Computer-Assisted Analysis of the Anderson-Hájek Ontological Controversy

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

The axioms in Gödel's ontological proof [7, 8] (cf. Appendix A) entail what is called *modal collapse* [18, 9]: the formula $\varphi \to \Box\varphi$, 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 [1, 3], has led to strong criticism of the argument and stimulated attempts to remedy the problem. Hájek [16, 13] proposed the use of cautious instead of full comprehension principles, and Fitting [11] suggested that greater care is necessary in the semantics of higher-order quantifiers in the presence of modalities. Others, such as Anderson [17] and Anderson and Gettings [15], **Hajek2002** and Bjordal Bjørdal [14], 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 [12] together with the automated higher-order reasoners Leo-II [6], Satallax [4], Metis [10], and Nitpick [5]. Our formalizations[1] employ the embedding of higher-order modal logic (HOML) in classical higher-order logic (HOL) as introduced in previous work [1, 3, 2]. 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 individuals actually exist in each world. Quantifiers

---

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`, `Hjek` and `Bjordal` at `github.com/FormalTheology/GoedelGod/blob/master/Formalizations/Isabelle/`.

are then uniformly formalized as *actualistic quantifiers* (i.e. guarded by the existence relation). Our main results are summarized here.

For both **constant domain semantics** and **varying domain semantics**, the following results hold for *Anderson's Emendation* (cf. Appendix B): 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[2] (the former in logic **K4B** and the latter already in **K**); a trivial countermodel (consisting of two worlds and two individuals) for MC is generated by Nitpick (for all mentioned logics); all axioms and definitions are shown to be mutually consistent.

The redundancy of A4 and A5 is particularly controversial. Magari [**Magari1988**] claimed that A4 and A5 are superfluous[3], arguing that T3 is true in all models of the other axioms and definitions by Gödel. Hájek [16, p. 5-6] investigated this further, refuted Magari's claim, but claimed that A4 and A5 are indeed superfluous under the assumption of an additional axiom (PEP, pointwise equality for positiveness), which Magari probably assumed silently. Moreover, Hájek Hájek [16, p. 2] cites his earlier work[4] [13], where he claims (in Theorem 5.3) that for Anderson's emended theory [17], A4 and A5 are not only superfluous, but also redundant. **AG1996**, **footnote 1**, **page 1**, in a footnote, rebutted Hájek's claim, arguing that the redundancy of A4 and A5 holds only under constant domain semantics, while Anderson's emended theory ought to be taken under Cocchiarella's semantics [**Cocchiarella**] (a varying domain semantics). Our results show that Hájek was originally right, under both constant and varying domain semantics.

Nevertheless, **Hajek2002** (in page 7) acknowledges Anderson's rebuttal, and apparently accepts it, as evidenced by his use[5] of A4 and A5, as well as varying domain semantics, in his new emendation (named $\mathcal{AOE}'$[**Hajek2002**], cf. Appendix C), which replaces Anderson's A:A1 and A2 by a simpler axiom H:A12. Surprisingly, the computer-assisted formalization of $\mathcal{AOE}'$shows that A4 and A5 are superfluous (not needed for proving T3). Moreover, A4 and A5 are independent[6] of the other axioms and definitions. Therefore, A4 and A5 are not redundant, despite their superfluousness.

Furthermore, he does show yet another emendation (his $\mathcal{AOE}'_0$, cf. Appendix D) where A4 and A5 are superfluous (though no claim is made w.r.t. to their redundancy), if A3 is replaced by a stronger axiom (H:A3) additionally stating that the property of actual existence is positive [**Hajek 2002**, **section 5**]. However, $\mathcal{AOE}'_0$is regarded as just an intermediary step towards his final emendation ($\mathcal{AOE}''$, cf. E), which do use A4 and A5, albeit in a modified form (i.e. H:A4 and H:A5).

---

[2] An axiom A is redundant w.r.t. a set of axioms S iff A is entailed by S.

[3] An axiom A is superfluous w.r.t. a set of axioms S iff T3 is entailed by $S \setminus \{A\}$.

[4] Although [13] precedes [16] in writing, it was published only 5 years later, in German.

[5] A4 and A5 are used by **Hajek2002** (page 11) in the proofs of, respectively, Lemma 4 and Theorem 4.

[6] An axiom A is *independent* of a set of axioms S iff there are models of S where A is true and other models of S where A if false.

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 [17, footnote 14] 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 [11]): 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.

1. Bjordal's Emendation (cf. Appendix F)

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

2. Conclusions

ToDo: disc
that Bjorda
more resist
than Ander

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] C. Benzmüller and Bruno Woltzenlogel Paleo. "Automating Gödel's Ontological Proof of God's Existence with Higher-order Automated Theorem Provers". In: *ECAI 2014*. Vol. 263. Frontiers in Artificial Intelligence and Applications. IOS Press, 2014, pp. 93 –98. DOI: 10.3233/978-1-61499-419-0-93. URL: http://christoph-benzmueller.de/papers/C40.pdf.

[2] C. Benzmüller and L.C. Paulson. "Quantified Multimodal Logics in Simple Type Theory". In: *Logica Universalis* 7.1 (2013), pp. 7–20. DOI: 10.1007/s11787-012-0052-y. URL: http://christoph-benzmueller.de/papers/J23.pdf.

[3] C. Benzmüller and B. Woltzenlogel-Paleo. "Formalization, Mechanization and Automation of Gödel's Proof of God's Existence". In: *arXiv:1308.4526* (2013). URL: http://arxiv.org/abs/1308.4526.

[4] C.E. Brown. "Satallax: an automatic higher-order prover". In: *J. Autom. Reasoning* (2012), pp. 111–117.

[5] 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. Springer, 2010, pp. 131–146. ISBN: 978-3-642-14051-8.

[6] 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*. LNCS 5195. Springer, 2008, pp. 162–170. DOI: 10.1007/978-3-540-71070-7_14. URL: http://christoph-benzmueller.de/papers/C26.pdf.

[7] K. Gödel. "Appx.A: Notes in Kurt Gödel's Hand". In: *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge U. Press, 2004, pp. 144–145. ISBN: 9781139449984. URL: http://books.google.de/books?id=ZQh8QJOQdOQC.

[8] D. Scott. "Appx.B: Notes in Dana Scott's Hand". In: *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge U. Press, 2004, pp. 145–146. ISBN: 9781139449984. URL: http://books.google.de/books?id=ZQh8QJOQdOQC.

[9] J.H. Sobel. *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge U. Press, 2004. ISBN: 9781139449984. URL: http://books.google.de/books?id=ZQh8QJOQdOQC.

[10] 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, pp. 56–68.

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

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

[13] P. Hájek. "Der Mathematiker und die Frage der Existenz Gottes". In: *Kurt Gödel. Wahrheit und Beweisbarkeit*. Ed. by B. Buldt, E. Köhler,

M. Stöltzner, P. Weibel, C. Klein, and W. Depauli-Schimanowich-Göttig. ISBN 3-209-03835-X. öbv & hpt, Wien, 2001, pp. 325–336.

[14] F. Bjørdal. "Understanding Gödel's Ontological Argument". In: *The Logica Yearbook 1998*. Ed. by T. Childers. Filosofia, 1999.

[15] 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*. Springer, 1996, pp. 167–172.

[16] P. Hájek. "Magari and others on Gödel's ontological proof". In: *Logic and algebra*. Ed. by A. Ursini and P. Agliano. Dekker, New York etc., 1996, 125–135.

[17] C.A. Anderson. "Some emendations of Gödel's ontological proof". In: *Faith and Philosophy* 7.3 (1990).

[18] J.H. Sobel. "Gödel's Ontological Proof". In: *On Being and Saying. Essays for Richard Cartwright*. MIT Press, 1987, pp. 241–261.

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:

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

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:

$$\exists xG(x) \rightarrow \Box\exists yG(y)$$

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

$$\Diamond\exists xG(x) \rightarrow \Box\exists yG(y)$$

T3 Necessarily, a God-like being exists:

$$\Box\exists xG(x)$$

Appendix B. Anderson's Emendation

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

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

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

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

T1 Positive properties are possibly exemplified:

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

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:

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

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

$$\varphi\ ess_A\ x \equiv \forall\psi[\Box\psi(x) \leftrightarrow \Box\forall y(\varphi(y) \to \psi(y))]$$

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

$$\forall x[G_A(x) \to G_A\ ess_A\ x]$$

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

$$NE_A(x) \equiv \forall\varphi[\varphi\ ess_A\ x \to \Box\exists y\varphi(y)]$$

A5' Necessary existence is a positive property:

$$P(NE_A)$$

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

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

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

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

T3' Necessarily, a God-like being exists:

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

Appendix C. Hájek's First Emendation $\mathcal{AOE}'$

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

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

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

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

T1 Positive properties are possibly exemplified:

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

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:

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

A4 Positive properties are necessarily positive:

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

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

$$\varphi\ ess_A\ x \equiv \forall\psi[\Box\psi(x) \leftrightarrow \Box\forall y(\varphi(y) \to \psi(y))]$$

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

$$\forall x[G_A(x) \to G_A\ ess_A\ x]$$

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

$$NE_A(x) \equiv \forall\varphi[\varphi\ ess_A\ x \to \Box\exists y\varphi(y)]$$

A5' Necessary existence is a positive property:

$$P(NE_A)$$

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

$$\exists x G_A(x) \to \Box\exists y G_A(y)$$

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

$$\Diamond\exists x G_A(x) \to \Box\exists y G_A(y)$$

T3' Necessarily, a God-like being exists:

$$\Box\exists x G_A(x)$$

Appendix D. Hájek's Second Emendation $\mathcal{AOE}'_0$

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:

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

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:

$$\varphi\ ess_A\ x \equiv \forall\psi[\Box\psi(x) \leftrightarrow \Box\forall y(\varphi(y) \rightarrow \psi(y))]$$

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

$$\forall x[G_A(x) \rightarrow G_A\ ess_A\ x]$$

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

$$NE_A(x) \equiv \forall\varphi[\varphi\ ess_A\ x \rightarrow \Box\exists y\varphi(y)]$$

A5' Necessary existence is a positive property:

$$P(NE_A)$$

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

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

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

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

T3' Necessarily, a God-like being exists:

$$\Box\exists x G_A(x)$$

Appendix E. Hájek's Third Emendation $\mathcal{AOE}''$

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

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

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

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

T1 Positive properties are possibly exemplified:

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

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:

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

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

$$\varphi\ ess_A\ x \equiv \forall\psi[\Box\psi(x) \leftrightarrow \Box\forall y(\varphi(y) \to \psi(y))]$$

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

$$\forall x[G_A(x) \to G_A\ ess_A\ x]$$

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

$$NE_A(x) \equiv \forall\varphi[\varphi\ ess_A\ x \to \Box\exists y\varphi(y)]$$

A5' Necessary existence is a positive property:

$$P(NE_A)$$

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

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

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

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

T3' Necessarily, a God-like being exists:

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

Appendix F. 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`