Gödel's God in Isabelle/HOL

Christoph Benzmüller and Bruno Woltzenlogel Paleo

September 18, 2020

A1	Either a property or its negation is positive, but not both:	$\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$
A2	A property necessarily implied by a positive property is positive:	$\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \to \psi(x)]) \to P(\psi)]$
T1	Positive properties are possibly exemplified:	$\forall\phi[P(\phi) \to \Diamond\exists x\phi(x)]$
D1	A *God-like* being possesses all positive properties:	$G(x) \leftrightarrow \forall\phi[P(\phi) \to \phi(x)]$
A3	The property of being God-like is positive:	$P(G)$
C	Possibly, God exists:	$\Diamond\exists x G(x)$
A4	Positive properties are necessarily positive:	$\forall\phi[P(\phi) \to \Box\, P(\phi)]$
D2	An *essence* of an individual is a property possessed by it and necessarily implying any of its properties:	$\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \to \Box\forall y(\phi(y) \to \psi(y)))$
T2	Being God-like is an essence of any God-like being:	$\forall x[G(x) \to G\ ess.\ x]$
D3	*Necessary existence* of an individual is the necessary exemplification of all its essences:	$NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \to \Box\exists y\phi(y)]$
A5	Necessary existence is a positive property:	$P(NE)$
T3	Necessarily, God exists:	$\Box\exists x G(x)$

Figure 1: Scott's version of Gödel's ontological argument [12].

1 Introduction

Dana Scott's version [12] (cf. Fig. 1) of Gödel's proof of God's existence [8] is formalized in quantified modal logic KB (QML KB) within the proof assistant Isabelle/HOL. QML KB is modeled as a fragment of classical higher-order logic (HOL); thus, the formalization is essentially a formalization in HOL. The employed embedding of QML KB in HOL is adapting the work of Benzmüller and Paulson [2, 1]. Note that the QML KB formalization employs quantification over individuals and quantification over sets of individuals (properties).

The gaps in Scott's proof have been automated with Sledgehammer [5], performing remote calls to the higher-order automated theorem prover LEO-II [3]. Sledgehammer suggests the Metis [9] calls, which result in proofs that are verified by Isabelle/HOL. For consistency checking, the model finder Nitpick [6] has been employed. The successfull calls to Sledgehammer are deliberately kept as comments in the file for demonstration purposes (normally, they are automatically eliminated by Isabelle/HOL).

Isabelle is described in the textbook by Nipkow, Paulson, and Wenzel [10] and in tutorials available at: http://isabelle.in.tum.de.

1.1 Related Work

The formalization presented here is related to the THF [14] and Coq [4] formalizations at https://github.com/FormalTheology/GoedelGod/tree/master/Formalizations/.
An older ontological argument by Anselm was formalized in PVS by John Rushby [11].

2 An Embedding of QML KB in HOL

The types i for possible worlds and μ for individuals are introduced.

typedecl i — the type for possible worlds
typedecl μ — the type for indiviuals

Possible worlds are connected by an accessibility relation r.

consts $r :: i \Rightarrow i \Rightarrow \mathit{bool}$ (**infixr** r 70) — accessibility relation r

QML formulas are translated as HOL terms of type $i \Rightarrow \mathit{bool}$. This type is abbreviated as σ.

type-synonym $\sigma = (i \Rightarrow \mathit{bool})$

The classical connectives $\neg, \wedge, \rightarrow$, and $\forall$ (over individuals and over sets of individuals) and $\exists$ (over individuals) are lifted to type σ. The lifted connectives are $m\neg$, $m\wedge$, $m\rightarrow$, $\forall$, and $\exists$ (the latter two are modeled as constant symbols). Other connectives can be introduced analogously. We exemplarily do this for $m\vee$, $m\equiv$, and $mL{=}$ (Leibniz equality on individuals). Moreover, the modal operators $\Box$ and $\Diamond$ are introduced. Definitions could be used instead of abbreviations.

abbreviation $\mathit{mnot} :: \sigma \Rightarrow \sigma$ $(m\neg)$ **where** $m\neg\ \varphi \equiv (\lambda w.\ \neg\ \varphi\ w)$
abbreviation $\mathit{mand} :: \sigma \Rightarrow \sigma \Rightarrow \sigma$ (**infixr** $m\wedge$ 65) **where** $\varphi\ m\wedge\ \psi \equiv (\lambda w.\ \varphi\ w \wedge \psi\ w)$
abbreviation $\mathit{mor} :: \sigma \Rightarrow \sigma \Rightarrow \sigma$ (**infixr** $m\vee$ 70) **where** $\varphi\ m\vee\ \psi \equiv (\lambda w.\ \varphi\ w \vee \psi\ w)$
abbreviation $\mathit{mimplies} :: \sigma \Rightarrow \sigma \Rightarrow \sigma$ (**infixr** $m\rightarrow$ 74) **where** $\varphi\ m\rightarrow\ \psi \equiv (\lambda w.\ \varphi\ w \longrightarrow \psi\ w)$
abbreviation $\mathit{mequiv} :: \sigma \Rightarrow \sigma \Rightarrow \sigma$ (**infixr** $m\equiv$ 76) **where** $\varphi\ m\equiv\ \psi \equiv (\lambda w.\ \varphi\ w \longleftrightarrow \psi\ w)$
abbreviation $\mathit{mforall} :: ('a \Rightarrow \sigma) \Rightarrow \sigma$ $(\forall)$ **where** $\forall\ \Phi \equiv (\lambda w.\ \forall x.\ \Phi\ x\ w)$
abbreviation $\mathit{mexists} :: ('a \Rightarrow \sigma) \Rightarrow \sigma$ $(\exists)$ **where** $\exists\ \Phi \equiv (\lambda w.\ \exists x.\ \Phi\ x\ w)$
abbreviation $\mathit{mLeibeq} :: \mu \Rightarrow \mu \Rightarrow \sigma$ (**infixr** $mL{=}$ 90) **where** $x\ mL{=}\ y \equiv \forall(\lambda\varphi.\ (\varphi\ x\ m\rightarrow\ \varphi\ y))$
abbreviation $\mathit{mbox} :: \sigma \Rightarrow \sigma$ $(\Box)$ **where** $\Box\ \varphi \equiv (\lambda w.\ \forall v.\ \ w\ r\ v \longrightarrow \varphi\ v)$
abbreviation $\mathit{mdia} :: \sigma \Rightarrow \sigma$ $(\Diamond)$ **where** $\Diamond\ \varphi \equiv (\lambda w.\ \exists v.\ w\ r\ v \wedge \varphi\ v)$

For grounding lifted formulas, the meta-predicate *valid* is introduced.

abbreviation $\mathit{valid} :: \sigma \Rightarrow \mathit{bool}$ ([-]) **where** $[p] \equiv \forall w.\ p\ w$

3 Gödel's Ontological Argument

Constant symbol P (Gödel's 'Positive') is declared.

consts $P :: (\mu \Rightarrow \sigma) \Rightarrow \sigma$

The meaning of P is restricted by axioms *A1(a/b)*: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$ (Either a property or its negation is positive, but not both.) and *A2*: $\forall\phi\forall\psi[(P(\phi)\wedge\Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$ (A property necessarily implied by a positive property is positive).

axiomatization where
A1a: $[\forall (\lambda\Phi.\ P\ (\lambda x.\ m\neg\ (\Phi\ x))\ m\!\rightarrow\ m\neg\ (P\ \Phi))]$ **and**
A1b: $[\forall (\lambda\Phi.\ m\neg\ (P\ \Phi)\ m\!\rightarrow\ P\ (\lambda x.\ m\neg\ (\Phi\ x)))]$ **and**
A2: $[\forall (\lambda\Phi.\ \forall (\lambda\Psi.\ (P\ \Phi\ m\wedge\ \Box\ (\forall (\lambda x.\ \Phi\ x\ m\!\rightarrow\ \Psi\ x)))\ m\!\rightarrow\ P\ \Psi))]$

We prove theorem T1: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$ (Positive properties are possibly exemplified). T1 is proved directly by Sledgehammer with command *sledgehammer* [*provers* = *remote-leo2*]. Sledgehammer suggests to call Metis with axioms A1a and A2. Metis sucesfully generates a proof object that is verified in Isabelle/HOL's kernel.

theorem *T1*: $[\forall (\lambda\Phi.\ P\ \Phi\ m\!\rightarrow\ \Diamond\ (\exists\ \Phi))]$
— sledgehammer [provers = remote_leo2]
by (*metis A1a A2*)

Next, the symbol G for 'God-like' is introduced and defined as $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$ (A God-like being possesses all positive properties).

definition $G :: \mu \Rightarrow \sigma$ **where** $G = (\lambda x.\ \forall (\lambda\Phi.\ P\ \Phi\ m\!\rightarrow\ \Phi\ x))$

Axiom *A3* is added: $P(G)$ (The property of being God-like is positive). Sledgehammer and Metis then prove corollary C: $\Diamond\exists xG(x)$ (Possibly, God exists).

axiomatization where *A3*: $[P\ G]$

corollary C: $[\Diamond\ (\exists\ G)]$
— sledgehammer [provers = remote_leo2]
by (*metis A3 T1*)

Axiom *A4* is added: $\forall\phi[P(\phi) \rightarrow \Box\ P(\phi)]$ (Positive properties are necessarily positive).

axiomatization where *A4*: $[\forall (\lambda\Phi.\ P\ \Phi\ m\!\rightarrow\ \Box\ (P\ \Phi))]$

Symbol *ess* for 'Essence' is introduced and defined as

$$\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$$

(An *essence* of an individual is a property possessed by it and necessarily implying any of its properties).

definition *ess* $:: (\mu \Rightarrow \sigma) \Rightarrow \mu \Rightarrow \sigma$ (**infixr** *ess 85*) **where**
$\Phi\ ess\ x = \Phi\ x\ m\wedge\ \forall (\lambda\Psi.\ \Psi\ x\ m\!\rightarrow\ \Box\ (\forall (\lambda y.\ \Phi\ y\ m\!\rightarrow\ \Psi\ y)))$

Next, Sledgehammer and Metis prove theorem *T2*: $\forall x[G(x) \rightarrow G\ ess.\ x]$ (Being God-like is an essence of any God-like being).

theorem *T2*: $[\forall (\lambda x.\ G\ x\ m\!\rightarrow\ G\ ess\ x)]$
— sledgehammer [provers = remote_leo2]
by (*metis A1b A4 G-def ess-def*)

Symbol *NE*, for 'Necessary Existence', is introduced and defined as

$$NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$$

(Necessary existence of an individual is the necessary exemplification of all its essences).

definition $NE :: \mu \Rightarrow \sigma$ **where** $NE = (\lambda x.\ \forall(\lambda\Phi.\ \Phi\ ess\ x\ m\rightarrow \Box\ (\exists\ \Phi)))$

Moreover, axiom *A5* is added: $P(NE)$ (Necessary existence is a positive property).

axiomatization where *A5*: $[P\ NE]$

The B axiom (symmetry) for relation r is stated. B is needed only for proving theorem T3 and for corollary C2.

axiomatization where *sym*: $x\ r\ y \longrightarrow y\ r\ x$

Finally, Sledgehammer and Metis prove the main theorem *T3*: $\Box\exists x G(x)$
(Necessarily, God exists).

theorem *T3*: $[\Box\ (\exists\ G)]$
— sledgehammer [provers = remote_leo2]
by (*metis A5 C T2 sym G-def NE-def*)

Surprisingly, the following corollary can be derived even without the T axiom (reflexivity).

corollary *C2*: $[\exists\ G]$
— sledgehammer [provers = remote_leo2]
by (*metis T1 T3 G-def sym*)

The consistency of the entire theory is confirmed by Nitpick.

lemma *True* **nitpick** [*satisfy*, *user-axioms*, *expect* = *genuine*] **oops**

4 Additional Results on Gödel's God.

Gödel's God is flawless: (s)he does not have non-positive properties.

theorem *Flawlessness*: $[\forall(\lambda\Phi.\ \forall(\lambda x.\ (G\ x\ m\rightarrow (m\neg\ (P\ \Phi)\ m\rightarrow\ m\neg\ (\Phi\ x)))))]$
— sledgehammer [provers = remote_leo2]
by (*metis A1b G-def*)

There is only one God: any two God-like beings are equal.

theorem *Monotheism*: $[\forall(\lambda x.\forall(\lambda y.\ (G\ x\ m\rightarrow (G\ y\ m\rightarrow (x\ mL{=}\ y)))))]$
— sledgehammer [provers = remote_leo2]
by (*metis Flawlessness G-def*)

5 Modal Collapse

Gödel's axioms have been criticized for entailing the so-called modal collapse. The prover Satallax [7] confirms this. However, sledgehammer is not able to determine which axioms, definitions and previous theorems are used by Satallax; hence it suggests to call Metis using everything, but this (unsurprinsingly) fails. Attempting to use 'Sledegehammer min' to minimize Sledgehammer's suggestion does not work. Calling Metis with *T2*, *T3* and *ess-def* also does not work.

lemma *MC*: $[\forall(\lambda\Phi.(\Phi\ m\rightarrow (\Box\ \Phi)))]$
— sledgehammer [provers = remote_satallax]
— by (metis T2 T3 ess_def)
oops

Acknowledgments: Nik Sultana, Jasmin Blanchette and Larry Paulson provided very important help on issues related to consistency checking in Isabelle. Jasmin Blanchette instructed us on producing Isabelle sessions and he showed us some useful tricks in Isabelle.

References

[1] C. Benzmüller and L.C. Paulson. Exploring properties of normal multimodal logics in simple type theory with LEO-II. In *Festschrift in Honor of Peter B. Andrews on His 70th Birthday*, pp. 386–406. College Publications.

[2] C. Benzmüller and L.C. Paulson. Quantified multimodal logics in simple type theory. *Logica Universalis (Special Issue on Multimodal Logics)*, 7(1):7–20, 2013.

[3] C. Benzmüller, F. Theiss, L. Paulson, and A. Fietzke. LEO-II - a cooperative automatic theorem prover for higher-order logic. In *Proc. of IJCAR 2008*, volume 5195 of *LNAI*, pp. 162–170. Springer, 2008.

[4] Y. Bertot and P. Casteran. *Interactive Theorem Proving and Program Development.* Springer, 2004.

[5] J.C. Blanchette, S. Böhme, and L.C. Paulson. Extending Sledgehammer with SMT solvers. *Journal of Automated Reasoning*, 51(1):109–128, 2013.

[6] J.C. Blanchette and T. Nipkow. Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In *Proc. of ITP 2010*, LNCS 6172, pp. 131–146. Springer, 2010.

[7] C.E. Brown. Satallax: An automated higher-order prover. In *Proc. of IJCAR 2012*, LNAI 7364, pp. 111 – 117. Springer, 2012.

[8] K. Gödel. *Appendix A. Notes in Kurt Gödel's Hand*, pp. 144–145. In [13], 2004.

[9] J. Hurd. First-order proof tactics in higher-order logic theorem provers. In *Design and Application of Strategies/Tactics in Higher Order Logics, NASA Tech. Rep. NASA/CP-2003-212448*, 2003.

[10] T. Nipkow, L.C. Paulson, and M. Wenzel. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic.* LNCS 2283. Springer, 2002.

[11] J. Rushby. The Ontological Argument in PVS. *CAV Workshop "Fun With Formal Methods'*, St. Petersburg, Russia, 13th of July 2013.

[12] D. Scott. *Appendix B. Notes in Dana Scott's Hand*, pp. 145–146. In [13], 2004.

[13] J.H. Sobel. *Logic and Theism: Arguments for and Against Beliefs in God.* Cambridge University Press, 2004.

[14] G. Sutcliffe and C. Benzmüller. Automated reasoning in higher-order logic using the TPTP THF infrastructure. *Journal of Formalized Reasoning*, 3(1):1–27, 2010.