An Object-Logic Explanation for the Inconsistency in Gödel's Ontological Theory

Christoph Benzmüller[1]* and Bruno Woltzenlogel Paleo[2]

[1] Freie Universität Berlin, Berlin, Germany
c.benzmueller@fu-berlin.de
[2] Australian National University, Canberra, Australia
bruno.wp@gmail.com

Abstract. This paper discusses the inconsistency in Gödel's ontological argument. Despite the popularity of Gödel's argument, this inconsistency remained unnoticed until 2013, when it was detected automatically by the higher-order theorem prover Leo-II [?]. Complementing the meta-logic explanation for the inconsistency available in our IJCAI 2016 paper [4], we present here a new purely object-logic explanation that does not rely on semantic argumentation.

1 Introduction

Kurt Gödel's ontological argument for the existence of God [7, 12] is amongst the most discussed formal proofs in modern literature. A rich body of publications – including very recent ones – present, discuss, assess, criticize, modify and improve Gödel's original work (see e.g. Sobel [13] and Oppy [10] and the references therein).

Scott's version of Gödel's argument was automatically reconstructed by the higher-order theorem prover Leo-II [3] and its correctness was verified step-by-step in the Coq proof assistant [?]. To bridge the gap between higher-order logics (HOL; cf. [1] and the references therein), as used by these systems, and higher-order *modal* logics (HOML; cf. [8] and the references therein), on which the ontological argument argument, the logic embedding approach [2, 3] was used.

However, Gödel's original axioms, as used in his manuscript [7], are inconsistent. This fact has remained unnoticed to philosophers until 2013, when Leo-II found a surprising refutation of the axioms.

In [4], we have extracted from Leo-II's machine-oriented refutation an informal and human-oriented intuitive explanation for the inconsistency, and we have reconstructed and verified it in the Isabelle proof assistant. However, that explanation relied on reasoning at the meta-logic (HOL) level, which was only possible because of the embedding. Here we complement that work with a purely object-logic (HOML) explanation.

Applications of (first-order) theorem proving technology in metaphysics were first reported by Fitelson, Oppenheimer and Zalta [6, 9]. Later on, Rushby [11] used the PVS proof assistant. Common to both works is a significant amount of proof-hand-coding work as well as their focus on a non-modal formalization of St. Anselm's simpler and older ontological argument.

* This work has been supported by the German Research Foundation DFG under grants BE2501/9-1,2

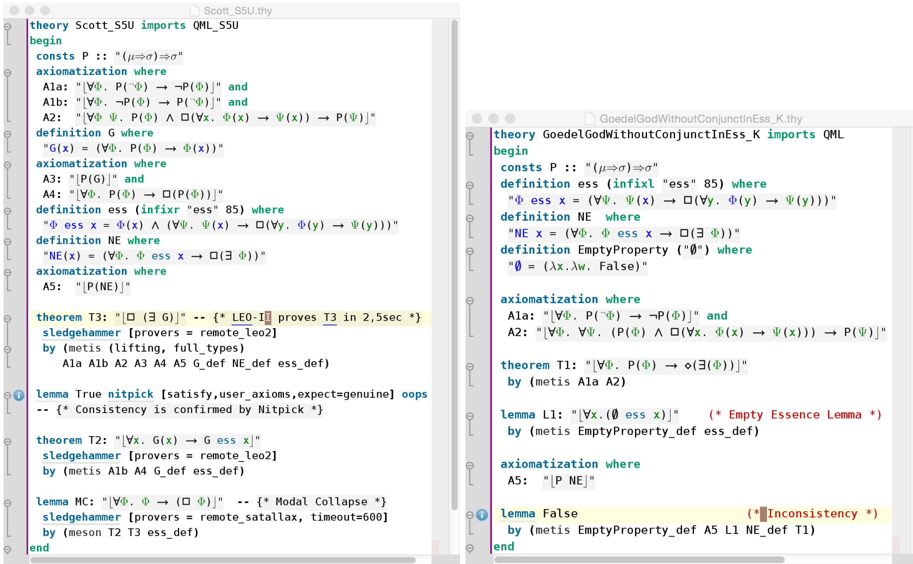


Fig. 1: Scott's consistent axioms (left) and proof of the inconsistency of (a subset of) Gödel's original axioms (right)

2 An Essential Difference in the Definitions of Essence.

Gödel's manuscript can be considered a translation of Leibniz's ideas on the argument into modern modal logic. Gödel discussed his manuscript with Scott, who shared a slightly different version with a larger public. Scott's version of the axioms and definitions, formalized in Isabelle, is shown in Fig. 1. The main difference to Gödel's version is an extra conjunct in the definition of *essence* (*ess*). For Scott, an essential property of an individual must be possessed by him/her. For Gödel, this is not required.

Gödel's omission has been considered inessential and merely an oversight by many. For more than four decades, its serious consequences remained unnoticed, despite numerous analyses and criticisms of the argument. However, as explained here, the extra conjunct is in fact crucial. Without it, Gödel's original axioms are inconsistent. With it, Scott's axioms are consistent (cf. Fig. 1 where the model finder Nitpick [5] confirms consistency). In personal communication, Dana Scott confirmed that he was unaware that Gödel's original axioms were inconsistent.

3 Automating HOML in HOL

In our experiments in this branch of metaphysics we utilise an embedding of HOMLs, such as **K**, **KB** and **S5** with various domain conditions (possibilist and actualist quantification), in HOL. More precisely, formulas in HOML are *lifted*, i.e., converted into predicates over worlds, which are themselves explicitly represented as terms. The logical constants of HOML are translated to HOL terms in such a way that, for instance, $\Box\varphi$

and $\Diamond\varphi$ (relative to a current world w_o) are mapped, respectively, to the HOL formulas $\forall w.(rw_0w){\rightarrow}(\varphi w)$ and $\exists w.(rw_0w) \wedge (\varphi w)$. This form of embedding is precisely the well-known standard translation, which is here intra-logically realized — and extended for quantifiers — in HOL by stating a set of equations defining the logical constants. The resulting object logic is the HOML **K** with rigid terms and constant domains (possibilist quantifiers). Other logics (e.g. **KB**, **S5**) are embedded by adding axioms that restrict the accessibility relation r. Varying domains and actualist quantifiers can be simulated by using an existence predicate to guard the quantifiers.

4 Intuitive Explanations for the Inconsistency

In the typical workflow during an attempt to prove a conjecture with a theorem prover, it is customary to check the consistency of the axioms first. For if the axioms are inconsistent, anything (including the conjecture) would be trivially derivable in classical logic (*ex falso quodlibet*). Surprisingly, when this routine check was performed on Gödel's axioms [3], the Leo-II prover claimed that the axioms were inconsistent. Unfortunately, the refutation generated by Leo-II was barely human-readable. The refutation was based on machine-oriented inference rules (the higher-order resolution calculus [14]), and the text file had 153 lines (with an average of 184 characters per line) and used a machine-oriented syntax (TPTP THF [15]).

Although Leo-II's resolution refutation is not easy to read for humans, it did contain relevant hints to the importance of the empty property[3] $\lambda x.\bot$ (also denoted $\emptyset$, as in HOL it is customary to think of unary predicates as sets)[4]. Based on this hint, we conceived the following informal explanation for the inconsistency of Gödel's axioms (reproduced without change from [4]):

1. From Gödel's definition of essence ($\phi\ ess\ x \leftrightarrow \forall\psi(\psi(x){\rightarrow}\Box\forall y(\phi(y){\rightarrow}\psi(y)))$) it follows that the empty property (or self-difference) is an essence of every individual (**Empty Essence Lemma**): $\forall x\ (\emptyset\ ess\ x)$
2. From theorem T1 (*Positive properties are possibly exemplified*: $\forall\phi[P(\phi){\rightarrow}\Diamond\exists x\phi(x)]$) and axiom A5 ("necessary existence" is a positive property: $P(NE)$), it follows that NE is possibly exemplified: $\Diamond\exists x[NE(x)]$
3. Expanding the definition of "necessary existence" ($NE(x) \equiv \forall\phi[\phi\ ess\ x{\rightarrow}\Box\exists y\phi(y)]$), the following is obtained: $\Diamond\exists x[\forall\varphi[\varphi\ ess\ x{\rightarrow}\Box\exists y[\varphi(y)]]]$
4. The sentence above holds for all φ and thus, in particular, for the empty property (or self-difference): $\Diamond\exists x[\emptyset\ ess\ x{\rightarrow}\Box\exists y[\emptyset(y)]]$
5. By the Empty Essence Lemma, the antecedent of the implication above is valid. Therefore, the sentence above entails: $\Diamond\exists x[\Box\exists y[\emptyset(y)]]$
6. By definition of $\emptyset$: $\Diamond\exists x[\Box\bot]$
7. As the existential quantifier is binding no variable within its scope, the sentence is equi-valid with: $\Diamond\Box\bot$

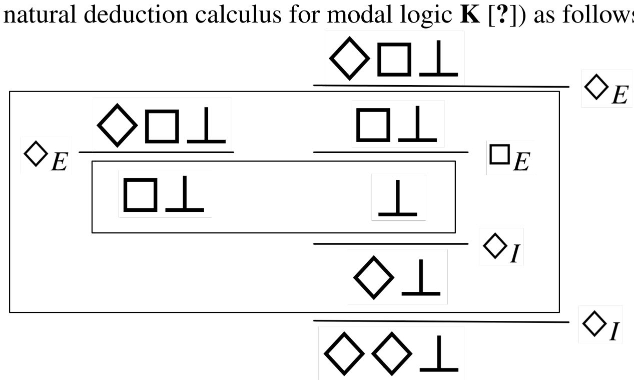
[3] Note that the terms for the empty property ($\lambda x.\bot$) and for the property of self-difference ($\lambda x.x \neq x$) have identical denotations in the logic setting with functional and Boolean extensionality assumed here. For the proof to go through it is irrelevant which property is used.

[4] An additional lambda abstraction occurs in the empty property in Leo-II's proof (and also in the reconstruction in Isabelle) because the embedding approach lifts the boolean type o to $\iota{\rightarrow}o$.

8. To see that the sentence above is contradictory, we may reason semantically, thinking of possible worlds. If w_0 is the arbitrary current world, the $\Diamond$ operator forces the existence of a world w accessible from w_0 such that $\Box\bot$ is true in w. But $\Box\bot$ can only be true in w, if there is no world w' accessible from w. In logics[5] with a reflexive or symmetric accessibility relation (e.g. **KB**), it is easy to see that there must be a world w' accessible from w: either w' itself, in case of a reflexive relation, or w_0, in case of a symmetric relation. In fact, even in **K**, with no accessibility condition, there must be a world w' accessible from w. The reason is that $\Diamond\Box\bot$ should be *valid* (true in all worlds). Therefore, it is true in w as well, where the existence of an accessible world w' is forced by the $\Diamond$ operator. As a model for $\Diamond\Box\bot$ (which is a consequence of Gödel's axioms) cannot be built, Gödel's axioms are inconsistent.

If we were to convert the informal proof above to a formal proof, the semantic reasoning in step 8 would require a leap to the meta-logic (HOL), in order to expand the definitions of the modal operators and reason directly about possible worlds. The following alternative derivation avoids this leap and remains purely within the object logic (HOML **K**):

8*. We must derive $\bot$ from $\Diamond\Box\bot$. In order to derive $\bot$, it suffices to show that there exists a derivable proposition such that its negation is also derivable. We choose $\Diamond\Diamond\bot$ as a candidate proposition, and hence we must show that:
   - $\neg\Diamond\Diamond\bot$ *is derivable:* this proposition is logically equivalent to $\Box\Box\top$, which is trivially derivable from $\top$ by two applications of the necessitation inference rule.
   - $\Diamond\Diamond\bot$ *is derivable:* and indeed, it can be derived (using a recently developed natural deduction calculus for modal logic **K** [**?**]) as follows:



An interesting, unusual and noticeable feature of the derivation shown above is that the leftmost $\Diamond_E$ (diamond elimination) inference derives a formula ($\Box\bot$) that is never used as a premise. This is necessary because of the *eigen-box* condition, which requires

[5] Interestingly, the refutation automatically generated by Leo-II uses a symmetric accessibility relation, and thus requires the modal logic **KB**. The informal, human-constructed refutations described here, on the other hand, requires only the weaker modal logic **K**. In our experiments Leo-II (like all other HOL provers) was still too weak to automatically prove the inconsistency already in logic **K**. Hence, this remains an open problem for automated theorem provers.

that every box must be accessed by exactly one *strong* modal rule. The purpose of the *strong* $\Diamond_E$ inference is merely to create and access the innermost box that is needed by the *weak* $\Box_E$ and $\Diamond_I$ inferences inside the outermost box.

```
Theorem dia_box_false_to_false_meta: [(dia (box mFalse))] -> [mFalse].
Proof.
intro H. intro w.
destruct (H w) as [w0 [R0 H0]]. destruct (H w0) as [w1 [R1 H1]].
box_elim H0 w1 HF.
unfold mFalse in HF. destruct HF as [p [HF1 HF2]].
contradiction.
Qed.

Lemma mimplies_to_mnot: [mforall p:o, (p m-> mFalse) m-> (m~ p)].
Proof. mv.
intro p. intro H. intro H0.
destruct (H H0) as [p0 [H1 H2]].
apply H2. exact H1.
Qed.

Lemma dia_not_not_box: [ mforall p, (dia (m~ p)) m-> (m~ (box p)) ].
Proof. mv.
intro p. intro H1. intro H2. dia_e H1. apply H. box_e H2 H3. exact H3.
Qed.

Theorem dia_box_false_to_false_object: [(dia (box mFalse))] -> [mFalse].
Proof.
intro H. intro w.
exists (dia (dia mFalse)).
split.
  dia_e (H w). dia_e (H w0). dia_i w0. dia_i w1. box_e H0 H3. exact H3.

  apply box_not_not_dia. box_i. apply box_not_not_dia. box_i.
  apply mimplies_to_mnot. intro H4. exact H4.
Qed.
```

5 Conclusion

The axioms and definitions in Gödel's manuscript are inconsistent; this was detected automatically by the prover Leo-II. We presented a rational reconstruction and verification of the inconsistency argument in Isabelle/HOL. This argument is valid in all normal HOMLs including base logic **K**.

Extending the presentation in our IJCAI 2016 paper [4], we have added a syntactical proof for the inconsistency of $\Diamond\Box\bot$. Previously we had reduced the inconsistency analysis to exactly this proposition and we had then argued semantically that this proposition must be false already in modal logic **K**.

Our work reveals a challenge for automated reasoning: the (so far partially manual) extraction of an informal argument from a formal proof. Without accompanying human-understandable explanations, the proofs generated by provers such as Leo-II or Metis, will presumably be only of limited value for philosophers, for whom intuitive arguments remain crucial for the acceptance of novel results.

Our work demonstrates the potential of higher-order automated theorem proving and the semantic embeddings approach for philosophy: this technology is, in its current state of development, already capable of contributing novel results to metaphysics and to conduct reasoning steps at granularity-levels beyond common human capabilities.

References

1. P.B. Andrews. Church's type theory. In E.N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition, 2014.
2. C. Benzmüller and L.C. Paulson. Quantified multimodal logics in simple type theory. *Logica Universalis*, 7(1):7–20, 2013.
3. C. Benzmüller and B. Woltzenlogel-Paleo. Automating Gödel's ontological proof of God's existence with higher-order automated theorem provers. In Torsten Schaub, Gerhard Friedrich, and Barry O'Sullivan, editors, *ECAI 2014*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 93 – 98. IOS Press, 2014.
4. Christoph Benzmüller and Bruno Woltzenlogel Paleo. The inconsistency in Gödel's ontological argument: A success story for AI in metaphysics. In *IJCAI 2016*, 2016.
5. J.C. Blanchette and T. Nipkow. Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In *ITP 2010*, number 6172 in LNCS, pages 131–146. Springer, 2010.
6. Branden Fitelson and Edward N. Zalta. Steps toward a computational metaphysics. *J. Philosophical Logic*, 36(2):227–247, 2007.
7. K. Gödel. *Appx. A: Notes in Kurt Gödel's Hand*, pages 144–145. In Sobel [13], 1970.
8. R. Muskens. Higher Order Modal Logic. In P. Blackburn et al., editor, *Handbook of Modal Logic*, Studies in Logic and Practical Reasoning, pages 621–653. Elsevier, Dordrecht, 2006.
9. P.E. Oppenheimer and E.N. Zalta. A computationally-discovered simplification of the ontological argument. *Australasian J. of Philosophy*, 89(2):333–349, 2011.
10. G. Oppy. Ontological arguments. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2015 edition, 2015.
11. J. Rushby. The ontological argument in PVS. In *Proc. of CAV Workshop "Fun With Formal Methods"*, St. Petersburg, Russia, 2013.
12. D. Scott. *Appx. B: Notes in Dana Scott's Hand*, pages 145–146. In Sobel [13], 1972.
13. J.H. Sobel. *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge U. Press, 2004.
14. N. Sultana and C. Benzmüller. Understanding LEO-II's proofs. In Konstantin Korovin, Stephan Schulz, and Eugenia Ternovska, editors, *IWIL 2012*, volume 22 of *EPiC Series*, pages 33–52, Merida, Venezuela, 2013. EasyChair.
15. G. Sutcliffe and C. Benzmüller. Automated reasoning in higher-order logic using the TPTP THF infrastructure. *J. of Formalized Reasoning*, 3(1):1–27, 2010.