How Gödelian Ontological Arguments Fail

Matthew W. Parker

Abstract.

Ontological arguments like those of Gödel (1995) and Pruss (2009; 2012) rely on premises that initially seem plausible, but on closer scrutiny are not. The premises have modal import that is required for the arguments but is not immediately grasped on inspection, and which ultimately undermines the simpler logical intuitions that make the premises seem plausible. Furthermore, the notion of necessity that they involve goes unspecified, and yet must go beyond standard varieties of logical necessity. This leaves us little reason to believe the premises, while their implausible existential import gives us good reason not to.

Gödel (1995) introduced a new class of formal arguments for the existence of God, appealing to a notion of "positive" property and applying modal logic. Gödel's premises were later shown to imply modal collapse, i.e., if they are true, then everything true is necessary (Sobel 1987), and then to be inconsistent (Benzmuller and Paleo 2016). However, Anderson (1990) and Pruss (2009; 2012) give simpler versions that avoid these problems.

We will focus here on one of Pruss's formulations, but our observations apply generally. We will see that Pruss's premises are not as innocent as they first appear. Once their modal import is unpacked, and their unclear foundations exposed, they are not very plausible at all.

◇ □ ◇

Pruss (2009) proposes the following axioms:

F_1: If A is positive, then ~A is not positive.

F_2: If A is positive and A entails B, then B is positive.

N_1: Necessary existence (NE) is positive.

N_2: Essential omnipotence (EO), essential omniscience, and essential perfect goodness are each positive.

He argues from F_1 and F_2 that any two positive properties are "compossible", i.e., it is possible something exists that has both properties. Thus, it is possible that something necessarily exists and is essentially omnipotent, and applying the axioms of S5, he deduces that something *does* necessarily exists and is essentially omnipotent. (For simplicity we will only discuss omnipotence here; the treatment of omniscience and goodness is entirely parallel. Also, we will not take issue with the axioms of S5.)

Pruss suggests several possible meanings for 'positive,' and we will consider a few of them below, but in fact, it does not matter much what 'positive' means. As long as we can find some class of properties that satisfies Pruss's axioms, the derivation goes through. One such class might consist of just NE, EO, and all the properties they imply. If these do not include two contradictory properties A and ~A, Pruss's axioms are satisfied.

This ought to be worrying. It seems all too easy to concoct a notion of positivity that will entail the existence of God, or for that matter, of anything, provided it is consistent with NE. And it seems that nearly any property would be consistent with NE. No standard system of logic rules out any possible being existing necessarily. Thus we can help ourselves to all

manner of parody arguments. If NE is consistent with being essentially a nine-legged frog (E9LF), then take all the properties implied by NE or E9LF, call them positive, and voilá, there necessarily exists a nine-legged frog. But surely the existence of a nine-legged frog would be an accident, so something has gone wrong.

Our mistake was to suppose that NE and E9LF were really consistent in the relevant sense. If we consider carefully what NE, 'essentially', and 'compossible' mean in this context, we will see that this is not plausible.

What is NE? Pruss gives an elaborate formulation:

D_1: x *necessarily exists* if and only if $\exists F[\Box Fx \,\&\, \Box \exists y Fy \,\&\, \Box \forall y(Fy \rightarrow y = x)]$.

All we need to note here is that, for any *x*, NEx is equivalent to $\Box$NEx. Pruss argues that NEx implies $\Box$NEx (2009, 352), and the converse is trivial. Let us say that a property is *necessitated* if it can be expressed in the form $\Box\varphi(x)$. So NE is a necessitated property.

The property of *essentially* having a property A is also a necessitated property. Pruss never defines 'essential,' but its meaning is clear from his argument (2009, 352): x has A essentially if and only if x has A in every accessible possible world where x exists, i.e., $\Box$(x exists $\rightarrow$ Ax). (Logicians differ on how to formulate 'exists' in this context, but that need not concern us here.) So essential properties are necessitated too, including EO and E9LF.

Necessitated properties are not simple, atomic properties that might or might not hold of anything. In S5, $\Diamond\Box$Ax implies $\Box$Ax, so it is not *possible* that a thing necessarily has A unless it *does* necessarily have A. It follows from this and Pruss's Lemma 4 (2009) that $\Diamond\exists x(NEx \,\&\, E9LFx)$ can only hold if $\Box\exists x(NEx \,\&\, E9LFx)$ does. In other words, necessary existence and being essentially a nine-legged frog are not compossible unless something

actually *has* both those properties necessarily. But Pruss argues that if two properties are not compossible, then one entails the negation of the other. So unless something really does have NE and E9LF, and has them necessarily, NE and E9LF are *not* consistent in the relevant sense. Intuitively they seem consistent, and indeed no standard system of logic tells us that they are contradictory, but they are not consistent in the sense we need unless they are necessarily co-instantiated, and that is quite implausible. There clearly does not have to be nine-legged frog. Thus the assumption that NE and E9LF could be contained in a class of properties satisfying Pruss's axioms is not nearly as plausible as it first seemed. We judged it plausible without attending to its full modal import in the context of S5.

Now, Pruss does not merely assume that NE and EO are consistent. He supposes his axioms are plausible and *derives* the result that NE and EO are compossible, and then that they are co-instantiated. However, I claim that he makes essentially the same mistake as we have with our nine-legged frog by supposing the axioms are plausible. The first clue that they are not plausible is that they imply a very strong, physically significant, intuitively contingent proposition, namely that an omnipotent being exists. But rather than rely on the implausibility of the conclusion, let us scrutinize the axioms, especially F_2.

What is the sense of 'entails' in F_2? Pruss writes, "If we assume toward contradiction that A and B are incompossible, then A entails ~B" (2009, 349). So the sense of 'entails' must be one that is implied by incompossibility. The strongest entailment implied by $\sim\Diamond\exists x(Ax \ \& \ Bx)$ is $\Box\forall x(Ax \rightarrow \sim Bx)$. Thus his 'entails' is no stronger than necessitated universal material implication. This is weaker than logical entailment, for the necessity involved cannot be necessity in virtue of any standard system of logic. After all, the existence of an essentially omnipotent being is not a truth of any such system. In other words, Pruss's '$\Box$' represents some unspecified notion of necessity, presumably metaphysical, that is not equivalent to validity in any standard system of logic. Put yet

another way, '□' represents truth in all "possible worlds" within some *intended* model or range of models rather than truth in all possible worlds in all models permitted by some standard semantics. Hence Pruss's 'A entails B' just means that, in every *intended* possible world, everything that has A has B.

<div align="center">□ ◇ □</div>

Given that, and assuming that positivity satisfies F_1 and F_2, is it plausible that NE and EO are positive? I say no.

One sense of 'positive' that Pruss mentions is, entailing no detraction or limitation. It might seem plausible at first that NE and EO are positive in this sense, and if so, it follows that everything they entail is positive. But if in fact there is no intended possible world where something has both NE and EO, then NE *does* "entail" some limitation or detraction, namely ~EO, and EO "entails" the limitation or detraction ~NE. And we have no reason to suppose that NE and EO are co-instantiated in any of the intended worlds, especially since we are given no hint what possible worlds are intended. So, though it was not initially apparent how NE or EO could entail anything negative, we can now see that the relevant kind of entailment is actually quite easy to come by. All it requires is that not every intended possible world contains an omnipotent being, and we have no independent reason to think every intended world does contain one. So we have no reason to believe that neither NE nor EO entails any detraction or limitation.

What about the Leibnizian sense of 'positive'? Gödel (1995) hints at such an interpretation, suggesting that 'positive' might mean pure "attribution" in the sense that the disjunctive normal form of the property contains at least one disjunct without negation. Two such properties cannot be logically contradictory if the elementary properties are all possible

and independent. If NE and EO are positive in this sense, then we can skip Pruss's axioms and infer compossibility directly, which leads quickly to the theistic conclusion. But as Adams points out in his introduction to Gödel 1995, this approach depends on "the controversial assumption that the only way in which properties can be incompatible is by formal contradiction arising from negation involved in their construction" (1995, 398). Given the modal demands of Gödelian arguments, this is not just a controversial assumption but flatly false. As we saw, two *necessitated* properties are contradictory if they are not actually co-instantiated, and this is not due to negation in their construction, but to their necessity operators. Nor can those operators be reduced to disjunctions and negations, for as noted, the necessity involved cannot be merely logical necessity.

<p style="text-align:center">◇ □ ◇</p>

The positivity and consistency of NE and EO seems plausible before we unpack (even in broad stokes, as we did here) the modal content of positivity, consistency, NE, and EO. Once we attend to that modal content, though, the plausibility collapses.

It might be tempting to view this as merely another case of, "One person's modus ponens is another's modus tollens," but there is more going on here. First, the premises appear plausible because we initially misconstrue them in simple logical terms, but in fact they pack modal import that makes them much less plausible. Second, judging the plausibility of the premises requires us to evaluate modal claims founded on an unspecified notion of necessity, which must go beyond standard theories of logical necessity, but for which we have been given no basis of evaluation.

Pruss is wise to say that his argument is not a proof of God's existence and merely confers some incremental credibility on the conclusion (2009, 352). But given that the

apparent plausibility of the axioms dissipates once we attend to their actual content, it is doubtful that such superficial plausibility confers any *rational* credence.

References.

Adams, R. M. (1994). Introductory note to Gödel *1970. Kurt Gödel collected works, 3, 388-402.

Anderson, C. A. (1990). Some emendations of Gödel's ontological proof. Faith and philosophy, 7(3), 291-303.

Benzmüller, C., & Paleo, B. W. (2016). The inconsistency in Gödel's ontological argument: A success story for AI in metaphysics. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*—IJCAI, Menlo Park, CA: AAAI Press.

Gödel, K. (1995). Ontological proof. Kurt Gödel collected works, 3, 403-404.

Pruss, A. R. (2009). A Gödelian ontological argument improved. Religious Studies, 347-353.

-----. (2012). A Gödelian Ontological Argument Improved Even More. Ontological Proofs Today, 50, 203.

Sobel, J. H. (1987). Gödel's ontological argument. On Being and Saying: essays for Richard Cartwright, MIT Press, Cambridge MA, 241-261.