Global Tides

Volume 8 Article 1

1-1-2014

Clarifying the Gettier Objection to Plantinga's Theory of Knowledge

Scott J. Woods

Pepperdine University, scott.woods@pepperdine.edu

In "The Inescapability of Gettier Problems," Linda Zagzebski claims that every theory of knowledge that does not raise the standard for justification problematically high, so as to rule out a tremendous proportion of beliefs we typically take to be justified, will inevitably face Gettier-type problems. That is because, she claims, Gettier problems arise from the gap of independence between truth and justification. In this paper, I defend and strengthen Alvin Plantinga's proper function theory of knowledge against an alleged Gettier case that Zagzebski provides by exploring the relationship between two specific components of his theory. Next, I consider a final objection to the reinforced framework of the proper function theory. I end with a dilemma that faces the proper functionalist and conclude that it is possible for the theory to achieve immunity to Gettier problems.

First, I will provide an overview of Plantinga's proper function theory. It defines knowledge as warranted true belief, where "warrant" encompasses a number of epistemically valuable states of affairs:

> A belief *B* has warrant for S if and only if the relevant segments (the segments involved in the production of *B*) are functioning properly in a cognitive environment sufficiently similar to that for which S's faculties are designed; and the modules of the design plan governing the production of *B* are (1) aimed at truth, and (2) such that there is a high objective probability that a belief formed in accordance with those modules (in that sort of cognitive environment) is true..."[1]

Warrant, then, involves a number of factors: proper functioning of relevant faculties; a proper cognitive environment; a truth-aimed design plan; and reliability. To gain a better understanding of these terms, let's take a look at a few attempted counter-examples that Plantinga discusses and why he believes they are unsuccessful.

First is one devised by Carl Ginet.[2] In driving through the Wisconsin countryside, you see a barn off the side of the road and form the belief that there is a barn there. Unbeknownst to you, the local inhabitants have constructed a massive number of fake barn façades, indistinguishable from real barns, which constitute 75% of the apparent barns in the area. At first glance, then, you have a warranted true belief that is not knowledge. However, as Plantinga explains, your belief is not actually warranted. Our faculties are not designed to operate in a cognitive environment (*CE*) in which our evidence is manipulated to intentionally induce false beliefs, and so the *CE* condition of warrant is not met.

The reader may now be wondering about the term *cognitive environment*, particularly what it is that differentiates a cognitive environment from an

[1] Plantinga, Alvin, *Warrant and Proper Function*, New York (Oxford UP, 1993), 19.
[2] Ibid., 33.

environment *simpliciter*. Although Plantinga never explicitly defines the term, it is clear that he intends for it to encompass a broad spectrum of one's physical environment, including, for example, other people and physical objects.[3] From his application of the term, I take it that one's cognitive environment consists of all factors external to an agent's own mind that influence the formation of a particular belief, including those that influence the reliability of the belief-forming process being used. As reliability is presumably determined modally (i.e. by the proportion of nearby possible worlds in which the belief-forming process yields a true belief), the relevant factors of *CE* will include more than just what influences the belief-forming process in the actual world. In this sense, the composition of the gasses in the air, noise level, lighting, and presence of barn façades can all be considered part of one's cognitive environment.

Take the classic Gettier example of a clock that just happened to stop running exactly twenty-four hours before you look at it. By looking at the clock, you form a true belief through properly functioning faculties in a non-manipulative environment, in accordance with a design plan aimed at truth. While it appears that this is a case of warranted true belief that is not knowledge, Plantinga contends that the belief in question is not warranted. Once again, there is a glitch in the cognitive environment. Our design plan does not intend for our faculties to operate in an environment in which belief-forming tools (such as clocks) are malfunctioning.[4] Once again, a glitch in *CE* prevents the belief from being warranted.

Finally, let's consider Zagzebski's example. It goes as follows: Mary, who has good vision, has seen and identified her husband sitting in his chair numerous times in the past. One night, unbeknownst to her, her husband's brother is visiting and sitting in her husband's chair. As they look sufficiently similar, she mistakes her brother-in-law for her husband, and so forms the belief that her husband is sitting in the living room. Her husband, coincidentally, is in fact sitting in the living room, just out of her view. Mary, then, has formed a warranted, true belief based on faculties functioning properly in an appropriate environment but does not have knowledge. Therefore, Plantinga's account does not avoid Gettier cases.

It seems to me, however, that Zagzebski has overlooked relevant parts of Plantinga's theory. First, she assumes that because "no one is dressing up as her husband to fool her," (208) Mary's environment is "appropriate." However, this is not a sufficient condition for an appropriate *CE*. Rather, for *CE* to be

[3] Plantinga, *Warrant and Proper Function*, 33-35.

[4] It might be easier to view such malfunctioning tools in light of the proper functioning condition instead of the favorable environment condition—that is, our faculties *and* the tools we use to form beliefs must be operating according to a reliable design plan aimed at truth. However, Plantinga does not so qualify proper functioning, so I will leave the thought here.

considered proper, it must be more than non-deceptive. But what other conditions must be met? Conveniently enough, Plantinga addresses a very similar case in the final volume of his warrant on trilogy. In his example, you glance across the street and see Paul, your neighbor, through his window. However, you are unaware that Paul's identical twin Peter happens to be visiting and is in the next room over, hidden from view. While you have what appears to be a warranted true belief, it just so happens to be true by chance, and thus you have a Gettier case for Plantinga's theory.

Once more, Plantinga credits the error to a glitch in the cognitive environment, though he gives a more specific account of it here. He first specifies that a cognitive environment is sufficiently warrant-yielding if it is favorabl*e* for the specific exercise of cognitive powers *(E)* used in producing the belief in question (*B*).[5] He refers to this specified environment as *MBE* (a cognitive environment with respect to *B* and *E*). More significantly, he provides the following definition of favorability, which I will call (F):

> *MBE* is favorable just if there is no state of affairs *S* included in *MBE* but not in *DMBE* [the conjunction of circumstances in *MBE* that are detectable by *E*] such that the objective probability of *B* with respect to the conjunction of *DMBE* and *S* falls below [a reasonably high objective probability].[6]

In other words, a favorable environment is one in which there are no factors that are undetectable to the belief-forming process used that, if known, would remove the believer's warrant for the belief being formed.

Note first that this clearly solves the problem with the cases of Mary and Peter and Paul. In each example, there is some evidence (namely, the presence of the person outside of view) not accessible to *E* (looking across the street, or into the living room) that, if made known to the agent, would render her belief unwarranted. According to (F) then, MBE in these circumstances is not favorable, and so the beliefs are not warranted.

Unfortunately, (F) introduces problems of its own. It is at heart a defeasibility requirement (a counterfactual amendment that restricts warrant to cases in which, if given access to all relevant evidence, the agent would continue to hold the belief in question). Zagzebski critiques such requirements in her article (successfully, in my assessment), and Plantinga's theory falls victim to the same concerns. The objection is that such stipulations link justification (or warrant) too closely with truth, and so render a great number of beliefs we

[5] It is important to see that we cannot assess the favorability an environment without considering it *in relation to the particular exercise of cognitive powers* (*E*) *and the belief being formed* (B), as this will become crucial later on.

[6] Plantinga, Alvin, *Warranted Christian Belief*, New York (Oxford UP, 2000), 159-60.

typically take to be warranted *un*warranted. Since, for practically any false belief, there will be some evidence of its falsehood undetected by the agent that, if known, would render that belief unwarranted, (F) makes truth a near prerequisite for warrant. This addition does not escape Zagzebski's attack and is not a satisfactory response.[7]

Is there a better defense one can afford Plantinga? I believe so. We need to push deeper into the notion of a favorable *CE*. Recall that one of Plantinga's criteria for warrant is that a belief be formed by faculties "functioning properly in a cognitive environment sufficiently similar to that for which S's faculties are designed."[8] While some conditions for a favorable environment seem to be all-or-nothing, others clearly admit of degrees—that is, certain environmental factors provide a range for a number of favorable environments. I believe it is this latter sort that concerns the examples we are considering, and so will now turn to consider it.

First, take an example. Suppose you are forming a belief as to whether or not your grandmother is in the dining room at her retirement center. Standing outside the dining room, you see a single walker along the wall and recognize it to be of the same unique color and style of your grandmother's. You thereby form the belief that your grandmother is in the dining room. Grant that *CE* is about as appropriate as it could be; the lighting is excellent, your field of view unobscured, no geriatric prankster is trying to play a joke on you, etc. Further, assume that all other warrant-criteria are met: your faculties are functioning properly according to a design plan successfully aimed at truth, and that such a belief formed under these conditions is highly likely to be true. We have, then, a situation in which your belief is warranted *in a highly favorable environment*.

Now, keeping all other factors constant, imagine a revision to the environment. The manager of the retirement center was so impressed with your grandmother's sense of style that he recently bought every resident a new walker identical in color and model to your grandmother's. Following the same process as before, your belief that your grandmother is in the dining room is *not* warranted. Presumably, this is because your faculties are not designed to identify a particular object in an environment in which there are a large number of objects identical (or nearly identical) to it. Your environment is no longer favorable.

Imagine one final change in the scenario. Although all the residents of the center now have identical walkers, due to a rather unfortunate but minor collision your grandmother's has a very specific pattern of scratches on one side of it (these scratches exist unnoticed in the previous example). In forming your belief that your grandmother is in the dining hall, you notice this specific pattern of scratches

[7] It is worth noting that Plantinga himself showed reservations about (F), though not exactly for the reasons specified here; WCB 160-161.
[8] Plantinga, *Warrant and Proper Function,* 19.

on the walker against the wall. Although *CE* is no more favorable in the case in which you notice the scratches than in the case in which you do not, your belief is warranted. Here, then, we have a situation in which your belief is warranted *in a significantly less favorable environment* than the first.

If my assessment of the above three examples is correct, then we have a range of cognitive environments in which a particular belief can be warranted. What is it that accounts for this range? It is the same thing that differentiates the second and third cases. Recall the previous mention of MBE. In order to see whether or not *CE* is favorable, we must first contextualize it with respect to *B* and *E*. *B* provides us with the end for which we are assessing favorability (are we looking to see if your grandmother is in the dining hall, or if the dining hall is open?), while *E* provides us with the method (a poorly lit room may be unfavorable to visual assessment, but pose no inhibitions to a tactile one). *E, B*, and *CE* are inextricably tied together in MBE, and it is only in MBE that favorability can be properly assessed.

In particular, the connection between *E* and *CE* concerns us here. As it is only for the conjunction of *E* and *CE* that favorability can be determined, a change in one can compensate for an opposite change in the other. This explains how the first and third scenarios above can both yield warranted beliefs while the favorability of *CE* shifts significantly. Your specific exercise of visual assessment *E* is much more precise in case three, with respect to the relevant evidence, than in case one. Thus, you are still able to have a warranted belief when the favorability of *CE* with respect to visual assessment declines. More significantly, this explains how the same belief being formed in the same cognitive environment by the same person has the potential to be either warranted or unwarranted. The more reliable *E* is, the less favorable *CE* must be, and vice versa.

One may be tempted to make a generality objection here. The classic objection to reliablism preys on the ambiguity of "belief-forming process." Does it include only internal faculties, or external tools and the environment as well? What determines which factors are relevant or not? If one defines the term narrowly enough, a belief-forming process could be so specifically indexed (to a time, a place, persons, etc.) as to only have a few possible instantiations. If one defines it too broadly, a single belief-forming process may encompass a number of unrelated instances. Either way, the reliability of any given process will depend heavily on this definition. Similarly, Plantinga must provide a satisfactory, non-arbitrary definition of *E* if his theory is not to fall prey to the same objection.

Such a generality objection does not apply to the proper function theory, however, as Plantinga has already adequately defined *E*. Recall that *E* is not defined as a *belief-forming process*, but much more narrowly as a *specific exercise of cognitive faculties*. The relevant *E* for any given belief-formation,

then, will be the type of faculties engaged, without reference to external conditions. *E* is just those cognitive faculties that are actually engaged in forming the belief in question. Further, reliability on Plantinga's theory is based on the conjunction of *E* and *CE* in reference to a specific belief (MBE). By specifically indexing reliability to the particular conditions for the formation of a particular belief, this theory avoids the ambiguity that typically plagues reliablist theories. As such, it does not succumb to generality objections.

I will clarify the notion of MBE with reference to Michael Strevens' approach to causal explanation.[9] In advocating for what he calls the kairetic account of causation, he uses the image of a brick breaking a window to explain how interdependent factors can be involved in causation:

> Suppose that I wish to explain the breaking of a window. As it happens, I threw a rather heavy brick at the window. Question: did the mass of the brick make a difference to the fact of the window's breaking? Answer: if you mean, did the exact mass of the brick make a difference, then no. The brick weighed 2 kg, but had it weighed 1 kg or 3 kg, the window would have broken just the same.[10]

This is analogous to the role of either *E* or *CE*. The *exact* favorability of one variable does not make or break whether or not a belief is warranted. Strevens continues:

> Assume that, in order to entail that the window breaks, a veridical causal model of the breaking must entail that the brick has a momentum of at least 3 kgms and no more than 20 kgms. Now suppose that a model M specifies that the brick has a mass between 1 kg and 5 kg. If M is to entail that the window breaks, it must limit the velocity of the brick to between 3ms and 4ms in order to guarantee that the brick's momentum falls within the required range. (The brick's momentum is the product of its mass and velocity.) Another veridical model, which stipulates a narrower range for the brick's mass—say, between 2 kg and 4 kg—can stipulate a wider range for the velocity... There is one way to describe the mass *m* and the velocity *v* of the brick so that the resulting model is veridical, entails the window breaking, and generates the models described above and any other competitors:
>
> $$3 \leq mv \leq 20$$
>
> A model incorporating this description, then, is the most abstract veridical, deterministic causal model for the window breaking...[11]

I believe this formula transfers well into Plantinga's theory. MBE is analogous to the brick's momentum, and its yielding warrant to the window breaking, while

[9] Strevens, Michael, The causal and unification approaches to explanation unified—causally, *Noûs* 38 (1):154–176.
[10] Ibid., 167.
[11] Ibid.,168-9.

mass and velocity are the equivalent of *E* and *CE*. When one of these values (or ranges) is specified, it restricts the value or range of the other.[12] Our formula for this part of the warrant equation (that dealing with CE and reliability), then, is something like the following:

$r \leq$ fMBE,

where *r* is a sufficiently high probability of a belief being true, and fMBE is the favorability of MBE, seen as a specific conjunction of *E, B,* and *CE*.

Admittedly, there is a significant difference between Strevens' original formulation and the one I provided above. While the mass and velocity of a brick are specific and quantifiable values, the favorability of *E* and *CE* (fE and fCE respectively) cannot properly be understood to hold values independent of each other. I do not think this speaks against my application, however. Instead of fMBE being the product of specific values, we can understand *E* and *CE* as holding independent *potentialities* for favorability—that is, for a specified *E*, it can be said how reliable the method is in any given environment, and vice versa for *CE*. Even if these potentialities are not quantifiable and we are thereby unable to precisely calculate fMBE, we can still claim that they in fact determine its value as well as the precision of the other value needed for warrant.

I need to say more about *E*'s role in this theory. First, I will refer to the *favorability* of *E* being its potential to yield a true belief without a specified *CE*, as explained above. Further, the type of faculties involved in *E* establishes the degrees of favorability with respect to fE. In other words, the faculties used in forming the belief (i.e. vision/sense perception, memory, etc.) determine how fE can move along the scale. In the retirement center example, using your sense of sight determines what it takes to improve or worsen fE: glancing at the walker on the wall would yield a lower value for fE than looking closely (and in doing so noticing the unique scratches on her walker). A better assessment of the situation can thus compensate for a less favorable environment.

I anticipate a vagueness objection here. Not only is it difficult (perhaps even impossible) to non-arbitrarily quantify the favorability of *CE* or *E*, but we must also quantify *r* to yield a functional formula. Unfortunately, I have no response here. Plantinga himself admits that his theory suffers from vagueness, and that problem carries over to the current argument. It is an admitted weakness, and not one that is within the scope of this paper to attempt to solve.

Now, let's consider the implications of the current theory on the Mary and Peter/Paul cases. For Mary, it seems that when her husband is alone in the living room, fCE is rather high, and so fE can be relatively low for MBE to be sufficient for warrant—she only need glance briefly to be warranted in her belief that he is in the living room. When her brother-in-law is visiting, however, fCE drops, and

[12] One notable disanalogy is that in relation to Plantinga's theory, we won't expect to have an upper bound (other than the conjunction of an ideal *CE* and *E*).

so fE must rise to compensate. She may, for example, need to look longer or move closer to her brother-in-law in order to sufficiently raise MBE. However, in doing so, she would recognize him to not be her husband. Similarly, if your neighbor's identical twin is visiting, CE is very unfavorable for forming the belief that Paul is home, and so it may not even be possible to achieve warrant by looking from across the street. In neither case, then, is the agent's belief warranted, and so both fail as Gettier problems.

I believe that what I provided above is the strongest defense that can be provided for Plantinga's theory against Gettier cases, and it certainly defeats Zagzebski's example in her paper. However, there is still a response that Plantinga's opponent can give. What the above framework does is explain how Gettier situations lower the favorability of one's environment and so remove the warrant one would normally have for a belief (unless the subject compensates with a sufficiently precise exercise of cognitive faculties). What the critic will rightly point out is that this does not preclude the possibility of Gettier problems. All a Gettier environment does is raise the bar for justification. As long as it is *possible* for an agent to meet this heightened requirement but still be mistaken-in other words, to have an accidentally true warranted belief in a Gettier environment-then a Gettier case can be devised.

Can Plantinga's theory escape a second time? I see one response he can give, and it is not an entirely satisfying one. If he is to avoid allowing Gettier cases, Plantinga must deny that it is possible for someone to have an accidentally true warranted belief in a Gettier environment. This would mean that acquiring warrant in such an environment *requires* the (non-accidental) truth of the belief in question. One concern here is that, prima facie, it seems to run right into the second horn of Zagzebski's original dilemma: it links warrant too closely with truth. However, my defense demonstrates how the proper function theory only links warrant this closely in extremely unfavorable cognitive environments such as those seen in Gettier cases. In typical belief-forming processes where fCE is not so low, the theory does not require such a strong connection with truth, and so it avoids the unpalatable consequences that defeasibility requirements and counterfactual conditions have on justification. In other words, it is a much smaller bullet to bite.

A further concern is that the above claim commits the *No True Scotsman* fallacy. One might contend that there is no good reason why a Gettier environment should require truth for warrant instead of just a very high degree of precision; Plantinga just adds an *ad hoc* addendum to his criteria for warrant that says "... and *B* is not formed in a Gettier environment (unless it's non-accidentally true!)." Unfortunately, I do not believe there is anything Plantinga can do here except dig in his heels. Getter environments are unfavorable to such a degree that fMBE cannot be sufficient for warrant unless *E* is precise enough to guarantee the

non-accidental truth of *B*. Unfortunately, the general vagueness of fCE, fE, and fMBE prohibits us from verifying this move with specific numbers. Nonetheless, it is the move Plantinga must make. When her brother-in-law is in the room, Mary must assess the environment well enough that she *cannot* be mistaken about the identity of the man in her husband's chair. In driving through fake-barn country, you cannot be warranted in believing any apparent barn is in fact a barn unless *E* is precise enough to guarantee that you would distinguish between a real barn and a façade. In short, given Plantinga's original criterion of a proper cognitive environment, the nature of the Gettier environment just *does* make non-accidental truth a prerequisite for warrant.

At this point, the defender of the proper function theory is left with a dilemma: either she adopts the above account of warrant in Gettier cases, or she admits that Gettier problems can be devised for her theory. While I suspect the former option to prove more undesirable than the latter, I will leave this for her to decide. Nonetheless, I have shown how a proper understanding of MBE, the joint relationship of one's cognitive environment and exercise of cognitive faculties, defeats Zagzebski's initial Gettier scenario, and I have clarified and strengthened Plantinga's theory against other potential objections. Although the proper function theorist may be in the best position to escape Gettier problems, she is not able to do so without cost.

Works Cited

Plantinga, Alvin. *Warrant and Proper Function*. New York: Oxford UP, 1993. Print.

Plantinga, Alvin. *Warranted Christian Belief*. New York: Oxford UP, 2000. Print.

Strevens, Michael (2004). The causal and unification approaches to explanation unified—causally. *Noûs* 38 (1):154–176.

Zagzebski, Linda (1994). The inescapability of Gettier problems. *Philosophical Quarterly* 44 (174):65-73.