Penultimate draft—please cite/quote from the published verison

On the PROVER9 Ontological Argument[1]

T. Parent (Virginia Tech)
parentt@vt.edu

1. Introduction

Should metaphysics be trusted to the machines? That is a larger question prompted by the (2011) paper by Oppenheimer & Zalta (henceforth, "O&Z"). There, the authors present their (1991) ontological argument, after being streamlined by PROVER9, an impressive automated reasoning engine. Like their (1991) argument, the authors end up rejecting the new version; however, I shall argue that the theist's case has been oversimplified. Is this the recurrent danger of automated processing? Perhaps. But the PROVER9 argument promises to be the most rigorous, most compelling version yet of the non-modal ontological argument.[2] Accordingly, in what follows, I first fill the gap in this exemplar of a philosophical argument.

But ultimately, like O&Z, I also attempt to diagnose the fundamental error of the argument. This is of interest, insofar as it promises to yield a more general lesson about argumentation. (It is also of interest, insofar as the non-modal ontological argument continues to have defenders; see e.g., Baker & Matthews 2010, Baker 2013). Here, O&Z's diagnosis is first shown applicable only to the oversimplified argument. The diagnosis then offered in its place is that the argument commits a fallacy that is common to ontological arguments (including modal ontological arguments, of which I say a word later on).

---

[1] My thanks to Ed Zalta and an anonymous referee for comments and discussion on this material.

[2] The PROVER9 argument is a *non-modal* ontological argument in that God is defined simply as a being that exists. The contrast is with the modal argument from, e.g., Plantinga (1974a, b), where God is a *necessarily* existing being.

For later reference, call the fallacy "the existential fallacy." The basic point is that one can define a term like 'God' however one wishes—yet it does not follow that *there is* an object satisfying the definition. This is what lies behind Gaunilo-type objections. Adapting an example from Salmon (1987), consider the following definition:

(D1) 'Existicorn' denotes *x* only if *x* is a unicorn that exists.

Here, the reasoning of the ontological argument would be, in (very) short form:

(a) 'Existicorns exist' is analytic. [from (D1)]

(b) 'Existicorns exist' is true. [from (a)]

(c) So, existicorns exist. [from the T-schema and (b)]

Thus, the argument putatively shows that a kind of unicorn exists. However, the mistake is that the definition does not render 'Existicorns exist' analytic, but rather a *conditional* claim: *If* anything is an existicorn, it exists.

If the PROVER9 argument commits this fallacy, it would be unexpected. For the reasoning has the virtue of being "existentially timid" twice over. For one, O&Z formulate it in a language governed by a free logic. Second, the language also includes terms that denote "nonexistent" objects. One might liken them to Anselmian objects that exist only "in the mind." (The metaphysics of such objects may be puzzling, but let that pass.) Concurrently, O&Z introduce a predicate '*Ex*' which is satisfied by all and only the existing objects in the domain. (Note, then, that the extension of '*Ex*' is narrower than the range of the existential quantifier.)

In the regimented language, O&Z also use a two-place predicate '*Gxy*', meaning "*x* is greater than *y*," and a one-place predicate '*Cx*' to mean "*x* is conceivable." In addition, let 'ϕ_1' be shorthand for the compound predicate '$Cx \wedge {\sim}\exists y(Gyx \wedge Cy)$' ("*x* is a greatest conceivable being”). Consider also the following lemma (proven in O&Z 1991, from uncontentious meaning-postulates about ‘greater than’):

(L2) $\exists x\phi_1 \supset \exists!x\phi_1$

This says that if there is a greatest conceivable being, it is unique. Hence, I typically talk of “*the* greatest conceivable being,” symbolized by ‘$\iota x\phi_1$’ and by the name ‘g’.

Three premises are offered at this point, and two are really theorems of a single axiom, the “Description Axiom,” meant to codify the Russellian analysis of definite descriptions:

(DA) $\psi[\iota x\phi/z] \equiv \exists y(\phi[y/x] \,\&\, \forall u(\phi[u/x] \supset u = y) \,\&\, \psi[y/z])$, *where* ψ *is an atomic or identity formula*.

The relevant theorems are as follows (“Description Theorem 2” and “Description Theorem 3”):[3]

(DT2) $\exists y(y = \iota x\phi) \supset \phi[\iota x\phi/x]$

(DT3) $\psi[\iota x\phi/z] \supset \exists y(y = \iota x\phi)$, *where* ψ *is an atomic or identity formula*.

The former is the truism that if there is a unique thing that is ϕ, then it is ϕ. Whereas, the latter says that if replacing the variable ‘z’ with an iota-term yeilds a truth, then the iota-term denotes an object in the domain (when ψ is an atomic predicate or an identity formula with at most z-free). Though again, the object may turn out to be a “non-existing object.”

The non-logical premise (labeled “Premise 2” in O&Z 2011) is:

(P) $\sim E\iota x\phi_1 \supset \exists y(Gy\iota x\phi_1 \wedge Cy)$

This is Anselmian idea that if the greatest conceivable being fails to exist in reality, then there is a conceivably greater being. The argument now runs as follows (2011, p. 345):

[3] O&Z’s Description Theorem 1, Lemma 1, and Premise 1 are omitted in this discussion. Part of PROVER9’s achievement was to show that these are unnecessary to the argument.

The PROVER9 Argument

1. $\sim E\iota x\phi_1$	[Assume for *reductio*]
2. $\exists y(Gy\iota x\phi_1 \wedge Cy)$	[from 1 and (P)]
3. $Gh\iota x\phi_1 \wedge Ch$	[from 2, '*h*' arbitrary]
4. $Gh\iota x\phi_1$	[from 3]
5. $\exists y(y = \iota x\phi_1)$	[from 4 and (DT3)]
6. $\phi_1\iota x\phi_1$	[from 5 and (DT2)][4]
7. $\sim\exists y(Gy\iota x\phi_1 \wedge Cy) \wedge C\iota x\phi_1$	[from 6 and def. of 'ϕ_1']
8. $\sim\exists y(Gy\iota x\phi_1 \wedge Cy)$	[from 7]
9. $E\iota x\phi_1$	[by *reductio,* contradiction at 2 and 8]
10. *E*g	[from 9, and the def. of 'g']

N.B., to imply at line 5 that 'ϕ_1' denotes may seem like an existential fallacy. But line 5 just says that 'ϕ_1' denotes *some*thing, albeit perhaps a non-existing thing. So as long as we do not presume the thing also exists in reality, the move to line 5 seems valid.

O&Z ultimately see the PROVER9 argument as a *reductio* on (P) rather than on '$\sim E\iota x\phi_1$'. But the falsity of (P) is noteworthy, they say, since it does not exhibit the typical mistakes of ontological arguments: "There is no de re/de dicto ambiguity in this premise, given that it has already been formally represented. Moreover, it doesn't presuppose, in the technical sense of presupposition, that anything answers to the description" (p. 348). Nevertheless, O&Z go on to give an independent case that (P) is either false or question-begging, and as things currently stand, their verdict strikes me as correct. (See also Garbacz 2011, p. 587.)

4 (O&Z harmlessly skip line 6, but I include it in case it promotes clarity.)

2. Further Dialectics

However, a defender of the ontological argument should balk at PROVER9's argument. For she typically does not leave (P) undefended; indeed, the additional defense is how she blunts O&Z's question-begging horn. Traditionally, (P) is justified by the idea that it is greater to exist both in the mind and in reality than just in the mind. Or as a weaker version of this: Assuming y exists in reality and z does not, if either y or z is the greatest conceivable being, then it is y. Symbolically:

(P*) $\forall y \forall z\{(Ey \wedge \sim Ez) \supset [(y = \iota x\phi_1 \vee z = \iota x\phi_1) \supset y = \iota x\phi_1]\}$

It is surprising that the PROVER9 argument neglects this well-known element of ontological arguments. Perhaps it was eclipsed by the "streamlining" goal—still, there is no advantage in paring down an argument if the result becomes question-begging.

Informally, the ontological arguer could argue (P) from (P*) as follows. Suppose for *reductio* that $\iota x\phi_1$ exists only in the mind. Then, it is either $\iota x\phi_1$ or h (where h is an arbitrary denizen of both the mind and reality). Yet (P*) says that if h really exists and $\iota x\phi_1$ does not, then h is the greatest conceivable being (since one of them is). Hence, $h = \iota x\phi_1$. But that contradicts the indiscernability of identicals, since ex hypothesi h exists in reality whereas $\iota x\phi_1$ does not. So by *reductio*, the consequent of (P) follows.

Toward a formal reconstruction, two non-logical premises are required beyond (P*)—but these are just the claims that (i) $\iota x\phi_1$ is conceivable, and (ii) something exists in reality. N.B., I call (i) a non-logical premise, even though it may seem definitionally true. However, we shall see that its truth has non-trivial effects within O&Z's logic—yet ignore that for now. As a prior matter, observe that (i) and (ii) are symbolized at 1 and 2, respectively, and enable us derive (P) in the manner indicated:

<u>The argument for (P)</u>

1. $C\iota x\phi_1$	[premise]
2. $\exists xEx$	[premise]
3. Eh	[from 2, '*h*' arbitrary]
4. $\sim E\iota x\phi_1$	[assume for conditional proof]
5. $\sim\exists y(Gy\iota x\phi_1 \wedge Cy)$	[assume for *reductio*]
6. $\exists z(z = \iota x\phi_1)$	[from 1 and (DT3)] [5]
7. $\iota x\phi_1 = \iota x\phi_1$	[from '$\forall x(x = x)$' and 6][6]
8. $h = \iota x\phi_1 \vee \iota x\phi_1 = \iota x\phi_1$	[from 7]
9. $Eh \wedge \sim E\iota x\phi_1$	[from 3 and 4]
10. $h = \iota x\phi_1$	[from (P*), 9, and 8]
11. $h \neq \iota x\phi_1$	[from 9, the Indiscernability of Identicals, and 6]
12. $\exists y(Gy\iota x\phi_1 \wedge Cy)$	[from 5–11, by *reductio*]
(P) $\sim E\iota x\phi_1 \supset \exists y(Gy\iota x\phi_1 \wedge Cy)$	[from 3–12, by conditional proof and ∃Elim]

[5] Note here that line 6 = line 5 of the PROVER9 argument. Accordingly, lines 3–5 of the PROVER9 argument can be eliminated as redundant, assuming the above argument for (P) precedes it. Even above, one could just start with line 6 as a premise ("the greatest conceivable being exists at least in the mind), thus rendering 1–5 idle. (Indeed, O&Z themselves took an equivalent to 6 as a premise in their 1991.) Regardless, I have protracted things a bit, since 1 in the argument for (P) is instructive when seeking the basic error in non-modal ontological arguments.

[6] Recall that free logic requires an additional premise for universal instantiation. Where τ is a name or descriptor for a (possibly nonexistent) object, the rule is '$(\forall x\phi \,\&\, \exists y\, y = \tau) \supset \phi(\tau/x)$'. This is why line 6 is needed to derive line 7 from '$\forall x(x = x)$' . Similarly, line 6 is needed to derive line 11 from the Indiscernability of Identicals.

If this further argument dispels the question-begging worry, how else might an opponent reply to the ontological arguer? One could mimic the earlier maneuver, where the argument is now taken to be a *reductio* on (P*) rather than on '$\sim E\iota x\phi_1$'. But the typical ontological arguer insists that rejecting (P*) is *incoherent*—for the *greatest* conceivable being exists in reality by definition, from which (P*) follows.

Yet if (P*) is definitional, then a question arises about the existential fallacy. To reiterate, if $\iota x\phi_1$ is defined as existing in reality, it is still open whether *there is* anything satisfying the definition, *a fortiori,* satisfying (P*). Still, we can agree that $\iota x\phi_1$ is conceivable; the "greatest conceivable being" is not gibberish, after all.[7] We can even allow that being exists in the mind, as the thing we all refer to as "the greatest conceivable being" (cf. Baker & Matthews 2010). So $\iota x\phi_1$ exists at least in the mind. But here is where the theist makes her move: If $\iota x\phi_1$ exists in the mind, and by the definitional (P*) it also exists in reality, then $\iota x\phi_1$ also exists in reality!

Compared to PROVER9's version, I submit that this is a more serious case for theism. Regardless, the argument still commits the existential fallacy. That's because for all that has been shown, the "greatest conceivable being" may be an *impossible* object in the manner of "the largest ordinal number."[8] Such a thing is conceivable, in the sense that its description is not

[7] Cf. Oppenheimer & Zalta (2007): "Anselm argues that there is something such that nothing greater can be conceived by appealing to the fact that we understand the definite description 'that than which nothing greater can be conceived'…We actually agree with him on this point" (p. 2).

[8] One difference between $\iota x\phi_1$ and "the largest ordinal" is that plausibly the former is not *logically* but only *metaphysically* impossible. Still, metaphysical impossibility would thwart the argument in roughly the same way.

gibberish. But like the ordinals, "greatness" may have no upper bound.[9] In which case, no *possible* being would be the greatest one could conceive of.

To say that $\iota x\phi_I$ is impossible is to say it cannot exist in reality. That is so, even if the thing is defined as existing in reality.[10] After all, if it is impossible on independent grounds, then "existing in reality" is just a *further* impossible condition on the object—it is impossible for an impossible object to exist in reality.

The criticism holds, even if the theist replies that it is greater to "non-impossibly exist in reality" than otherwise. Here too, if $\iota x\phi_I$ is impossible for a prior reason, that just compounds its impossibility even more. (If $\iota x\phi_I$ is antecedently impossible, it cannot be non-impossible.[11]) The same point applies if $\iota x\phi_I$ is defined as "non-impossibly, non-impossibly existing in reality," and so on. More broadly, if a definition has an unsatisfiable predicate, adding more predicates will not change that fact. So for all that has been shown, (P*) is not satisfied by anything in reality,

---

[9] Aquinas once raised a similar doubt: "For any given thing, in the imagination or in reality, one greater can be conceived" (*Summa Contra Gentiles* 1a.q2). But Baker (2013, p. 29) reads this as implying that $\iota x\phi_I$ is not conceivable. And that is false, assuming (per usual) that a non-gibberish descriptor is sufficient for conceivability. Regardless, my point is not a challenge to the *conceivability* of $\iota x\phi_I$; it is rather that if "greatness" lacks an upper bound, then $\iota x\phi_I$ is *not possible*, even if conceivable.

It might be said that *conceivable* greatness has an upper bound even if greatness *per se* does not. Human beings may well be limited in how great they can conceive something to be. But this would imply that something possible might be greater than $\iota x\phi_I$, and the ontological argument cannot allow this (cf. Millican 2004; 2007). Accordingly, the discussion above assumes that $\iota x\phi_I$, if there is such a thing, is also the greatest being possible.

[10] If $\iota x\phi_I$ exists in the mind, then there is a very weak sense in which it is "possible." But that is a nonstandard sense, for it would make the largest ordinal "possible" as well. I thus ignore this sense of the word.

[11] This point most likely assumes axiom S4. Still, if the ontological arguer replies by rejecting S4, her position becomes significantly more contentious than has been hitherto recognized.

despite being definitionally true. It remains that $\iota x\phi_I$ might "exist in reality" only *per impossibile*. (And that is a way of *not* existing in reality.)

Still, this could strike one as idle skepticism. That $\iota x\phi_I$ can (non-impossibly, non-impossibly, non-impossibly…) exist in reality may seem obvious. But conceivability is no guide to possibility, or so the slogan goes. Besides, the possibility of $\iota x\phi_I$ is regularly contested in modal ontological arguments (cf. Leibniz, Tooley 1981), where its possibility is an explicit premise. The premise is suppressed in non-modal arguments—yet I want to contest it here too, or at least, highlight the lack of argument.[12] Though questioning an upper bound to "greatness" seems like a novel way to challenge modal arguments as well.

3. An Explanation

Yet there may remain something puzzling. We saw that O&Z's free logic is "existentially timid" twice over: Some terms denote objects just in the mind; others fail to denote at all. This made the logic admirably vigilant in avoiding the existential fallacy. However, the objection is that the fallacy occurs regardless—for nothing secures that '$\iota x\phi_I$' denotes. Yet if the logic is timid in the ways indicated, how could it allow the fallacy to occur?

The answer lies in O&Z's "Description Axiom," or more precisely its corollary (DT3), repeated here for convenience:

(DT3) $\psi[\iota x\phi/z] \supset \exists y(y = \iota x\phi)$, *where* ψ *is an atomic or identity formula*.

---

[12] Unlike PROVER9, and to their credit, Baker & Matthews (2010) consider the impossibility of $\iota x\phi_I$ in their non-modal ontological argument. But they simply declare that no reason has been given to doubt its possibility. Yet this misplaces the burden of proof. Besides, the above remarks can be seen as substantiating the doubt. On reflection, I have little idea whether there is an upper bound to "greatness;" as it stands, $\iota x\phi_I$ may indeed be impossible.

The fact is that (DT3) licenses inferences of this fallacious type. Consider in this connection the predicate ‘Cx’. In the argument for (P), ‘$C\iota x\phi_1$’ is said to be true, since ‘$\iota x\phi_1$’ translates the non-gibberish descriptor ‘the greatest conceivable being’. But as ‘the largest ordinal’ shows, some non-gibberish descriptors are impossible to satisfy, *yet they still occur in true atomic sentences*. The following is example, where ‘$\iota x\omega_1$’ translates ‘the largest ordinal’:

(*) $C\iota x\omega_1$

Given the truth of (*), the problem is that (DT3) permits us to derive ‘$\exists y(y = \iota x\omega_1)$’. Yet since it ends up being *impossible* to satisfy ‘$\iota x\omega_1$’, the inference here is fallacious.[13,14]

In their defense, O&Z may suggest that the inference to ‘$\exists y(y = \iota x\omega_1)$’ is valid, on the supposition that the descriptor denotes an impossible object (one that exists only in the mind). This impossibilist view may comport with Zalta’s (1983; 1988) well-known Meinongianism. But ultimately, this only re-locates the fallacy. Consider here the following descriptor:

(D2) ‘$\iota x\omega_2$’ denotes y iff Ey and y is the largest ordinal.

Note that like ‘$\iota x\omega_1$’, this descriptor is also non-gibberish, meaning it is true that:

(**) $C\iota x\omega_2$

---

[13] It might be argued that ‘is conceivable’ should not be symbolized as atomic, given the modality expressed by the suffix ‘-able’. But we could make the same point using instead the predicate ‘Dx’, meaning “x is conceived.” Indeed, the ontological arguer is committed to ‘Dx’ having satisfiers; that is her point in observing that a descriptor is “non-gibberish.” (Her point is that, in understanding the descriptor, the object is *thereby* conceived.)

[14] The fallacy here might not be an *existential* fallacy, since ‘$C\iota x\omega_1$’ was never said to be a *definitional* truth. But if (DT3) permits this fallacy, then it permits the existential fallacy. After all, the inference licensed by (DT3) remains a non-sequitur even if we suppose in addition that ‘$C\iota x\omega_1$’ is true by definition. This is how (DT3) can explain why the existential fallacy occurs in O&Z’s free logic.

But from (**), (DT3) allows us to derive '$\exists y(y = \iota x\omega_2)$'. And here, the existential statement does not reflect just that some impossibilium exists in the mind. For in light of (D2), it follows that:

(*X*) $E\iota x\omega_2$

That is, by (D2), the largest ordinal exists in reality. And even if this ordinal exists in the mind, there is certainly no such ordinal *in reality*. (The argument here is basically a Gaunilo-type maneuver, which also exploits that impossibilia can "exist in reality" only *per impossibile*.) Yet since (D2) is definitional, it then follows that '$\exists y(y = \iota x\omega_2)$' is false. So given the truth of (**), (DT3) licenses an existential fallacy.[15]

Before closing, let me acknowledge that O&Z entertain the possibility of an atomic formula falsifying (DA) or (DT3).[16] But they purportedly rule it out via a *metaphysical* thesis, the correspondence theory of truth:

> The correspondence theory of truth, which presumes truth to be a kind of correspondence between language and the world, places the following constraint on the truth of atomic and identity formulas ϕ: ϕ is true if and only if every term in ϕ has a denotation and the denotations stand in the correct relationships. Otherwise, how could an atomic or identity

---

[15] Another objection: If O&Z restrict the domain to possible objects, then (*) and (**) are each false in O&Z's logic. In which case, they are not truths that show (DT3) fails to be truth-preserving. Reply: If (*) or (**) are evaluated as false, that just shows the logic to be unsuitable for the ontological arguer. After all, the descriptors in (*) and (**) are not gibberish (at least, no more so than '$\iota x\phi_1$'), and the ontological arguer has presumed that non-gibberish description suffices for conceivability.

[16] Since (DT3) is false, it may be best to say that the ontological argument is unsound rather than invalid. For (DT3) functions like a false premise in the argument. If you prefer to describe things that way, be my guest. My own tendency is to think of the argument as invalid, given that the logic is partly defined by (DA)'s axiomatic status, and (DT3) originates in (DA). But ultimately, nothing dialectically important hangs on the issue.

formula having a non-denoting term be true in virtue of some feature of the world? (n. 9 of O&Z 1991)

In light of (*) or (**), however, this seems like cold comfort. Who knows *how* a thing like (*) is true, but *that* it is true seems hard to deny, at least if one thinks that '$C\iota x\phi_1$' is true.

4. Closing

Broadly speaking, the lesson is that a so-called "non-modal" ontological argument must be modal in part. The possibility of $\iota x\phi_1$ must be defended, to avoid "proving" the impossible. Despite its exciting potential, then, PROVER9 leaves matters oversimplified. It is accordingly hasty to conclude that "The question of the soundness of the [non-modal] ontological argument…reduces to the question of the truth of [(P)]" (O&Z 2011, p. 345). One can understand the temptation: When an automated system shows that (P) is the only non-logical premise required, that suggests the whole debate reduces to (P). But unexpectedly, the case for theism can be made *stronger* by adding further non-logical premises. That is most clearly the case when the theist needs to fend off the charge of question-begging. So, although PROVER9 clearly has something to offer, it is a different sort of fallacy to think that a valid argument from fewer premises must be superior.

When filling out the argument, moreover, we found that the controversy hangs on at least four more questions:

a. Is (P*) true?
b. Does (P*) entail (P)?
c. Is $\iota x\phi_1$ possible?
d. Is axiom (DA) true?

Not coincidentally, this is where we found a more compelling objection. The fact is that the answer to (c) is unknown. And if the answer is 'no', then nothing is $\iota x \phi_1$. That is so, even if the answer to both (a) and (b) is 'yes'. Again, if an object is impossible for an independent reason, it just adds to its impossibility to define it as "existing in reality." Nonetheless, O&Z's logic licenses the existential fallacy at this point, despite its safeguards against it. Briefly, that is because the answer to (d) is 'no'.

[2793 words]

References

Baker, L. (2013). Updating Anselm Again. *Res Philosophica*, 90(1), 23–32.

Baker, L. & G. Matthews. (2010). Anselm's Argument Reconsidered. *The Review of Metaphysics*, 64(1), 31–54.

Broad, C.D. (1953). *Religion, Philosophy and Psychical Research.* London: Routledge.

Garbacz, P. (2012). Prover9's Simplification Explained Away. *Australasian Journal of Philosophy*, 90(3), 585–592.

Lewis, D. (1970). Anselm and Actuality. *Noûs*, 4(2), 175–188.

Malcolm, N. (1960). Anselm's Ontological Arguments. *Philosophical Review*, 69, 41–62.

Millican, P. (2004). One Fatal Flaw in Anselm's Argument. *Mind*, 114, 437–476.

____. (2007). Ontological Arguments and the Superiority of Existence: Reply to Nagasawa. *Mind*, 116, 1041–1053.

Oppenheimer, P. & E. Zalta. (1991). On the Logic of the Ontological Argument. *Philosophical Perspectives*, 5, 509–529.

____. (2007). Reflections on the Logic of the Ontological Argument. *Studia Neoaristotelica*, 4(1), 28–35.

____. (2011). A Computationally-Discovered Simplification of the Ontological Argument. *Australasian Journal of Philosophy*, 89(2), 333–349.

Plantinga, A. (1974). *The Nature of Necessity*. Oxford: Clarendon Press.

Salmon, N. (1987). Existence. *Philosophical Perspectives*, 1, 49–108.

Tooley, M. (1981). Plantinga's Defence of the Ontological Argument. *Mind*, 90(359), 422–427.

Zalta, E. (1983). *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Dordrecht: Kluwer Academic Publishers.

____. (1988). *Intensional Logic and the Metaphysics of Intentionality*. Cambridge, MA: MIT Press.