemma for the proper functionalist.

Warrant, Knowledge, Proper Functionalism, Reliabilism, Generality-Problem, Swampman

1. Introduction

Proper functionalism is currently one of the more popular externalist theories of warrant. This is mostly due to Alvin Plantinga's innovative and substantive defense of this theory throughout the past three decades.[1] By "warrant," I mean that which grounds one's having knowledge that p so long as one believes p and p is true.[2] While proper functionalism includes a truth-conduciveness/reliability condition, its hallmark feature is an additional proper function condition that we can formulate as follows:

> DP A belief b has warrant only if b is formed in a way that proceeds according to the subject's truth-aimed cognitive design plan.[3]

According to Plantinga, a design plan for some entity x is a "set of specifications" for how x *ought* to work, i.e., a description of what it is for x to function properly given its purpose or aim (1993b:14). According to proper functionalists, creators (e.g., God) or processes (e.g., evolution) are typical sources of cognitive design plans.[4]

For the past two and a half decades, the most common strategy for attacking proper functionalism involves constructing hypothetical cases of putatively warranted belief formation that don't follow any sort of truth-aimed design plan provided by either a designer or designing process. I call such cases *DP problem cases*.[5] Swampman is arguably the most famous DP problem case.[6] The case of Swampman features a neurophysiological duplicate of a normal human being who is instantaneously formed from a random chemical/quantum event in a swamp. Being an exact neurophysiological duplicate of a normal human, Swampman goes on to form judgments "just like" normal humans do. Many have the intuition that Swampman can gain at least some kinds of knowledge—most plausibly, this includes simple perceptual knowledge. But, presumably, Swampman's cognitive faculties *weren't designed* in any way. Hence, Swampman is a paradigmatic DP problem case.

---

[1] The combination of Plantinga's (1993a) and (1993b) constitute the *locus classicus* of his proper functionalist theory of warrant.

[2] In other words, warrant is the entity that "takes" true belief to knowledge. This functional sense of "warrant" was popularized by Alvin Plantinga (1993b: v). Various philosophers have also presented proper functionalist accounts of epistemic *justification* (Bergmann (2006: 130-3), Graham (2012), (2014)). In this paper, I am only engaging an argument against proper functionalism about *warrant* in the above sense. I'll remain neutral regarding the implications that my arguments have for proper functionalist accounts of justification.

[3] See Plantinga (1993b: 19, 46-7) for the classic formulation of Alvin Plantinga's version of proper functionalism. Of course, there are other possible formulations of theories that would plausibly fit under the broad designation *proper functionalist theory of warrant*. Indeed, any account that includes DP or any sufficiently similar condition would fit under this designation.

[4] Plantinga notes that, "Human beings are constructed according to a certain design plan. This terminology does not commit us to supposing that human beings have been literally designed—by God, for example" (1993b:13). He grants that those conceiving of design plans as conferred by evolutionary forces are not incompetent users of the concept *design plan* (13). However, in a later chapter of that same book, he argues that design is best understood in terms of *agential design* (194-215).

[5] See Feldman (1993: 47) and Taylor (1991: 187-193) for early presentations of DP problem cases.

[6] Ernest Sosa has championed the Swampman case as highlighting critical problems for DP. See Sosa's (1993) for a full treatment. Sosa adapts the Swampman case from Donald Davidson's discussions of mental content (1987).

Dialectically, DP problem cases might just seem like simple counter examples to proper functionalism and DP, but the situation is actually more complicated. Some proper functionalists accept warrant verdicts in DP problem cases, explaining them as instances of a subject's acquiring a new or different warrant-enabling cognitive design plan—even though this design plan was *not* acquired through a designer or designing process like evolution.[7] Such proper functionalists thereby adopt what I call the *non-standard account* of design, which claims that design plans needn't emerge from some agent or process.

However, some philosophers, like James Taylor, have argued that taking this warrant-affirming approach to the DP problem cases—and thereby affirming the non-standard account of design—commits the proper functionalist to a rather unattractive result: DP is now rendered explanatorily dispensable in a theory of warrant.[8] So, according to Taylor, DP problem cases are properly seen as posing an interesting dilemma for the proper functionalist. First, if proper functionalists maintain that design plans must have designers (or designing processes) while *denying* that subjects have warrant in the DP problem cases, then they concede that their theory cannot accommodate the intuitive verdicts on these cases. On the other hand, if proper functionalists *affirm* warrant verdicts in DP problem cases and adopt the non-standard account of design, DP is rendered explanatorily superfluous in their theory.[9]

Proper functionalists have responded to both horns of Taylor's dilemma. First, Kenny Boyce and Andrew Moon have recently argued that our intuitive inclination to affirm warrant verdicts in DP problem cases is motivated by a tacit commitment to a substantive thesis—a thesis that they argue is false. In §2, I present Boyce and Moon's argument, and show how it falls prey to a reasonable counter-argument. I also show how the cases of my counter-argument undermine a straightforward approach to using Peter Graham's recent account of proper function in order to accommodate the intuitive verdicts on DP problem cases.[10]

I discuss the warrant-affirming horn of Taylor's dilemma in §3. There, I reconstruct Taylor's argument for the dispensability of DP. I then present Plantinga's objection to Taylor's dispensability argument. Plantinga argues that externalist theories of warrant *that lack DP* will struggle to accommodate intuitive verdicts on a particular set of cases.[11] I argue that Plantinga is mistaken, and that his concerns stem from a particularly narrow conception of how the reliability condition on warrant works. I show that a more robust reliability condition can elegantly account for Plantinga's problem cases.

I conclude that Boyce and Moon's argument and Plantinga's argument fail to undermine either horn of Taylor's DP problem case dilemma for proper functionalism. As a result, cases like Swampman remain a serious challenge to proper functionalism.

---

[7] See Bergmann (2006: 147-9). There, he argues that it's a live option for the proper functionalist to see Swampman as having a normal human design plan, thus avoiding the counter example to DP. Also, Plantinga (1991: 208-9) suggests that entities *might* be able to adopt design plans in ways that don't involve designers or evolutionary processes.

[8] Taylor (1991: 185) and Feldman (1993: 44).

[9] While not exactly the same, this formulation closely mirrors Taylor's presentation of this dilemma (1991: 185).

[10] See Graham (2014).

[11] See Plantinga (1991: 210-14), (1993c: 73-5) and (1995: 431).

2. Defending Swampman's Warrant

2.1 Boyce and Moon's argument

After considering other approaches proper functionalists might take in responding to DP problem cases, Boyce and Moon conclude that the best option is to find something wrong with the warrant-affirming intuition in these cases (2016: 2995). In turn, they attempt to challenge the underlying assumptions that motivate these warrant-affirming intuitions.

Boyce and Moon characterize our thought process on DP problem cases as follows:

> Our intuitive inclination is to believe that if we were to form our beliefs in the way that these creatures [e.g., Swampman] do, in similar environments, then *our* beliefs would be warranted. And we are also inclined to infer from this fact that *their* beliefs are warranted.
>
> Thus, it seems that the central intuition underlying our intuitive responses to Swampman-like scenarios may be stated (roughly) as follows:
>
> (CI) If a belief B is warranted for a subject S and another subject S* comes to hold B in the same way that S came to hold B in a relevantly similar environment to the one in which S came to hold B, then B is warranted for S*. (2016: 2990)

In other words, a tacit or explicit commitment to CI generates the intuition that Swampman has warrant. Hence, if one comes to reject CI, she'll no longer have a warrant-affirming intuitive response to DP problem cases like this.

Boyce and Moon set out to establish a counterexample to CI. They defend and present this counter example with a progression of cases. They start with a case that has a particular intuitive verdict, and then present subsequent cases that are relevantly similar to the first one. Given that we ought to reason similarly about relevantly similar cases, they argue that the reader should have an intuitive verdict on the last case that, in essence, constitutes seeing a counterexample to CI. The sort of belief-formation that occurs in these cases is what they call "unlearned doxastic responses" rather than learned cognitive processes (2991, 2995). Their cases involve judgments of object permanence, i.e., judgments on whether some object continues to exist even when it passes out of the agent's sensory range (2996). The progression of cases begins here:

> Case 1: Billy is a human infant who is an unknowing participant in an experimental program concerning early childhood cognitive development. In the process of being experimented upon, he sees a red ball go behind a screen, out of sight, and, via an unlearned doxastic response, forms the belief that the round object he just saw is behind the screen. (2997)

Boyce and Moon report that the clear intuitive upshot of this case is that Billy's belief has warrant (2997). Not so with the second case:

> Case 2: Billy sees a red ball go behind a screen. But due to a genetic birth defect, he has an abnormal doxastic response to that input and forms the belief that the round object he just saw has ceased to exist. (2997)

Boyce and Moon report that, "[I]t is intuitively clear that Billy's belief is not warrant-appropriate in this case since Billy does not have the right sort of doxastic response required for knowledge" (2997). As they note, Billy in Case 2 is a significant aberration from the typical way humans develop in their early months of life. Insofar as Billy's doxastic response deviates from this standard human cognitive practice, it fails to be warrant-appropriate.

Things get interesting in the next case:

> Case 3: All goes as it did in Case 2. However, owing to the design of the cognitive experiments to which Billy is being subjected, red objects that pass behind the screen placed in front of Billy are (as soon as they are behind the screen and out of Billy's sight) instantly annihilated by a powerful laser (which is also behind the screen and out of Billy's sight). We'll also add that Billy only tends to have the sort of abnormal doxastic response described in Case 2 when he sees *red* objects being occluded. In other cases when he sees objects of different colors occluded, he believes in their continued existence just as any other human child would. (2997)

Boyce and Moon think that Billy clearly lacks warrant in this case too. According to them, Billy lacks warrant even though his "belief-forming tendency is reliable in the experimental environment he is in" (2997). According to their reasoning, the same sorts of concerns that we had about Billy's aberrant nature in Case 2 should apply *mutatis mutandis* to how we judge his warrant in Case 3.

> Case 4: Zork, unlike Billy, is an alien child who lives on another planet, although members of his species are very human-like (so human-like, in fact, that even down to the molecular level, it is difficult to distinguish them from human beings). Furthermore, Zork's planet is very much like Earth, with one important exception. Owing to some odd features of its magnetic field and some currently unknown (to Earth scientists) quantum phenomenon, strange, macroscopic sized, red, spherical particles frequently pop into existence in the presence of observers on Zork's planet and then disappear. These particles have a further odd property. If one of them pops into existence in the presence of a group of observers, then if any one of those observers stops observing that particle, it immediately ceases to exist. Zork (who is also an unknowing subject of cognitive experiments on early childhood development being conducted by adult members of his species) sees one of these particles go behind a screen and via an unlearned doxastic response, forms the belief that the round object he just saw has ceased to exist. (2997-8)

Boyce and Moon think that Zork's beliefs ascribing non-existence to red balls are warranted, and their reasoning is as follows:

> It seems that this belief is warrant-appropriate for members of Zork's species. We think we should grant that if Billy's belief in Case 1 is warranted, then Zork's belief could also be warranted; any objection to thinking that one is warranted is also an objection to thinking that the other is. (2998)

The last two steps of the argument, as well as Boyce and Moon's intervening commentary, proceed as follows:

> Case 5: Billy has been abducted by aliens of Zork's species and taken to Zork's planet. Furthermore, Billy has exactly the same sorts of belief-forming tendencies that he was stipulated to have in Case 3. Billy sees one of the strange, red, spherical particles from Zork's world pass behind a screen and believes that it has ceased to exist.
>
> If we say of Case 3 that Billy's belief is not warrant-appropriate, then we should also say the same about Billy's belief in this case. So, given what we said in Case 3, we should also deny that Billy's belief has warrant in this case.
>
> Case 6: Billy and Zork are both present in a laboratory on Zork's planet. Both simultaneously observe one of the strange, red, spherical particles from Zork's planet pass behind a screen. All goes for Zork as it did in Case 4 and all goes for Billy as it did in Case 5. Consequently, both Zork and Billy form the belief that the round object that they just saw has ceased to exist.
>
> Given all of the above, we should accept that Zork's belief is warranted and Billy's is not. (2998)

Here at the end, Boyce and Moon think that Case 6 gives us a counterexample to CI, so long as we build in the stipulation that "Billy forms his beliefs 'in the same way' that Zork does" (2998). They say that this stipulation faces no significant challenges, as it's reasonable to read Billy's birth defect as altering his psychology and neural structure such that the psychological and physical-causal descriptions of both Billy and Zork's belief-forming processes will be the same.[12]

At the outset, one might wonder whether a (tacit or explicit) commitment to CI is actually motivating the warrant-affirming intuition on DP problem cases, and as a result, wonder

[12] Boyce and Moon write,

> Billy forms his beliefs "in the same way" …that Zork does…[W]e can add that there is no intrinsic difference between the internal physical processes relevant to the formation of Billy and Zork's beliefs, including those occurring in their brains and other parts of their nervous systems. (2998)

whether a counterexample to CI actually helps the proper functionalist in any way.[13] I'm inclined to agree with Boyce and Moon regarding the role CI plays in our intuitive responses to DP problem cases.[14] But critics of proper functionalism who are unsure of CI's psychological (and conceptual) role should still be concerned by Boyce and Moon's argument. At the end of their paper, they explain that Case 6 itself provides the material for an argument in favor of proper functionalism:

> Since Billy and Zork form beliefs in the same cognitive environment and in "the same way", it seems that the only facts left to explain the differences in warrant-appropriateness in their beliefs are facts about proper function and species membership. This indicates either that proper function is necessary for warranted belief or that the appropriate process of belief-formation required for warrant is relative to the species of the believer. (2016: 2999)

2.2 Case 3 underspecified

While Boyce and Moon's argument is a sophisticated and thought-provoking response to the warrant-denying horn of the DP problem case dilemma, it is not without its flaws. To start, I think that Case 3 has been importantly underspecified in a way that makes it unclear whether Billy's belief satisfies the reliability necessary condition on warrant.

Before proceeding, it is important to fill out Case 3 with the details that Boyce and Moon in fact added in, but not until after their presentation of Case 6. First, given the "in the same way" qualification, we must read Billy, in Case 3, as *basing* his judgment on the same kinds of doxastic or experiential mental states on which Zork bases his judgment. Secondly, in a footnote, Boyce and Moon claim that neither Zork nor Billy possess defeaters for their beliefs ascribing the non-existence of some red object.[15] Here, I take this to mean that neither Zork nor Billy possess reasons to doubt the truth or rationality of the target belief *nor* are they disposed to acquire such reasons. These qualifications are crucial, because the mental-state bases on which one forms a belief, and whether one has defeaters, crucially determine whether one's belief has warrant.

Next, Boyce and Moon state that "Billy's belief-forming tendency is reliable in the experimental environment he is in" (2997). But it's unclear what exactly the experimental environment is, and hence, unclear whether "reliability in that environment" is sufficient for satisfying the reliability necessary condition on warrant. To illustrate, we can fill out Case 3 in at least two different ways:

---

[13] Boyce and Moon's justification for this psychological thesis is anecdotal: they see it reflected in the comments of those who present the DP problem cases as reasons to reject DP (2016: 2990).

[14] I agree with this insofar as the "in the same way" and "in a relevantly similar environment" qualifications in CI jointly entail that S and S* have the same dispositional (counterfactual) profiles before and throughout their belief-forming processes. I would imagine that Boyce and Moon would be friendly to this reading of CI.

[15] See fn. 15 on pg. 2998.

Case 3a The scientists at Billy's experimental facility are committed to locking Billy in the facility for his entire life. Through elaborate locking mechanisms and deception, he will never get out. Their goal is to destroy every red ball that becomes obscured to Billy throughout his entire life.

Case 3b In one week's time, the scientists at Billy's facility will let Billy go free in the outside world, where he will proceed to make all sorts of incorrect *non-existence* ascriptions to every single red ball that he sees passing behind an obstruction.

If we fill in the case with 3b, it's much clearer that Billy's belief lacks warrant. This is because Billy's belief-forming process does *not* seem to be reliable in 3b. But I contend that when we fill in the case with 3a, it is significantly less obvious that Billy lacks warrant. I think the intuitive appeal of the *no-warrant* verdict on Case 3 has its pull because one might be inclined to tacitly build in background information like 3b.

But why think Billy's belief-forming process is unreliable in 3b? Of course, a fully satisfying answer to this question would require giving an informative account of the reliability necessary condition on warrant. As many have pointed out, any effort to precisely formulate this condition must answer the famed generality problem. In brief, the generality problem asks us to state the relevant ways of categorizing belief-forming process tokens so as to clarify how exactly the warrant-determining reliability measurements are taken from these tokens.[16] It's common to view answers to the generality problem as having two distinct parts: a statement of a token's relevant *type* description, and a statement of the *reference class* across which a token's relevant type is measured for reliability.[17] Presenting and defending a fully informative solution to the generality problem is certainly beyond the scope of this paper. Nevertheless, more modest insights regarding epistemic reliability can highlight plausible ways in which the process in 3b is unreliable.

One can formulate partial answers to the generality problem by describing the types and reference classes with respect to which a given token must be reliable in order to *confer warrant* on its resultant belief. That said, I think the following necessary condition accommodates our intuitions on a variety of cases:

Content-Basis Reliability (CBR)

> Necessarily, where S's token process t involves forming a belief that p on the (complete) basis of mental states m, t confers warrant only if the type T, [forming beliefs in propositions sufficiently similar to p on the basis of mental states sufficiently similar to m] has a high reliability measurement across a

[16] See Feldman (1985) and Conee and Feldman (1998) for classic statements of the generality problem for reliabilism.

[17] See Conee and Feldman (1998: 3) and Wallis (1994:251). Wallis uses the phrase "relevance class" to refer to the reference class. In the context of belief-forming process reliability, reliability measurements are ultimately ratios of true-belief to false-belief, taken across some class of possible belief-forming events. This class is called the reference class.

> reference class of all possible belief-forming events sufficiently close to t in which S undergoes T processes.

Granted, CBR is rather vague, but useful enough for our purposes.[18] According to CBR, reference classes are counterfactual in nature, comprised of event possibilities centered on the actual token process and arranged according to a similarity relation.[19] Importantly, *the subject from the token case* undergoes a belief-forming process (of some type) in each of these nearby possible events. Hence, whether a token satisfies CBR crucially depends on the subject's *dispositions* at or around the time of belief formation. Furthermore, CBR construes relevant types as [content-description, basis-description] pairs.[20]

Today, many philosophers view the sort of reliability required for warrant as a matter of an agent's possessing a certain sort of dispositional profile, where these dispositions are spelled out in terms of what happens in *nearby possibilities.*[21] Consider Carl Ginet's famous fake barn county case and how, intuitively, the subject lacks warrant.[22] A counterfactual account of reliability explains this case as follows: even though the subject forms a true visually-based

---

[18] Importantly, CBR just states *one* form of reliability that's necessary for warrant. A token might need to exemplify other forms of reliability in order to confer warrant.

[19] Here, the notion of modal nearness/distance is taken from Lewis and Stalnaker's semantics for counterfactuals. On their model, sentences of the form *If event X had occurred, then event Y would have occurred,* are true iff the nearest possible worlds in which X occurs are also worlds in which Y occurs. (Stalnaker (1968), Lewis (1973)) Of course, the notion of modal distance is figurative, corresponding to some measure of modal similarity or difference between possible situations/worlds.

[20] CBR is similar to Juan Comesaña's (2006) solution to the generality problem. Comesaña claims that,

> Given that there will always be some evidence that the belief is based on, the process that generates the belief will always instantiate a case of the type-schema *producing a belief that p based on evidence E*. (2006: 37)

Comesaña's solution, as stated, seems to type processes too narrowly. Given that "p" in his relevant type schema is filled in by the *specific* proposition believed in the token case, Comesaña's solution to the generality problem entails that any belief in any necessary truth will automatically have a maximal degree of reliability. This is because necessary truths are true in *every* possible situation. Hence, a process type's truth ratio, taken across any reference class of possibilities, will necessarily be 100% according to Comesaña's solution. But it's rather implausible that any correct judgment on a necessary truth (automatically) has maximal epistemic reliability for humans like us. CBR avoids this problem by typing processes more broadly. According to CBR, relevant type descriptions only hold fixed that the subject forms a belief regarding content that is "sufficiently similar" to the target propositions p from the token case.

[21] Interestingly enough, Pritchard's formulation of the safety condition on knowledge is quite similar to CBR:

> If a believer knows that p, then in nearly all, if not all, nearby possible worlds in which the believer forms the belief that p in the same way as she does in the actual world, that belief is true (2005: 163).

Given that reliability measurements are taken across reference classes of possible situations, we can interpret Pritchard's safety condition as a particular kind of reliability condition. Sosa (1999: 142) also defends a safety condition on knowledge. Also, see Goldman (1988: 63) for his presentation of a counterfactual based account of epistemic reliability.

[22] See Goldman (1976) for the classic presentation of Carl Ginet's original fake barn county case. See Prichard (2012: 251, 254) for a counterfactual explanation of the fake barn case.

barn-ascribing belief in the actual token case, there are many nearby possibilities in which the subject makes false visually-based barn-ascriptions.[23]

CBR elegantly captures our intuitions on 3a and 3b as follows: in 3b, there are a great many modally nearby judgments in which Billy falsely ascribes *ceased existence* to objects on the basis of seeing their occlusion. The contents and bases of these possible judgments are very similar to those of the actual token judgment in 3b. Moreover, not much would have to change from the token case in order for these possibilities to be brought about—just the passing of one week's time. According to CBR, these modally nearby false judgments populate the reference class and drive down the degree of reliability corresponding to Billy's token process in 3b. On the other hand, similar false belief-forming events in 3a are modally much further away from the token case. Due to this increased modal distance, they don't populate the token's reference class, and hence have no negative impact on the token's degree of reliability in 3a.

All this to say, on 3a it's much less plausible that Billy's belief lacks warrant. And if Billy *has* warrant in 3a, it seems as if there's no reason to deny that he has warrant in Cases 5 and 6, where false belief-forming events are similarly (modally) far away from the token case. But if we fill in the case with 3b and preserve the no-warrant verdict, the clear warrant-undermining factor (counterfactual *un*reliability) vanishes once we consider Cases 5 and 6. Either way, the chain of reasoning throughout Cases 1-5 won't as clearly deliver Boyce and Moon's intended intuition on Case 6. Of course, many—especially those sympathetic to proper functionalism—will have reservations regarding Billy's warrant in 3a. However, I contend that a warrant verdict in 3a is supported by a progressive-case argument of its own, which I present below.

2.3 Un-designed cognitive improvements and a counter-argument

Before discussions of the Swampman, James Taylor presented his own set of DP problem cases for the proper functionalist. Each of these involve un-designed cognitive improvements to already existing agents. For example, Taylor asks us to consider Clarence, who after taking a rough hit to the head acquires a very reliable clairvoyance capacity (1991: 190). On the basis of vivid clairvoyant-phenomenal imagery of events occurring in far off places, Clarence reliably

[23] There are good reasons to believe that modal distance, rather than other forms of distance, e.g., spatial distance, determine whether a given belief-forming possibility comprises a token's reference class. Consider two agents standing in a county filled with only real barns, forming beliefs ascribing the presence of barns. Suppose on the other side of the planet, there's a nearly indistinguishable county that contains only barn facades. Neither agent knows about this other county. In agent 1's case, she has no means of leaving the current county, let alone making it to the other side of the world. On the other hand, agent 2 has a tele-transporter that sends her around the world multiple times a day to make observations of barns. She has no idea that the fake-barn county is her next destination! Interestingly, the possibility of finding oneself in fake barn county and making false barn-ascriptions seems completely irrelevant to the warrant of agent 1's barn-ascribing beliefs in the normal county. But intuitively, this possibility *does* undermine the (or reduce the degree of) warrant of agent 2's beliefs in the normal county. It seems too likely that agent 2 forms an incorrect visual barn ascription, and we can infer that these belief-forming possibilities from the other side of the planet are members of the reference classes of agent 2's process tokens from the normal county—but this is not the case for the reference classes of agent 1's tokens! This odd phenomenon obtains even though, *spatially,* these possibilities are just as far away for either agent as they stand in normal county. But clearly, these possibilities are modally much closer to agent 2—very little has to change about the world in order to agent 2 to find herself making a barn-ascriptions on the other side of the planet. Not so with agent 1. Principles like CBR nicely capture this result.

forms true beliefs about these events. Furthermore, we can build-in that Clarence lacks defeaters for this sort of basing process as well.[24] He claims that such beliefs plausibly have warrant, even though they don't arise from belief-forming processes that follow a truth-aimed design plan. Neither evolution nor God nor anyone else designed Clarence's mind to form these clairvoyance-based beliefs.

Granted, clairvoyance is a bit mysterious and perhaps intuitions about it shouldn't be trusted. We can, however, think of less bizarre un-designed cognitive improvements. Consider the following:

> PRECISION
>
> > Liz is a normal human being. One day, she takes some hard hits to the eye sockets. These impacts alter the constitution of her neural cortex and color-sensing retinal cones, effectively allowing her to visually represent *double* the number of precise color shades that any other human can. On the basis of her sharp visual experiences, she reliably forms beliefs ascribing (demonstratively) these ultra-precise color shades.

Ceteris Paribus, it seems as if Liz's new and super fine-grained color beliefs can have warrant—especially once we build in that Liz lacks any disposition to acquire defeaters for these beliefs. Furthermore, consider this analogous case:

> UV
>
> > Cindy is a normal human being who takes a significant blow to the head and gains the ability to visually represent ultra-violet light, much like reindeer can. On the basis of these new visual experiences, she reliably forms beliefs that ascribe these UV properties.

Insofar as Liz's new capacity seems to be warrant-conferring, so does Cindy's. Nonetheless, neither of their new capacities were *designed* to be aimed at truth in any interesting way.

Plantinga's discussion of these un-designed cognitive enhancement cases is instructive. He begins by entertaining the idea that such cognitive alterations *couldn't possibly* occur without some sort of agential design—an issue to which I'll return in 3.2.[25] Plantinga goes on to grant the possibility of these un-designed changes for the sake of argument, but then suggests that a warrant-affirming verdict on these cases is consistent with DP. He compares cases like Clarence to the following: imagine having a radio that is dropped off the roof of a house and crashes to the ground. The impact, lo and behold, caused the radio to be able to pick up far-off radio stations with clarity that it never had before (Plantinga 1991: 208-9). Plantinga explains this sort of case as follows:

> 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 works when it is working properly. And can't we think of Taylor's [case]…along the same lines? Clarence's cognitive system now (by chance) works differently from the way it did before: this new way of

[24] Taylor also mentions that we can build in that Clarence has verified the outputs of his clairvoyant processes independently (1991: 192-3). This would serve to satisfy those who think there are internalist conditions on knowing, requiring that we have higher order awareness of the reliability of our belief-forming processes.
[25] See Plantinga (1991: 208).

> working gets adopted by us (and no doubt by Clarence and possibly by God) as a new design plan…and hence his precognition and clairvoyance can constitute knowledge…[T]wo 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. (209)

It seems as if Plantinga is floating the possibility of *un-designed design plan acquisition* in order to accommodate the warrant verdicts in cases of un-designed cognitive improvement. Such an explanation of these cases invokes the non-standard account of design, according to which having a design plan does not entail *being designed* by some agent or process.

I contend that cases similar to these three constitute a progressive-case argument for a warrant verdict on Case 3a. Importantly, my argument does not presuppose CI nor the un-designed design plan acquisition *explanation* for the warrant verdicts on these cases. I merely claim that the intuitive warrant verdicts on the following cases constitute a counter-argument to Boyce and Moon's original argument for the falsity of CI.

My counter-argument begins with the following case:

> SHADE
>
> > Tim is a member of a species that is very similar to humans with a very similar evolutionary history, but with one crucial difference. Creatures like Tim are by and large disposed to correctly (visually) represent colors, except for two very precise shades: Things that are *green-12* look to be *blue-47* to him, and things that are *blue-47* look to be *green-12*. Oddly enough, in the species' distant past, there was a slight evolutionary advantage conferred by this disposition to incorrectly represent these two shades. One day, Tim takes a hard blow to the head. The impact causes a slight reconfiguration to his cerebral cortex which now gives him the disposition to correctly visually represent these two shades.

Ceteris Paribus, it seems like Tim can now have warrant for his visual *green-12* judgments. SHADE seems importantly analogous to UV and PRECISION in how they are each cases of warrant-conferring un-designed cognitive improvements. Of course, SHADE is unique in how the cognitive improvement was a *correction* of a previously unreliable disposition. But this fact doesn't seem to disqualify it from being a warrant-enabling improvement. Now, consider a slightly altered version of SHADE.

> SHADE ENVIRONMENT
>
> > Hold fixed the details from SHADE except alter the explanation for the species' unreliability (with respect to the two shades) as follows: the reason creatures like Tim are naturally unreliable in this way is that, due to some quantum phenomenon, charged particles in the atmosphere alter the light reflecting off of *green-12* things to make them look like they're *blue-47* (and vice versa). This phenomenon only affects these two shades. But one day, a solar storm blows through the atmosphere and clears away these charged particles, and now *green-12* things actually look *green-12* to Tim.

Here, this is an un-designed change to the cognitive *environment* rather than the cognitive apparatus. Nonetheless, it's a change that seems to be warrant-conferring for Tim. Indeed,

in some ways this seems no different than how one's *barn* ascriptions go from being unwarranted to warranted as soon as one leaves fake barn county and finds herself deep in the heart of normal barn county.

Then, consider this alteration of Case 4:

> Case 7
>
> > Zork is born with a cognitive abnormality for his species that gives him a disposition to think that red spheres (just like any colored sphere) continue existing even when they become obscured. He forms these sorts of judgments "in the same way" any normal human would. Interestingly enough, unbeknownst to any other member of his species, a year before Zork was born, a solar storm blew through their atmosphere and changed the planet's magnetic field properties. Now, red spheres *don't disappear* when no one's looking—however, everyone on the planet (except for Zork) still mistakenly thinks that they go out of existence. Zork is rather introverted and keeps to himself, choosing to live far away from other members of his species. He never realizes (and lacks any disposition to realize) that the rest of his species radically disagrees with him about the persistence of obscured red spheres.

Once again, it seems as if Zork has warrant for his belief that red spheres persist while obscured. Case 7 involves *both* a cognitive apparatus alteration and an environmental alteration (from the way the environment was when evolution was shaping the cognitive faculties of Zork's species). In this sense, it's a combination of the factors involved in SHADE and SHADE ENVIRONMENT. But the fact that *both* kinds of changes occur doesn't seem to undermine the warrant-conferring nature of the combined alteration in Zork's case. Now, consider this final case.

> Case 8
>
> > Zork has the same cognitive abnormality as Case 7, but the solar storm never blew through. In addition, Zork was born into a large but closed-off facility on his planet. Unbeknownst to everyone, including Zork, the radioactive material in the walls of this facility counter-acts the planet's odd quantum-phenomenon, so that *inside* the facility, red spheres behave just like any other colored object: they remain in existence when obscured. Zork lives out his days in this facility. In fact, he has no idea that there's even an outside world, and he also has no way to get out.

Upon reflection, Zork seems to have warrant in Case 8 just like Case 7. Here, the main difference between these cases is that there are some *spatially* nearby possible belief-forming events in which Zork misjudges persistence. But these possibilities are *modally* far away, and as I suggest above, it's modal distance that determines warrant-enabling reliability.

I contend that Case 8 and Case 3a are analogous in the following sense: if Zork has warrant in Case 8, then so does Billy in Case 3a. Of course, the subject in 3a denies the persistence of some object while the subject in Case 8 ascribes persistence. But this difference doesn't seem like it could ground a difference in warrant. Ultimately, this new progressive-case argument shows a warrant verdict on 3a to be *at least* as plausible as a warrant-denying verdict. But this result cancels out the probative force of Boyce and Moon's argument, seeing as how a warrant-denying verdict on 3a is essential for yielding their counterexample to CI.

In addition, my counter-argument highlights that there are other DP problem cases that are more compelling than Swampman. Boyce and Moon are keen to point out that their argument doesn't trade on our intuitions regarding overly bizarre cases like persons popping into existence in a pond. Instead Cases 1-6 are "closer to home" and our everyday experiences (2999). But insofar as these cases are importantly closer to home for our intuitions, so are DP problem cases like UV, PRECISION, and SHADE. Given that it's at least somewhat contentious whether the Swampman's coming into existence is even metaphysically possible, I think that critics of proper functionalism should follow the lead of James Taylor and return to using more down to earth DP problem cases.

2.4 Graham's approach to function

As I noted in the introduction, one approach to preserving proper functionalism, while affirming warrant verdicts in DP problem cases, is to adopt the non-standard account of design. According to the *non*-standard account of design, design plans can emerge without any sort of designing activity from a designing agent or process. As I'll argue in the next section, this move comes at a high cost. But before engaging this latter option, I'll briefly explore a potential path for splitting the horns of Taylor's dilemma. An anonymous referee insightfully points out that Peter Graham's recent work on function in epistemology could provide the conceptual resources for both affirming warrant verdicts in DP problem cases while maintaining that design plans must emerge from some design-conferring process.

Graham has recently defended what he calls a functionalist account of warrant.[26] Importantly, Graham uses the term "warrant" the way most epistemologists use "justification" (2014: 16). So, strictly speaking, Graham's views don't directly engage our current debate regarding the nature of epistemic warrant (i.e., the state that takes true belief to knowledge). However, I will present and evaluate what I take to be the most straightforward method of applying Graham's ideas to the debate at the center of this essay.

Graham writes that "epistemic [justification] consists in the normal functioning of the belief-forming process when the process has forming true beliefs reliably as a function" (2014: 16).[27] Graham endorses an etiological theory of functions according to which "functions turn on histories that explain why the item exists or operates the way it does" (2014: 16). While he cites evolution as one possible source of an organism's functions, he thinks that an organism's normal metabolism can ground the emergence of functions as well (18). Graham sees metabolism as "a feedback mechanism" that "explains the continued existence of traits within individuals over cycles in virtue of their beneficial effects" (29).

The case of Swampman brings the importance of metabolic feedback mechanisms into sharp relief since evolutionary etiology cannot be a source of function for his organs or mental processes. Graham claims that "Swampman at the moment of his creation has no functions; over time Swampman's organs and traits acquire functions" (31). Only once an organ "starts to benefit the organism and thereby contribute to its own self-replication and repair" does it acquire a function with effects that "play an explanatory role" in the life of the organism (30). The same could also be said for Swampman's belief-forming processes and their justification-

[26] See Graham (2012), (2014).

[27] Throughout, Graham uses the phrase "normal functioning" rather than proper functioning, but it's plausible that he doesn't intend to communicate anything substantially different than proper functioning.

conferring powers. For example, Swampman's perceptual capacities can acquire the function of *reliably producing true beliefs* once the capacity's track record of "reliably [producing] true beliefs benefits the creature" and "the creature possesses such a capacity *because* reliably [producing] true beliefs benefits the creature" (28, emphasis mine).

Now, let's apply this treatment of the Swampman case to the property of epistemic *warrant*. Suppose that metabolic etiology can be a source of a new, truth-aimed, warrant-enabling perceptual capacity. This means that, after enough metabolic cycles in which Swampman's true perceptual beliefs bring about beneficial results for Swampman, his perceptual capacity can begin conferring warrant on the resultant beliefs. Seeing as how a metabolic process plays an *essential* role in the emergence of Swampman's new warrant-enabling capacity and design plan, Graham's approach presents us with a strategy for affirming warrant verdicts in DP problem cases without committing to the non-standard account of design.

However, there are serious problems facing this straightforward application of Graham's views to the current debate about warrant.[28] First, it appears that agents like Liz and Cindy in PRECISION and UV can have warrant for their beliefs *immediately after* their transformations occur. There doesn't seem to be any sort of required lag time before their new belief-forming processes confer warrant.[29] In addition, there's further reason to think that Graham's metabolic conditions are *un*necessary for the emergence of new *warrant*-enabling cognitive capacities. Consider a slight variation of PRECISION, in which Liz's new belief-forming capacity confers no practical benefit to her and even makes her disposed to experience intense pain and digestive organ failure for long periods of time after she uses the cognitive capacity. From her perspective, she has no idea why these painful episodes arise. Nevertheless, it seems just as plausible that Liz's new fine-grained color beliefs count as knowledge, even though using her new capacity to form these beliefs is detrimental to her overall flourishing as a biological organism.[30]

Here, I've only raised problems for this straightforward application of Graham's theory to Taylor's DP problem case dilemma. There could very well be other promising avenues for marshalling Graham's ideas to the aid of proper functionalism about warrant. Unfortunately, a full treatment of Graham's approach to function is beyond the scope of this paper. I now turn to the warrant-affirming horn of Taylor's DP problem case dilemma. There, I'll defend Taylor's claim that adopting the non-standard account of design renders DP explanatorily superfluous.

---

[28] Of course, all that follows from my criticisms is that justification, in Graham's sense, is unnecessary for warrant.

[29] Insofar as one *is* inclined to think that some post-transformation time-lag is required before the new capacity can yield knowledge, there is a purely reliabilist way of accommodating this idea. While it's certainly contentious, some reliabilists could require, in addition to CBR, a belief-forming process type to have a sufficiently good *track record* in the subject's actual experience in order to confer warrant. See Goldman (1986: 45) for a discussion of this suggestion. The requirement of a strong track record would entail that Liz, upon *first* using her new capacity, couldn't gain warrant for the resultant beliefs. All this to say, this is an approach to explaining a time-lag requirement in a way that doesn't invoke the notions of design, design plan acquisition, or function.

[30] One might argue that forming a preponderance of true beliefs is *in itself* a substantial contribution to the overall flourishing of an organism. But if this is the case, the acquisition of a reliable belief-forming mechanism entails the acquisition of a capacity that thereby confers a "benefit" on the creature. This latter condition on warrant would simply collapse into the reliability condition—a pattern that I'll discuss in more detail in §3.

3. Design is dispensable

3.1 Reduction

Admittedly, the non-standard account of design and proper function is rather perplexing. Ernest Sosa notes 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'" (1993: 57). Many will be inclined to agree with Sosa, and dismiss out of hand any notion of proper functionalism that relies on this non-standard account of design. But some will be less troubled, and perhaps see the idea of un-designed design plan acquisition as a plausible means of salvaging proper functionalism in the face of DP problem cases.[31]

However, James Taylor argues that this latter strategy comes at a rather steep cost:

> If the concept of proper functioning is not analyzed in terms of an actual designing process or agent, then Plantinga's account either offers no clear advantage over reliabilism, or it is not clearly distinguishable from reliabilism. (1991: 185)

As Taylor further explains the Clarence case, he clarifies that, by "not clearly distinguishable from reliabilism," he means "*reduces to* reliabilism."[32]

> Not every newly-acquired un-designed cognitive faculty would count as part of the design plan. Why should Clarence's clairvoyant ability qualify? Presumably, it qualifies because the design plan, on the liberal [non-standard] construal under consideration, sanctions any cognitive faculties, including those un-designed and acquired, which are reliable and produce beliefs on the basis of the appropriate sorts of grounds…If this is the case, the appeal to proper functioning and design is *superfluous*: A belief is warranted only if it is produced by a properly functioning cognitive faculty…A cognitive faculty functions as it was designed to function if and only if it is reliable and produces beliefs on the basis of appropriate sorts of grounds…The appeal to proper functioning and design simply *drops out of the picture*. (1991: 191-2 emphasis mine)[33]

From this passage it's clear that, by "reliabilism," Taylor just means a theory of warrant that *does not* invoke the concepts of design or proper function, in which a reliability necessary condition features prominently. For short, I'll call any condition on warrant that does not

---

[31] For example, see Bergmann (2006: 147-9).

[32] Much thanks to an anonymous referee for pushing me to focus on Taylor's reduction thesis and present it as the central move in Taylor's defense of the warrant-affirming horn of his dilemma.

[33] William Alston, Richard Feldman, and Ernest Sosa all make a similar claim. Alston surveys an assortment of Plantinga's putative counterexamples to reliabilism, where Plantinga takes these cases to suggest a proper function condition on warrant. Alston concludes that the best explanation for the no-warrant verdicts on these cases more has to do with the reliability condition being unsatisfied (1995: 397-400). While these cases exemplified some sense of reliability, it wasn't the epistemically *relevant* kind of reliability that can confer warrant (401). Feldman concurs. After considering DP problem cases and Plantinga's approach to handling them, Feldman suspects that "it is the reliability requirement as Plantinga construes it that gives his theory whatever merit it has. The proper function requirement is not really doing any work in the theory" (1993: 44). In addition, Sosa contends that a similar account with just as much explanatory power as proper functionalism can be formulated (without mentioning design) by simply invoking reliable *faculties*. Sosa says "[this account] reduces to a form of 'reliabilism' in the broad sense" (1993:58).

invoke design or proper function a *non-design* condition. In addition to reliability, Taylor suggests a further non-design condition, stating that only certain sorts of experiences or beliefs can serve as warrant-conferring *bases* for certain sorts of resultant beliefs (191, 199).

For simplicity, I take Taylor to be defending the following thesis:

> Reduction (R)
>
> > Any correct theory of warrant that includes DP, as construed according to the non-standard account of design, reduces to a correct theory of warrant (that includes a reliability condition) that doesn't include DP or any other condition that invokes design.[34]

If R is true, then those who accept a non-standard account of design can simply remove DP from their theory of warrant without any important explanatory loss. In what follows, I'll explain and expand Taylor's argument for R. After that, I'll present Plantinga's criticism of Taylor's argument. Lastly, I'll demonstrate a critical flaw with Plantinga's criticism.

3.2 Taylor's abductive argument for R

Unfortunately, Taylor does not rigorously present or explain an argument for R. However, certain key passages provide clues for how such an argument might go. First, consider how Taylor discusses the way in which one seems to recognize instances of proper function—like in the cases of UV or Plantinga's radio—once we're in the mindset that accepts the non-standard account of design:

> Design plans can be considered in isolation from external designing agents or processes. They can be considered as something internal to the cognizer - as instantiations of a set of specifications for ideal cognitive operation. It seems that we can often determine the design plan of a thing just by observing it in operation. For instance, one can find out a lot about the design plan of an automobile simply by *observing its operation under various conditions*. (1991: 195, emphasis mine)

In other words, we're inclined to think that agents like Cindy have cognitive design plans because of the way their minds operate across various counterfactual situations. But what sorts of counterfactual "operations" are particularly salient for determining whether an agent has a truth-aimed cognitive design plan? At this point, Taylor would most likely highlight this pattern shared by each DP problem case: these agents all gain dispositional profiles for

---

[34] Initially, Plantinga suggested interpreting Taylor's reduction thesis in terms of theoretical equivalence:

> 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. (Plantinga 1991: 213)

Plantinga then points out that, strictly speaking, Taylor framed his initial argument (which I won't reconstruct here) in terms of material equivalence, but what he needed to establish was a relation of broadly logical (i.e., metaphysical) equivalence (213). In their next exchange, Taylor confirmed that he was suggesting a relation of metaphysical equivalence between a correct (reliabilist) theory of warrant that doesn't invoke design and one that does invoke design (Taylor 1995: 424-5).

forming beliefs in certain *ways* across nearby possible worlds, and for a sufficiently high percentage of these beliefs to be true across nearby possible worlds. Here, Taylor's reasoning most likely embodies an abductive inference. First, it seems that a variety of informative non-design conditions, including reliability, are well suited to describing these dispositional profiles that enable warrant. This being the case, it appears as if a simple and explanatorily powerful set of necessary and sufficient conditions for warrant can be constructed with these non-design conditions.[35]

Of course, this sort of inference to the best explanation is standard in epistemological theorizing. Interestingly enough, inference to the best explanation plays a central role in Plantinga's own defense of proper functionalism. After presenting a number of cases that he takes to be counterexamples to the sufficiency of various extant internalist and externalist theories of warrant, Plantinga concludes, "[A]ll [of these cases] are such that their beliefs lack warrant for them. In each case the reason, I suggest, is *cognitive malfunction,* failure of the relevant cognitive faculties to function properly, to function as they ought to" (1993b: 4). However, someone like Taylor might introduce his counter-abduction as follows: now that we see that *having a truth-aimed cognitive design plan* doesn't require an actual designing process or agent—but simply amounts to having the right sort of dispositional profile—we can account for warrant more elegantly with a set of conditions that doesn't invoke design. If this counter-abduction is cogent, we should expect to find warrant-undermining features in Plantinga's famous counterexamples—features that can be described in dispositional terms that don't invoke design. I contend that this is what we *do* in fact find. While it's beyond the scope of this paper to engage all of Plantinga's cases, I'll demonstrate this pattern on Plantinga's most famous case: the serendipitous lesion.[36]

In this case, the subject, who Taylor later calls Jane, suffers from an odd brain lesion.

> This lesion wreaks havoc with [Jane's] noetic structure, causing [her] to believe a variety of propositions, most of which are wildly false. It also causes [her] to believe, however, that [she] is suffering from a brain lesion. [Jane] has no evidence at all that [she] is abnormal in this way, thinking of [her] unusual beliefs as resulting from an engagingly original turn in mind…Surely [Jane] does not know that [she] is suffering from a brain lesion. [She] has no evidence of any kind—sensory, memory, introspective, whatever—that [she] has such a lesion. (Plantinga 1993a: 195)

Plantinga thinks that Jane's beliefs clearly lack warrant. However, this case has been underspecified in important respects. As we'll see, on the specifications of the case in which Jane clearly lacks warrant, we can readily account for this lack by invoking other non-design conditions on warrant. First, as Plantinga formulates the case, it doesn't look like Jane's belief that L, *I have a lesion,* has any mental state basis (either experience or belief). The belief just seems to pop up along with the lesion. But as Taylor suggests, beliefs must be based on the

---

[35] Of course, we currently lack a robust and fully detailed formulation of these conditions. But this doesn't mean that we lack any sort of grasp of how, generally, these formulations might go or what sorts of things these conditions could explain. Philosophers like Taylor would most likely contend that we only need this sort of grasp of these non-design conditions in order to competently judge whether a further design condition is explanatorily superfluous. Furthermore, now that design is no longer understood in terms of a designing process or the intentions or a designing agent, it does not seem as if the envisaged proper functionalist has any firmer grasp of the notions *design plan* and *proper function.*

[36] For Plantinga's other classic cases that appear to feature cognitive malfunction of some sort, see (1993a: 190-208).

right sorts of mental states—either doxastic or experiential—in order to have warrant (1991: 191, 199). We can call this the *right basis* condition on warrant.[37] There does, however, seem to be a way of modifying the case such that it would satisfy the right-basing condition. Given that Jane's belief is about some inner state of hers, we could liken this case to one of typical proprioception and build in that Jane bases her judgment on the right sort of *introspective experience*.

However, even with this modification, we could read Jane's case as failing other non-design conditions. Upon reflection, it's clear that agents must not only *in fact* base their judgments in the correct kinds of mental-state bases, but they must also be disposed—in a stable fashion—to do so. Call this the *dispositional right basis* condition on warrant. For example, suppose at the time Jane forms her introspective belief that L, the lesion also has her disposed—with respect to nearby possibilities—to believe all sorts of mathematical propositions (e.g., 2+2=4, Gödel's first incompleteness theorem, etc.) on the basis of introspective experiences alone. Reasonably, her L belief would lack warrant in this case, given the preponderance of nearby possibilities in which she uses introspective bases for judgments that are never warranted on the basis of introspection alone.

But even if her L belief satisfies the *right basis* and *dispositional right basis* conditions, it might fail to have warrant due to unreliability. Suppose Jane is disposed as follows: In nearby possibilities, she comes to believe propositions on the basis of introspective experiences that are mostly false (e.g., she'll introspectively/proprioceptively believe that her skull is below her toes, when it really isn't).

If we augment the case further and remove these problematic dispositions from her, it's much less clear that Jane fails to have warrant. First of all, there doesn't seem to be anything impossible about being able to warrantedly introspect all sorts of things that typical humans can't, including whether one has a brain lesion. But here we can just apply the general pattern that we observe with the DP problem cases: if a creature acquires the right sorts of cognitive dispositions, one's beliefs carried out according to those dispositions can have warrant. Just like Cindy acquired the cognitive dispositions of other creatures who can warrantedly represent UV properties, and thereby acquired the capacity to warrantedly represent these properties herself, why not think that Jane has undergone a similar experience with respect to introspective lesion representation? Surely, the mere fact that the new cognitive dispositions come about through the removal of brain tissue doesn't seem to undermine Jane's warrant. Imagine a modified case of UV, such that part of Cindy's neuro-physical alteration involved the destruction of one small region of her cerebral cortex. All else being equal, this particular description of Cindy's alteration doesn't undermine the warrant she has for her new beliefs that represent UV light.[38] All this to say, if we modify Jane's case accordingly, it seems like Jane actually has warrant. Moreover, it's unclear how Plantinga could modify the lesion case

---

[37] In an illustrative case, Chris Tucker asks us to consider a subject S who is reliably disposed to form a belief that p, *Gödel's first incompleteness theorem is true*, solely on the basis of the belief that q, *People hate Keystone Light beer* (2014: 3349). Plausibly, no matter what S's other dispositions are, basing p on q can never yield a warranted belief that p. This holds even though q reliably indicates p in the trivial sense that all nearby worlds in which q is true, p is true as well—because p is true in *all* possible worlds. We can multiply cases like this. Consider someone with a reliable disposition to form the belief *1+1=2* solely on the basis of a visual experience representing the color *orange*. This is not the sort of basis that could confer warrant on that sort of belief. Tucker suggests that these cases illustrate that the right-basing condition cannot be analyzed in terms of reliability (3327).

[38] This sort of suggestion is made by Taylor (1991: 197) and Alston (1995: 399-400).

such that it would clearly highlight the joint-insufficiency of all the non-design conditions on warrant, thus supporting an additional DP condition.

As I mentioned, Plantinga entertains the idea that the cognitive alterations in DP problem cases *cannot possibly* occur without some sort of agential design. Plantinga's worries draw upon broader theistic commitments:

> According to theism, God not only creates the world, but constantly upholds and sustains it…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?…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. (1991: 208)

In this passage, Plantinga seems to be considering the following thesis:

> Exhaustive Creation and Providence (ECP):
>
> Necessarily, for every contingent state of affairs A that obtains, God intends for A to obtain and brings about A.[39]

On ECP, the apparent "freak accidents" of Liz, Cindy, and Tim were actually just the extraordinary means by which God brought about the new cognitive capacities that God intended them to have. Hence, these agents possess (new) cognitive design plans in the *standard* sense of design. One might think that this strategy clearly avoids Taylor's abductive argument for R, as there's no longer any reason to think that having a design plan *reduces* to having a certain dispositional profile.

However, this strategy undermines the explanatory power of DP in a slightly different fashion. On ECP, the set of cognitive dispositions that some creature in fact possesses is trivially the set of cognitive dispositions that it was designed to have. But this means that cognizing *according to a design plan* cannot differentiate between cases of warrant-conferring and warrant-undermining belief formation. Here, we now want our theory of warrant to tell us what a *warrant-enabling* design plan amounts to. We can once again consult the DP problem cases and recognize the same pattern highlighted by Taylor: satisfying the non-design dispositional conditions looks to be *sufficient* for one's having a warrant-enabling design plan. Hence, DP is still dispensable even if we assume ECP.

3.3 Plantinga's response to Taylor

Ultimately, I find Taylor's abductive argument to be a strong reason to accept R and view DP as superfluous in a theory of warrant once one accepts the non-standard account of design. Plantinga, however, is not convinced. He thinks that reliability is not up to the task of capturing all of the externalist features constitutive of warrant:

---

[39] ECP is a rather controversial thesis amongst theists, which explains why Plantinga himself does not commit to it, but rather considers it for the sake of argument. While he states that it isn't clear to him that a randomized un-designed state of affairs *is* possible on a theistic picture of the world, he is quick to note, "But it also isn't clear that it *isn't* [possible]" (1991:208).

> 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? …But …even if the set of cognitive powers is reliable *überhaupt*, it doesn't follow that every belief produced by them has warrant…Shall we say that the belief has warrant if and only if the cognitive process or module is functioning reliably in this very instance …But this entails once more that beliefs produced by the serendipitous [lesion] 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 (1991: 214).[40]

In essence, Plantinga suggests that reasonable externalist accounts of warrant simply cannot do without notions like design, proper function, and malfunction. Plantinga seems to think that *no matter how narrowly or broadly* the reliability condition types belief-forming process tokens, a high reliability measurement, combined with satisfying the other non-design conditions, is still insufficient for warrant. This result highlights how a set of non-design conditions could never—in any elegant fashion—constitute sufficient conditions on warrant.

3.4 Robust reliability and a response to Plantinga

By way of response, I think Plantinga's pessimism about the reliability condition's explanatory power is rooted in an unfounded (but commonly held) presupposition pertaining to the workings of the reliability condition: namely, that whether a belief satisfies the reliability condition depends on just one type and one corresponding reliability measurement.[41] But why couldn't the reliability condition depend on multiple types and multiple reference classes? At first glance, there doesn't seem to be any independent reason to think that the reliability condition only invokes one type and one reliability measurement. Moreover, I take the sorts of cases Plantinga cites as considerations *in favor of* a multi-type/multi-reference class reliability

---

[40] Plantinga (1995: 431), (1993c: 74) makes almost the exact same reply to both Alston's (1995: 398-401) and Feldman's (1993: 44) suggestion that a reliability condition could handle all of the cases Plantinga uses to motivate a proper function condition.

[41] Interestingly enough, as Conee and Feldman present and explain what the generality problem *is,* they frame it as the task of finding "a specification of *the* relevant type" (1998: 3 emphasis mine). In other parts of the essay, they criticize various theories of type relevance for not identifying *one* type as relevant for epistemic evaluation (e.g., pg. 10). This way of setting-up the generality problem presupposes that epistemic reliability is simply a matter of having a high reliability measurement with respect to *one* type description. Conee and Feldman don't present any argument for this presupposition. Given that the problem was framed in this way, it's not surprising that almost every subsequent theory of type relevance aimed at solving the generality problem presupposes that just *one type* determines epistemic reliability. However, there are some exceptions. Most notably, Wunderlich (2003) defends a multi-type theory of reliability called *vector-reliability*. Also, Goldman (1986) suggests that warrant requires both "local" and "global" reliability. According to Goldman, the difference between these two kinds of reliability is "*the range of uses* for which the process is reliable" (44-5 emphasis mine). Here, one might interpret Goldman as proposing a relevance theory according to which multiple reference classes determine whether a token confers warrant.

condition on warrant. I'll also show that this sort of reliability condition can be articulated in an intuitively elegant manner.

To begin, the lesion case highlights how knowledge requires a process token to be reliable with respect to *multiple process types,* each with a different degree of descriptive broadness/narrowness. Interestingly enough, the relevant type schema from CBR provides an easy way to capture this sort of structure. [Content description, basis description] pairs can have different degrees of broadness/narrowness. In the lesion case, we see that the token t_l's counterfactual reliability with respect to the type *S's forming introspectively-based judgments on whether S possesses a brain lesion* is insufficient for conferring warrant. While reliability with respect to this narrower type is plausibly necessary for conferring warrant, it seems that t_l must also have reliability with respect to the broader type, *S's forming introspectively-based judgments regarding states of S's body*. We can multiply cases like this. Imagine a subject who is reliably disposed such that, whenever he considers the formula 4+27, he'll intuit that the answer is 31. However, due to some cognitive glitch, he's disposed to form incorrect judgments 50% of the time on almost every other simple addition formula when he considers them. Here, we see that counterfactual reliability with respect to the narrower type *S's forming intuitive judgments on the sum of 4+27* is insufficient for conferring warrant. It seems that reliability with respect to some broader type is required too. Cases like this suggest that, generally, in order for a token to confer warrant, it must be counterfactually reliable with respect to *multiple* types, where these types have varying degrees of broadness/narrowness.

Cases like the drunk vision case suggest that, in order for a token to confer warrant, its relevant types must each be reliable with respect to *multiple* counterfactual reference classes—each with varying degrees of broadness/narrowness. Let t_v be a token visual belief-forming process for S during an occasional wild night of heavy drinking. Plantinga correctly notes that the type T_v, *S's forming visually based beliefs pertaining to S's environment* is, *ceteris paribus*, reliable with respect to a broad class B of nearby possibilities. Plausibly, in order for t_v to confer warrant, T_v must be reliable with respect to class B. But Plantinga is correct in noting that this sort of reliability is insufficient for conferring warrant. This is because, throughout the particular evening of heavy drinking, S's vision is wildly misleading. A multi-reference class reliability condition can straightforwardly account for this case. As I said, all of the possibilities within B are modally close to t_v—let's say they all lie within some modal distance X from t_v. But within B, we can focus on the sub-class C of possibilities that only involve S's drunken visual judgments *throughout the night of heavy drinking*. Importantly, all of the possibilities contained in C are *modally closer* to t_v. In other words, each of the C possibilities lie within a modal distance Y from t_v, where $Y < X$. We can imagine B and C as spheres of possibility space with t_v at the center, where C is a concentric sphere of B. It's in this sense that class C is a descriptively *narrower* class than B, because *all* of the possibilities contained in C are much more similar to t_v than some of the possibilities contained in B. Clearly, the type T_v is unreliable with respect to class C, given that S's vision went haywire the night that t_v occurred. I contend that this fact explains why t_v fails to confer warrant in this case.

Plantinga describes the drunken vision case as one of reliable belief formation that *malfunctions*. But here we see that we can account for this no-warrant verdict just by invoking reliability: token t_v's relevant type T_v is reliable with respect to one of its (broader) warrant-determining reference classes, but unreliable with respect to another one of its (narrower) warrant-determining reference classes.[42] Of course, we can multiply cases like this, and they'll

[42] Duncan Pritchard also seems to think that a correct theory of warrant must take into account the varying degrees of modal distance (from the token) at which possible error can occur (2012: 179-80).

all suggest that in order for a token to confer warrant, its relevant types must have reliability with respect to *multiple* reference classes—each with varying degrees of counterfactual narrowness/broadness.[43]

Seeing as how the broadness/narrowness of a reference class is determined by *modal* distance, a more detailed reliability condition could analyze the various reference classes determinative of warrant in terms of a particular counterfactual ordering relation O_1—which is ultimately a relation of similarity that comes in many degrees. Importantly, another relation like this could also specify the varying degrees of broadness/narrowness characterizing the [content, basis] pair types that are determinative of warrant. Descriptively narrower [content, basis] pairs have a greater degree of similarity to the precise description of the content of S's belief, and the basis of S's belief, in the token case. Just like the reference classes could be analyzed in terms of similarity relation O_1, these various relevant types could be analyzed in terms of some similarity relation O_2. Given our general competence in deploying various counterfactual notions in everyday life, we have no reason to believe that we lack any intuitive grasp of relations like O_1 and O_2. Hence, there's no good reason to doubt that an informative and elegant multi-type/multi-reference class theory of warrant could be formulated in terms of these two similarity relations.

While capturing these varying degrees of broadness/narrowness in an informative reliability condition will still be challenging, it's important to point out that Plantinga's own formulation of proper functionalism faces a similar theoretical challenge once one accepts the non-standard account of design. Here's a contemporary rendering of Plantinga's account:

> S's belief B has warrant iff
>
> 1. B has been produced by a segment/module M of the agent's cognitive design plan that is aimed at truth and functioning properly in an environment appropriate for the (the functioning of) that module.
> 2. There is a high statistical probability that beliefs produced by M in the maxienvironment, in which B is formed, will be true.
> 3. There is a high statistical probability that beliefs produced by M in the minienvironment, in which B is formed, will be true.[44]

The maxi and minienvironment distinction was made to capture intuitively correct warrant verdicts on Gettier-style cases similar to the fake barn county case.[45] A maxienvironment is a broad description of the environment in which B is formed, where this description characterizes the environment in which S's faculties were "designed to function" (313). B's minienvironment, on the other hand is a "much less global" and "much more specific and detailed state of affairs" describing the context of B's formation (313-14). So, we see here that

---

[43] A similar dialectic ensues between Plantinga and Ernest Sosa over subsequent discussions pertaining to the Swampman case. Sosa (1996: 275-8) presents a truth-tracking reductive account of proper function in terms of counterfactual conditions like the safety condition. Plantinga (1996: 368-70) responds by noting that these truth-tracking conditions are nowhere near sufficient for warrant, for similar reasons that narrow-type reliability theories cannot deliver the correct warrant verdict in the lesion case. I do, however, think that Sosa's broader epistemological framework has the resources to furnish a reply very similar to the one I make here. One can view Sosa's truth tracking conditions as securing the narrowest sort of epistemically relevant reliability that's necessary for warrant. Then, Sosa could invoke his account of intellectual virtue, which is analyzed in terms of [content, circumstance] pairs to capture the broader sort of reliability that's necessary for warrant (1991: 138).

[44] Combined from Plantinga (1993b: 46), Plantinga (1996: 313-14), and Plantinga (2000: 160).

[45] See Plantinga (1996: 312).

Plantinga's account invokes truth/falsity ratios taken across classes of possibilities where these classes have varying degrees of broadness/narrowness.

But these characterizations of the maxi and minienvironments are rather vague. One might have initially thought that the creative intentions of the designer (e.g., God) straightforwardly determine the specific descriptions of the environments for which the subject's faculties were "made to function." But once one adopts the non-standard account of design, one cannot appeal to this strategy for fixing maxi and minienvironment descriptions. As things stand, Plantinga's proper functionalism faces a no less challenging task of specifying the varying degrees of broadness and narrowness across which warrant determining truth/falsity ratios are taken.

Importantly, we've seen that capturing these different degrees of broadness/narrowness within a reliabilist condition on warrant can account for Plantinga's problem cases without invoking notions like design or proper function. Hence, we're left with no reason to think that these sorts of cases weaken Taylor's abductive argument for R and the dispensability of DP.

4. Conclusion

Above, we examined contemporary objections to both horns of Taylor's DP problem case dilemma for proper functionalism, and found each of these objections to have serious problems. First, we saw Boyce and Moon's progressive-case argument against the warrant-denying horn to be importantly underspecified in key respects. Moreover, the progressive-case argument is itself subject to its own progressive-case counter-argument. On the warrant-affirming horn, Plantinga argued that any informative account of warrant must invoke DP in some way, making DP explanatorily indispensable. However, we saw that Plantinga's argument presupposed a single-type/single-reference class approach to the reliability condition. As it turns out, there's good reason to believe that warrant requires high reliability measurements across multiple types and multiple reference classes with varying degrees of broadness and narrowness, and that a multi-type/multi-reference class reliability condition could elegantly accommodate intuitions on Plantinga's cases. As things stand, both horns of Taylor's dilemma seem just as troubling for the proper functionalist as ever. Hence, we should still hold that cases like Swampman present a serious problem for proper functionalism.[46]

References

Alston, William. (1995), "Epistemic Warrant as Proper Function." *Philosophy and Phenomenological Research.* 55.2: 397-402.

Beddor, Bob. (2014), "Process Reliabilism's Troubles with Defeat." *Philosophical Quarterly*. 65.259: 145-159.

[46] Much thanks to Andrew Moon for many helpful conversations pertaining to the arguments in this article, and to an anonymous referee for numerous insightful comments on each part of this article.

Bergmann, Michael. (2006), *Justification Without Awareness*. Oxford: OUP.

Boyce, Kenny and Andrew Moon. (2016), "In Defense of Proper Functionalism: Cognitive Science Takes on Swampman." *Synthese*. 193: 2987-3001.

Conee, Earl and Richard Feldman. (1998), "The Generality Problem for Reliabilism." *Philosophical Studies*. 89.1:1-29.

Comesaña, Juan. (2006), "A Well-Founded Solution to the Generality Problem." *Philosophical Studies*. 129.1: 27-47.

Davidson, Donald. (1987), "Knowing One's Own Mind," *Proceedings and Addresses of the APA*. 60.3: 441-458.

Feldman, Richard. (1985), "Reliability and Justification." *The Monist*. 68.2: 159-174.

———(1993), "Proper Functionalism." *Noûs*. 27.1: 34-50.

Goldman, Alvin. (1976), "Discrimination and Perceptual Knowledge," *The Journal of Philosophy*. 73: 771-791.

——— (1986), *Epistemology and Cognition*. Harvard: Harvard University Press.

———(1988), "Strong and Weak Justification." *Philosophical Perspectives*. 2: 51-69.

Graham, Peter J. (2012), "Epistemic Entitlement." *Noûs*. 46.3: 449-482.

———(2014), "Functions, Warrant, History." *Naturalizing Epistemic Virtue*. Cambridge: CUP. 15-35.

Lewis, David. (1973), *Counterfactuals*. Cambridge: Harvard University Press.

Plantinga, Alvin. (1991), "Warrant and Designing Agents: A Reply to James Taylor." *Philosophical Studies*. 64.2: 203-215.

———(1993a), *Warrant: the Current Debate*. Oxford: OUP.

———(1993b), *Warrant and Proper Function*. Oxford: OUP.

———(1993c), "Why We Need Proper Function." *Noûs*. 27.1: 66-82.

———(1995), "Reliabilism, Analyses, Defeaters." *Philosophy and Phenomenological Research*. 55.2: 427-464.

———(1996), "Respondeo." *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*. Ed. by Jonathan L. Kvanvig. New York: Rowman and Littlefield. 307-378.

———(2000), *Warranted Christian Belief*. Oxford: OUP.

Pritchard, Duncan. (2005), *Epistemic Luck*. Oxford: OUP.

———(2012), "A Defense of Modest Anti-Luck Epistemology." *The Sensitivity Principle in Epistemology*. Ed. by Kelly Becker and Tim Black. Cambridge: CUP. 173-192.

Sosa, Ernest. (1991), *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: CUP.

———(1993), "Proper Functionalism and Virtue Epistemology." *Noûs*. 27.1: 51-65.

———(1996), "Postscript to 'Proper Functionalism and Virtue Epistemology.'" *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*. Ed. by Jonathan L. Kvanvig. New York: Rowman and Littlefield. 271-80

———(1999), "How to Defeat Opposition to Moore." *Philosophical Perspectives*. 13:137-49.

Stalnaker, Robert. (1968), "A Theory in Conditionals." *Studies in Logical Theory, American Philosophical Quarterly*. 2, 98-112.

Taylor, James. (1991), "Plantinga's Proper Functioning Analysis of Epistemic Warrant." *Philosophical Studies*. 64.2: 185-202.

———(1995), "Plantinga on Epistemic Warrant." *Philosophy and Phenomenological Research*. 55.2: 421-426.

Tucker, Christopher. (2014), "On What Inferential Justifies What: the Vices of Reliabilism and Proper Functionalism." *Synthese*. 191.14: 3311-3328.

Wallis, Charles. (1994), "Truth-Ratios, Process, Task, and Knowledge." *Synthese*. 98: 243-269.

Wunderlich, Mark E. (2003), "Vector Reliability: A new Approach to Epistemic Justification." *Synthese*. 136.2: 237-262.