Computer-Assisted Analysis of Emendations of Gödel's Ontological Proof

C. Benzmüller, L. Weber and B. Woltzenlogel Paleo

The axioms in Gödel's ontological argument [11, 16] (cf. Appendix A) entail what is called *modal collapse* [17, 18]: the formula $\varphi \rightarrow \Box\varphi$, abbreviated as MC, holds for any formula φ and not just for $\exists x. God(x)$ as intended. This fact, which has recently been confirmed with higher-order automated theorem provers [3, 6], has led to strong criticism of the argument and stimulated attempts to remedy the problem. Hájek [14, 13] proposed the use of cautious instead of full comprehension principles, and Fitting [10] suggested that greater care is necessary in the semantics of higher-order quantifiers in the presence of modalities. Others, such as Anderson [2, 1] and Bjordal [7], proposed slight emendations of Gödel's axioms and definitions. They require neither comprehension restrictions nor more complex semantics. Therefore, they are technically simpler to analyze with computer support. We have formalized them using the proof assistant Isabelle/HOL [15] together with the automated higher-order reasoners Leo-II [5], Satallax [9], Metis [12], and Nitpick [8]. Our formalizations[1] employ the embedding of higher-order modal logic (HOML) in classical higher-order logic (HOL) as introduced in previous work [3, 6, 4]. We explored the effect of different domain conditions on the provability of lemmas, theorems and even axioms. This was motivated by a controversy between Hájek and Anderson regarding the redundancy of some axioms in Anderson's emendation. In *constant domain semantics*, the individual domains are the same in all possible worlds. In *varying domain semantics*, the domains may vary from world to world. This variation is technically encoded with the help of an existence relation expressing which

This work was supported by German National Research Foundation (DFG) under grants BE 2501/9-1 and BE 2501/11-1.

[1] The formalizations are available in the subdirectories `Anderson` and `Bjordal` at `https://github.com/FormalTheology/GoedelGod/blob/master/Formalizations/Isabelle/`.

individuals actually exist in each world. Quantifiers are then uniformly formalized as *actualistic quantifiers* (i.e. guarded by the existence relation). Our main results are summarized in the following sections.

1. Anderson's Emendation (cf. Appendix B)

For both **constant domain semantics** and **varying domain semantics**, the following results hold: the theorems T1, C and T3' can be quickly automated (in logics **K**, **K** and **KB**, respectively); the axioms A4 and A5 are proven redundant (the former in logic **K4B** and the latter already in **K**); a trivial countermodel (consisting of two worlds and two individuals) to MC generated by Nitpick (for all mentioned logics), which also shows the consistency of the axioms.

The redundancy of A4 and A5 is particularly controversial. Magari [**?**] claimed that the redundancy occurs already with Gödel's original axioms and definitions. Hájek [**?**] investigated this further and claimed that Magari's claim is invalid, but true under the assumption of an additional axiom (PEP). Moreover, Hájek [**?**] claimed that the redundancy occurs for Anderson's emended axioms and definitions [**?**]. Anderson and Gettings [**?**, **?**, **?**] rebutted Hájek's claim, arguing that it holds only under constant domain semantics, while the emended argument by Anderson ought to be taken under Cocchiarella's semantics [**?**] (a particular kind of varying domain semantics). Hájek [**?**, **?**] acknowledges this rebuttal, and apparently accepts it, as evidenced by his use of A4 and A5 in his new small emendation (replacing A:A1 and A2 by a new axiom H:A12) of Anderson's variant with varying domains [**?**, **?**]. Nevertheless, he does show yet another emended version where A4 and A5 are redundant, but A3 is replaced by a stronger axiom additionally stating that the property of actual existence is positive [**?**, **?**]. Our results show that Hájek was originally right, under both constant and varying domain semantics with no need to strengthen A3 (though this need might exist when H:A12 is used instead of A:A1 and A2). It should be noted, however, that [**?**, **?**] vaguely remarks that only the quantifiers in T3' and in A:D2 need to be interpreted as actualistic quantifiers, while other may be taken as possibilistic quantifiers. We have inv mixed variant.

ToDo: tell the story of the controversy here, and then link with the mixed variant.

Mixed variant. (varying domain quantifiers are used only in the definitions of essence and NE; cf. Fitting's comments to Anderson in [10]): Also in this setting we obtain the same results as above. However, if a varying domain quantifier is used only in the definition of NE, then the situation changes slightly. Now axiom A5 is no longer provable and a countermodel is reported by Nitpick. The remaining results are as before.

2. Bjordal's Emendation (cf. Appendix C)

For both **constant domain semantics** and **varying domain semantics**, the following results hold:

Gödel's axiom A2, A3 can be quickly automatically derived in logic **K** from Bjordal's definition B:D. A4 can be proved in logic **KT** (reflexivity). Proving Gödel's D1 from B:D is possible in logic **K4**. Conversely, the proof that B:D follows from D1, A2, A3 and A4 is possible already in logic **K**. Hence, Bjordal's lemma B:L1 holds in logic **S4**. The provers also show that theorem T3 follows from B:D, B:A1 and B:A2 already in logic **KB**. Modal collapse does not follow in Bjordal's setting as Nitpick demonstrates with a countermodel (consisting of two worlds and one individual).

@Bruno: we
tually also a
for MC1 we
ger counterm

ToDo: disc
that Bjorda
more resist
than Ander

3. Conclusions

Anderson's emendation (cf. Appendix B; , which we have analysed for different domain conditions. These variations were motivated by various comments on Anderson's work in the literature.

In this approach full comprehension is naturally "built-in" since the underlying HOL supports λ-abstraction.

Summary (what else can we say here, feel free to add): Using our approach, the formalization and (partly) automated analysis of different variants of Anderson's and Bjordal's emendations of Gödel's ontological argument has been surprisingly straightforward. The provers confirmed the claimed results and in a few cases they have even contributed some novel insights. The weakening of the comprehension principles would clearly constitute another interesting parameter for further experiments. However, this seems hard to achieve in our approach, since full comprehension is naturally built-in.

References

[1] A.C. Anderson and M. Gettings. Gödel ontological proof revisited. In *Gödel'96: Logical Foundations of Mathematics, Computer Science, and Physics: Lecture Notes in Logic 6*, pages 167–172. Springer, 1996.

[2] C.A. Anderson. Some emendations of Gödel's ontological proof. *Faith and Philosophy*, 7(3), 1990.

[3] C. Benzmüller and B. Woltzenlogel Paleo. Automating Gödel's ontological proof of God's existence with higher-order automated theorem provers. In *ECAI 2014*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 93 – 98. IOS Press, 2014.

[4] C. Benzmüller and L.C. Paulson. Quantified multimodal logics in simple type theory. *Logica Universalis*, 7(1):7–20, 2013.

[5] C. Benzmüller, F. Theiss, L. Paulson, and A. Fietzke. LEO-II - a cooperative automatic theorem prover for higher-order logic (system description). In *Proc. of IJCAR 2008*, number 5195 in LNCS, pages 162–170. Springer, 2008.

[6] C. Benzmüller and B. Woltzenlogel-Paleo. Formalization, mechanization and automation of Gödel's proof of God's existence. *arXiv:1308.4526*, 2013.

[7] F. Bjørdal. Understanding gödel's ontological argument. In T. Childers, editor, *The Logica Yearbook 1998*. Filosofia, 1999.

[8] 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*, number 6172 in LNCS, pages 131–146. Springer, 2010.

[9] C.E. Brown. Satallax: an automatic higher-order prover. *J. Autom. Reasoning*, pages 111–117, 2012.

[10] Melvin Fitting. *Types, Tableaus, and Gödel's God*. Kluwer, 2002.

[11] K. Gödel. *Appx.A: Notes in Kurt Gödel's Hand*, pages 144–145. In [18], 2004.

[12] 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*, pages 56–68, 2003.

[13] P. Hájek. Magari and others on gödel's ontological proof. In A. Ursini and P. Agliano, editors, *Logic and algebra*, page 125–135. Dekker, New York etc., 1996.

[14] P. Hájek. Der mathematiker und die frage der existenz gottes. In B. Buldt, E. Köhler, M. Stöltzner, P. Weibel, C. Klein, and W. Depauli-Schimanowich-Göttig, editors, *Kurt Gödel. Wahrheit und Beweisbarkeit*, pages 325–336. öbv & hpt, Wien, 2002. ISBN 3-209-03835-X.

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

[16] D. Scott. *Appx.B: Notes in Dana Scott's Hand*, pages 145–146. In [18], 2004.

[17] J.H. Sobel. Gödel's ontological proof. In *On Being and Saying. Essays for Richard Cartwright*, pages 241–261. MIT Press, 1987.

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

I think we compact pr the proofs.

Appendix A. Scott's version of Gödel's ontological argment

A1 Either a property or its negation is positive, but not both:

$$\forall\varphi[P(\neg\varphi) \leftrightarrow \neg P(\varphi)]$$

A2 A property necessarily implied by a positive property is positive:

$$\forall\varphi\forall\psi[(P(\varphi) \wedge \Box\forall x[\varphi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$$

T1 Positive properties are possibly exemplified:

$$\forall\varphi[P(\varphi) \rightarrow \Diamond\exists x\varphi(x)]$$

D1 A *God-like* being possesses all positive properties:

$$G(x) \equiv \forall\varphi[P(\varphi) \rightarrow \varphi(x)]$$

A3 The property of being God-like is positive:

$$P(G)$$

C Possibly, a God-like being exists:

$$\Diamond\exists x G(x)$$

A4 Positive properties are necessarily positive:

$$\forall\varphi[P(\varphi) \rightarrow \Box\, P(\varphi)]$$

D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties:

$$\varphi\ ess\ x \equiv \varphi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\varphi(y) \rightarrow \psi(y)))$$

T2 Being God-like is an essence of any God-like being:

$$\forall x[G(x) \rightarrow G\ ess\ x]$$

D3 *Necessary existence* of an individual is the necessary exemplification of all its essences:

$$NE(x) \equiv \forall\varphi[\varphi\ ess\ x \rightarrow \Box\exists y\varphi(y)]$$

A5 Necessary existence is a positive property:

$$P(NE)$$

L1 If a god-like being exists, then necessarily a god-like being exists:

$$\exists x G(x) \rightarrow \Box\exists y G(y)$$

L2 If possibly a god-like being exists, then necessarily a god-like being exists:

$$\Diamond\exists x G(x) \rightarrow \Box\exists y G(y)$$

T3 Necessarily, a God-like being exists:

$$\Box\exists x G(x)$$

Appendix B. Anderson's Emendation

A:A1 If a property is positive, its negation is not positive:

$$\forall\varphi[P(\varphi) \rightarrow \neg P(\neg\varphi)]$$

A2 A property necessarily implied by a positive property is positive:

$$\forall\varphi\forall\psi[(P(\varphi) \wedge \Box\forall x[\varphi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$$

T1 Positive properties are possibly exemplified:

$$\forall\varphi[P(\varphi) \rightarrow \Diamond\exists x\varphi(x)]$$

A:D1 A *God-like* being necessarily possesses those and only those properties that are positive:

$$G_A(x) \equiv \forall\varphi[P(\varphi) \leftrightarrow \Box\varphi(x)]$$

A3' The property of being God-like is positive:

$$P(G_A)$$

C Possibly, a God-like being exists:

$$\Diamond\exists xG(x)$$

A4 Positive properties are necessarily positive:

$$\forall\varphi[P(\varphi) \rightarrow \Box\ P(\varphi)]$$

A:D2 An *essence* of an individual is a property that necessarily implies those and only those properties that the individual has necessarily:

$$\varphi\ \mathit{ess}_A\ x \equiv \forall\psi[\Box\psi(x) \leftrightarrow \Box\forall y(\varphi(y) \rightarrow \psi(y))]$$

T2' Being God-like is an essence of any God-like being:

$$\forall x[G_A(x) \rightarrow G_A\ \mathit{ess}_A\ x]$$

D3' *Necessary existence* of an individual is the necessary exemplification of all its essences:

$$NE_A(x) \equiv \forall\varphi[\varphi\ \mathit{ess}_A\ x \rightarrow \Box\exists y\varphi(y)]$$

A5' Necessary existence is a positive property:

$$P(NE_A)$$

L1' If a god-like being exists, then necessarily a god-like being exists:

$$\exists xG_A(x) \rightarrow \Box\exists yG_A(y)$$

L2' If possibly a god-like being exists, then necessarily a god-like being exists:

$$\Diamond\exists xG_A(x) \rightarrow \Box\exists yG_A(y)$$

T3' Necessarily, a God-like being exists:

$$\Box\exists xG_A(x)$$

Appendix C. Bjordal's Alternative

In Bjordal's emendation G (God-like) is taken as primitive and P (Positive) is defined (cf. definition D).

B:D A formulas ϕ is positive iff it is necessarily the case that anything which is God-like has the property ϕ.

$$P(\phi) \equiv \Box\forall x(G(x) \to \phi(x))$$

B:L1 D is logically equivalent in S4 with the union of Gödel's definition D1 and axioms A2, A3 and A4.

$$D \leftrightarrow D1 \wedge A2 \wedge A3 \wedge A4$$

The proof splits into the two implication directions B:L1$^{\to}$ and B:L1$^{\leftarrow}$. B:L1$^{\to}$ can be further split into four single steps.

B:D2 ϕ is a maximal composite of object x's positive properties iff x has ϕ and ϕ is positive and all positive properties ψ which x has are such that is necessarily the case that all objects which have ϕ also have ψ.

$$MCP(\phi, x) \equiv (\phi(x) \wedge P(\phi)) \wedge \forall\psi((\psi(x) \wedge P(\psi)) \to \Box\forall y(\phi(y) \to \psi(y)))$$

B:D3 x has the N-property iff x is such that if ϕ is a maximal composite of x's positive properties then it is necessary that some object y has the property ϕ.

$$N(x) \equiv \forall\phi(MCP(\phi, x) \to \Box\forall y\phi(y))$$

B:A1 If a property is positive, its negation is not positive:

$$\forall\varphi[P(\varphi) \to \neg P(\neg\varphi)]$$

B:A2 The N-property is positive.

$$P(N)$$

T3 Necessarily, a God-like being exists:

$$\Box\exists x G(x)$$

C. Benzmüller
Dep. of Mathematics and Computer Science, Freie Universität Berlin, Germany
e-mail: c.benzmueller@fu-berlin.com

L. Weber
Dep. of Mathematics and Computer Science, Freie Universität Berlin, Germany
e-mail: leon.weber@fu-berlin.de

B. Woltzenlogel Paleo
Room HA0402, Favoritenstraße 9, 1040 Wien, Austria
e-mail: bruno.wp@gmail.com