

Sobel on Gödel's Ontological Proof

Robert C. Koons

Department of Philosophy

University of Texas at Austin

koons@mail.utexas.edu

February 14, 2005

Kurt Gödel left with his student Dana Scott two pages of notes in which he sketched a new version of Anselm's ontological proof of God's existence. In his most recent book, Howard Sobel spends the greater part of a chapter dedicating his considerable talents to an elucidation and critique of Gödel's argument, as well as to an emended version of that argument proposed by Anthony Anderson.

The ontological argument has garnered quite a bit of attention in the last fifty years. In most cases, philosophers have agreed that the argument is unsuccessful but have disagreed vigorously over where exactly the fatal flaw lies. This paper, will to some extent, follow the familiar pattern. I will argue that Gödel's argument is unsuccessful, but I hope to show that Sobel and Anderson have both misdiagnosed its failure, and, consequently, Anderson's attempted

repairs are likewise unsuccessful. However, I will close with a sketch of my own proposed repair of Gödel's argument, and I will suggest that, although the repaired argument is not by itself a successful theistic proof, it may represent a fruitful matter for future investigation.

Technically speaking, Gödel's argument requires second-order quantified modal logic, with a single third-order predicate of properties, P , intended to signify that a property is "positive". Gödel uses a standard modal logic, including axioms 5 and T (or at least B). Gödel's notion of a "positive" property seems to have two distinct bases: an axiological understanding of positivity, and a purely logical understanding. The axiological understanding is something like this: a property F is positive iff having F is compatible with being perfect (in a moral and aesthetic sense). The logical notion is something like this: when the logical form of the property is correctly analyzed (using a logically perfect language in Russell's sense, a language whose lexical primitives correspond perfectly to ontological primitives) the prenex, disjunctive normal form of the correct formulation of the property has at least one negation-free disjunct.

These two conceptions are, of course, not entirely unrelated. The Neoplatonic and Augustinian tradition in Western philosophy and theology has long maintained a "privation theory of evil": that every defect in a thing, whether moral, aesthetic or whatever, consists in the thing's lacking some positive quality. As it is often put, being and goodness are convertible. To be good is simply to be: to fail to be good is (in some relevant respect) to fail to be. On this

conception of evil (that is, of imperfection), the two conceptions of positivity coincide perfectly: a property is incompatible with perfection just in case it entails that its bearer is lacking in some positive quality, and this entailment occurs just in case the correct formulation of the property contains negations in every disjunct.

In my view, this privative theory of evil is a reasonably plausible one, so I will not fault Gödel's argument for presupposing it.

Gödel's proof depends on five axioms and three definitions:

A1. $P(\neg F) \leftrightarrow \neg P(F)$

A2. $(P(F) \& \Box(F \rightarrow G)) \rightarrow P(G)$

Axiom A1 tells us that a property is positive iff its negation is negative. This makes sense on both the logical and the axiological understandings. If the correct formulation of F contains a negation-free disjunct, then every disjunct in the formulation of $\neg F$ contains a negation, and vice versa. On the axiological understanding, it is clear that if F is incompatible with perfection, then $\neg F$ must be compatible with it (this is the right to left direction of A1). As Anderson pointed out, the left-to-right direction of A1 isn't so obviously true on the axiological interpretation: couldn't both F and $\neg F$ be compatible with perfection? However, if goodness and being are truly convertible, then, since at least one of F or $\neg F$ must entail a degree of negativity or privation, they can't both be compatible with absolute perfection.

Axiom A2 indicates that, if F is positive, then any property that F necessitates must also be positive. This clearly makes sense under both interpretations of positivity. Axioms A1 and A2 has an important corollary:

Th. 1. $P(F) \rightarrow \Diamond \exists x Fx$

If a property F is positive, it must be possibly instantiated, since a property that is not possibly instantiated necessitates every property (vacuously), and then, by Axiom A2, it would follow that every property is positive, which is clearly ruled out by A1.

Gödel defines Godlikeness as the possession of all positive properties:

Def. G: $G(x) \leftrightarrow \forall F(P(F) \rightarrow F(x))$

A3. $P(G)$

Axiom A3 asserts that G is positive. This makes good sense under both interpretations of positivity: if no positive property entails any negativity or privation, then G (which is, in effect, the infinitary conjunction or intersection of all the positive properties) must also be potentially negation-free. Similarly, to possess all the properties compatible with perfection is surely itself compatible with perfection. From A3 and Th. 1, it follows that G is possibly instantiated.

A4. $P(F) \rightarrow \Box P(F)$

Axiom A4 simply states that being positive is an essential attribute of every positive property, an unexceptionable claim.

Def. Ess. $F\text{Ess } x \leftrightarrow F(x) \& \forall G[G(x) \rightarrow \Box(F \rightarrow G)]$

A property is an “essence” of a thing in Gödel’s idiosyncratic sense just in case the thing has the property, and the property necessitates all of the thing’s other actual properties. An essence is something like a total individual concept in Leibniz’s sense: the property of a thing that encompasses all of its actual properties. If we assume that two properties are identical if each entails the other, then it is easy to show that each thing has at most (and presumably exactly) one essence.

It is easy to prove that Godlikeness is an essence (in this sense) of anything that has it:

Th. 2. $G(x) \rightarrow G \text{Ess } x$

Next, Gödel introduces a definition of “necessary existence”. Again, Gödel uses this phrase in a non-standard sense. What he calls “necessary existence” is really something like a contingency-free existence: having only those “essential” or total properties that are necessarily instantiated by something or other.

Def. NE. $NE(x) \leftrightarrow \forall F[F \text{Ess } x \rightarrow \Box \exists x F(x)]$

Gödel’s NE is much stronger than necessary existence, as it is ordinarily understood. An object x might exist even though its essence (its total or individual concept) might not have been instantiated: this will happen whenever the necessarily existing thing has even one contingent property. (If being identical to

x is a property of x , then Gödel's NE property does entail necessary existence. If, however, we don't count such things a properties strictly speaking, then it would be possible for a contingent being to have Gödel's NE property, so long as, in every possible world, something exactly like it exists.)

Finally, Gödel assumes that "necessary existence" in this sense is a positive property, from which it follows that Godlikeness is necessarily instantiated, and, if we assume axiom T of modal logic, that Godlikeness is actually instantiated.

A5. $P(NE)$

Th. 3. $\Box\exists xGx$

A Godlike this has every positive property, including "necessary existence". From Theorem 2, we know that G is an essence of any Godlike thing, so, by the definition of necessary existence, it follows that if anything is a Godlike thing, Godlikeness is necessarily instantiated. We know that it is at least possible that there be a Godlike thing (since Godlikeness is positive, and, by Theorem 1, any positive property is possibly instantiated). So, it is possible that Godlikeness is necessarily instantiated. By axiom 5 of standard modal logic, it follows that Godlikeness is necessarily instantiated.

If being identical to x counts as a property of any Godlike thing x , we can prove that there is exactly one Godlike thing, since being identical to Godlike thing x must be a positive property (since, otherwise, not being identical to Godlike thing x would be positive, and thing x would, being Godlike, have to

have it). But this means that every Godlike thing must be identical to thing x , so there can exist only one Godlike thing. Thus, we conclude that God (i.e., the unique godlike thing) exists.

The argument I just gave can be extended (as Sobel proves) to show that God can have only positive properties:

Th. 4. (Sobel) $G(x) \rightarrow \forall F[F(x) \rightarrow P(F)]$

Sobel's principal objection to Gödel's argument is that it engenders "modal collapse": we can use Gödel's axioms to prove that every actual truth is necessarily true – that there is absolutely nothing is contingently true, a disastrous result.

Here is Sobel's proof of modal collapse (p. 157): first, we prove that all of God's properties are necessarily instantiated. Suppose that a Godlike being exists and has property F . Call the Godlike being j . We know, from theorem 2, that G is the essence of j . This means that G necessitates all of j 's actual properties. Since j has F , G must necessitate F , and since G is necessarily instantiated, F must also be necessarily instantiated. In fact, the conjunction of F and being identical to j is necessarily instantiated: so j has F in every possible world.

For the proof of modal collapse, let Q be some arbitrary truth. We will show that $\Box Q$. We know, from Gödel's theorem 3, that a Godlike being exists: call it j again. So, we know $G(j)$. We also know, from theorem 2, that G is the essence of j . This means that G necessitates all of j 's actual properties. Since Q is true,

j has the property of being such that Q (i.e., from $(Q \& j = j)$, we can deduce that j has the property $\hat{x}[Q \& x = x]$). Thus, being G must necessitate being such that Q . Since G is instantiated in every world, it follows that something is such that Q is true in every world. Hence, $\Box Q$.

Of course, the crucial question here is: what are the properties in the domain of Gödel's second-order quantifiers? Sobel assumes that properties are nothing more than functions from possible worlds to sets of things, an extremely liberal conception. On such a conception, the property of being such that Q is unproblematic, since it corresponds to a function from worlds to sets of individuals of the following kind: if Q is true in world W , then $f(W)$ is the set of all individuals in W ; if Q is false in W , then $f(W)$ is the empty set. Sobel's liberal interpretation of properties corresponds to his acceptance of an abstraction schema for properties: if μ is an open formula, with free variable x , then there exists a property $\hat{x}[\mu]$.

On the one hand, Gödel's proof does not require anything so powerful as a generic abstraction schema. In fact, nothing in the proof seems to depend any instance of the schema. On the other hand, Gödel does in fact assume that there is a property of being self-identical and a property of being non-self-identical. This suggests that he might have accepted something like Sobel's conception of properties.

If so, Sobel's proof seems to show that such a liberal conception of properties forces Gödel's system into a disastrous collapse. Before rejecting or emending

one or more of Gödel's axioms, the most conservative response is to restrict the domain of properties. This could be done in a number of ways. We might insist that Gödel's property-variables stand only for a thing's intrinsic properties. The class of intrinsic properties is the class of properties that are qualitative and non-relational: that pertain or fail to pertain to a thing because of its internal constitution. There is nothing in Gödel's argument that rules out this interpretation of his second-order variables. To make his proof work under this interpretation, we need only the following properties of the set of intrinsic properties:

IN1. If F is intrinsic, so is $\neg F$.

IN2. The conjunction of a set of intrinsic properties is itself intrinsic.

IN3. Everything has at least one intrinsic property in every world (satisfied if the property of being self-identical counts as intrinsic), and an impossible property (such as being non-self-identical) counts as intrinsic.

These are quite plausible assumptions. Furthermore, Gödel's axioms make perfect sense under this new interpretation. We can apply both the logical and the axiological interpretation of positivity to the case of intrinsic properties. Being Godlike is intrinsic, as well as positive, since it consists in an infinite intersection of intrinsic properties. Finally, necessary existence is an intrinsic property, since it consists simply in not having certain intrinsic properties (namely, those that are not necessarily instantiated).

Under this interpretation, Sobel's modal collapse proof does not go through, since being such that snow is white no longer counts as a property (under the intended interpretation) We still have the conclusion that God has all of His intrinsic properties necessarily, but this conclusion would not be unwelcome to theists of an Anselmian or Neo-Platonic bent. Classical theists like Thomists describe God as a being of "pure act", a being whose intrinsic character is utterly free of contingency, and hence absolutely changeless. This of course raises the question of how such a God could know about or care about contingent matters of fact (such as the plight of the victims of hurricane Ivan or the South Asian tsunami). The standard scholastic answer to this question consists in the claim that God's knowledge about and concern for His creatures requires no internal modification of His being. His love for us simply consists in the loving actions that flow inevitably from God's being to us, and there is no real distinction between God's knowledge of a contingent fact and that fact itself. These are, admittedly, counterintuitive, even paradoxical claims, but to object to the ontological argument on the grounds that it supports the standard, scholastic version of theism, as opposed to a more commonsensical version of it, seems seriously misplaced.

I should also mention, however, that there is one corollary of Gödel's argument that cannot be sustained under the interpretation that limits properties to intrinsic properties. We can no longer prove that there can be only one God-like being. If j is a Godlike being, then being identical to j (and, equivalently,

distinct from everything other than j) cannot plausibly be thought an intrinsic property of j . However, there are other arguments that can be used to rule out, on plausible grounds, the existence of two or more godlike beings. For example, the existence of two omnipotent beings is logically impossible. In addition, trinitarian Christians might find it advantageous to abandon too rigorous a proof of the absolute unicity of God.

If one finds the scholastic model of an impassible and immutable God unattractive, there is at least one more plausible interpretation of Gödel's property variables. We can take a 'property' to be something like belonging to or not belonging to a natural kind, or, in Aristotelian terms, a genus or differentia in the category of substance. Let's call F a $\{i\}_i$ natural-kind property/ $i\}_i$ just in case F can be defined as a logical complex, built up from genera and species. Again, the facts we need are readily available:

NK1. If F is a natural-kind property, so is $\neg F$.

NK2. The conjunction of a set of natural-kind properties is itself a natural-kind property.

NK3. Everything belongs to at least one natural kind in every world (satisfied if the property of being self-identical counts as a natural kind), and an impossible property (such as being non-self-identical) counts as a (vacuous) natural-kind property.

Consequently, being Godlike surely qualifies as a natural-kind property, since

it is the conjunction of a set of natural kinds. Similarly, necessary existence consists in not belonging to any natural kind that is possibly uninstantiated. Given N1 and N2, this is itself surely a natural kind.

On this interpretation, Sobel's modal collapse argument clearly fails. Being such that snow is white is certainly not a natural kind. There is nothing especially shocking about the conclusion that God belongs to whatever natural kinds He does as a matter of necessity.

Thus, Sobel seems to have erred in finding fault with Gödel's arguments on these grounds, and Anderson's emendations, designed to avoid the collapse by substantially modifying Gödel's axioms and definitions, were entirely unnecessary. Nonetheless, I believe that there is a fatal flaw in Gödel's argument, one that both Sobel and Anderson overlooked. The flaw concerns axiom A5, the positivity of necessary existence. Sobel thinks that A5 is plausible under the logical interpretation of positivity: "there seems to be 'no privation' about it." (p. 125). This was an injudicious concession on Sobel's part.

Axiom A5 states simply that "necessary existence", in Gödel's sense, is a positive property. Gödel's necessary existence is provably equivalent to the condition below, the condition of being "contingency free" (or CF).

Def. CF $CF(x) \leftrightarrow \forall F[F(x) \rightarrow \Box \exists x F(x)]$

Equivalently:

$$CF(x) \leftrightarrow \forall F[\Diamond \neg \exists x F(x) \rightarrow \neg Fx]$$

It is easy to prove that CF and NE are necessarily coextensive (on the assumption, which Sobel rightly endorses, that everything necessarily has at least one essence). So, NE is positive if and only if CF is. CF is the property of having only necessarily instantiated properties. This entails not having any property that is possibly uninstantiated. CF is the equivalent to the infinite conjunction of the members of a set of properties – namely, the set of complements of those properties that are not necessarily instantiated (i.e., that are possibly uninstantiated). CF is positive iff none of the properties that are possibly uninstantiated are themselves positive. If, instead, there is a positive property F that is not necessarily instantiated, then CF entails not having F, which would make CF a negative property (any property that entails not having some positive property must itself be negative).

Thus, whether CF (and NE) are positive depends on whether it is true that all positive properties are necessarily instantiated. If some positive property is possibly uninstantiated, then CF and NE are clearly themselves negative. Thus, we have no reason to accept Axiom 5, unless we already believe that all the positive properties (including of course *G*) are necessarily instantiated. We have no reason to accept Axiom 5 unless we know that God exists necessarily.

Why were Gödel (as well as Sobel and Anderson) taken in by Axiom 5? I think the error lies in confusing a positive property with a property picked out by a positive second-order condition. The condition by which CF is defined is purely positive: in order to belong to the set of which CF is the conjunction, a

property must satisfy only the purely positive condition of being a property that is necessarily instantiated. However, this is certainly not sufficient to make CF itself a purely positive property. Consider the property "self-identity completeness". This property consists in having every property that is self-identical:

Def. SIC $SIC(x) \leftrightarrow \forall F[F = F \rightarrow F(x)]$

The second-order condition by which we define SIC is paradigmatically positive: the property of being self-identical. Yet, SIC itself is certainly negative, since it is logically impossible to have all properties (including, for every F, having both F and not-F). In fact, SIC is paradigmatically negative, since it is equivalent to the first-order property of being non-self-identical, $\hat{x}[x \neq x]$. I think that it's plausible to think that it was just such a confusion between being a positive property and being a property defined by a purely positive condition that led Gödel into the error of proposing Axiom 5 as part of his proof.

So, we don't need to worry about global collapse, and there's nothing seriously wrong with Axioms 1-4. However, without Axiom 5, Gödel's ontological proof is unsuccessful. There is, however, a simple repair that might do the job: replace Axiom 5 with Axiom 6:

A6. $P(F) \rightarrow P(\Box F)$

If a property F is positive, then so is the property of being F in every possible world. Since Godlikeness is positive, it follows that being Godlike in every possible world is also positive. Positive properties are always possibly

instantiated, so being necessarily Godlike is possibly instantiated. In S5, it follows that Godlikeness is necessarily, and thus also actually, instantiated.

Does A6 suffer from exactly the same flaw as A5? Unfortunately, the answer is Yes. If there is a positive property F that is possibly uninstantiated, then A6 will fail in that case, since in that case $\Box F$ or, more precisely, $\hat{x}\Box F(x)$ (being F in every possible world), will be an impossible property, and so negative rather than positive. Thus, A6 presupposes that every positive property (including the conjunction of all of them) is instantiated of necessity, but this is just what the ontological argument was supposed to establish.