# 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* [20, 9]: the formula $\varphi \rightarrow \Box\varphi$, abbreviated as MC, holds for any formula $\varphi$ 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 [17, 14] proposed the use of cautious instead of full comprehension principles, and Fitting [11] took greater care of the semantics of higher-order quantifiers in the presence of modalities. Others, such as Anderson [18], Hájek [12] and Bjørdal [15], proposed 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 [13] 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 are then uniformly formalized as *actualistic quantifiers* (i.e. guarded by the existence relation). Our main results are summarized here.

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

For all emendations and variants discussed here, the axioms and definitions have been shown to be consistent and not to entail modal collapse.

For both **constant domain semantics** and **varying domain semantics**, the following results hold for *Anderson's Emendation* (cf. Appendix B): 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 (with 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 [19] 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 [17, p. 5-6] investigated this further, and claimed that Magari's claim is not valid, but is nevertheless true under additional silent assumptions by Magari. Moreover, Hájek [17, p. 2] cites his earlier work[4] [14], where he claims (in Theorem 5.3) that for Anderson's emended theory [18], A4 and A5 are not only superfluous, but also redundant. Anderson and Gettings [16, footnote 1 in p. 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 [21] (a varying domain semantics). Our results show that Hájek was originally right, under both constant and varying domain semantics.

Nevertheless, Hájek [12, p. 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 $\mathcal{AOE}'$[12, sec. 4], 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 still superfluous. Moreover, A4 and A5 are independent[6] of the other axioms and definitions. Therefore, A4 and A5 are not redundant, despite their superfluousness.

Although, Hájek did not notice the superfluousness of A4 and A5 in his $\mathcal{AOE}'$, he did describe 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 when it comes to God-like beings [12, sec. 5]. Formalization of $\mathcal{AOE}'_0$ shows that A4 is not only superflous, but also redundant. For A5, no conclusive results were achieved; neither a proof nor a countermodel could be automatically generated. Surprisingly, a countermodel for the weaker A3 was succesfully generated. This

[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 [14] precedes [17] in writing, it was published only 5 years later, in German.
[5] A4 and A5 are used by Hájek [12, p. 11] in, 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.

is somewhat unsatisfactory (for theistic goals), because it shows that $\mathcal{AOE}'_0$ does not entail the positiveness of being God-like.

Nevertheless, $\mathcal{AOE}'_0$ is explicitly regarded by Hájek [12, p. 12] as just an intermediary step towards a more natural theory, based on a more sophisticated notion of positiveness. That is his final emendation ($\mathcal{AOE}''$, cf. E), which restores A3 and does use A4 and A5, albeit in a modified form (i.e. H:A4 and H:A5). The formalization of $\mathcal{AOE}''$ shows that H:A4 is independent. The old A5 is independent as well, and both H:A4 and H:A5 are superfluous, but no conclusive results were achieved regarding independence or redundancy of H:A5.

Additionally, Anderson [18, footnote 14] remarks that only the quantifiers in T3' and in A:D2 need to be interpreted as actualistic quantifiers, while others may be taken as possibilistic quantifiers. Our computer-assisted study of this mixed variant shows that A4 is still redundant, but A5 becomes independent (hence not redundant). Unfortunately, a countermodel for T3 can then be found.

The controversy over the superfluousness of A4 and A5 indicates a trend to reduce the ontological argument to its bare essentials. In this regard, already Anderson [18, p. 7] indicates that, by taking a notion of *defective* as primitive and defining the notion of *positive* upon it, axioms A:A1, A2 and A4 become derivable. These claims have been confirmed by the automated theorem provers (in logic **K4B**). Within the same trend, the alternative proposed by Bjørdal [15] (cf. Appendix F) achieves a high level of minimality. He takes the property of being God-like as a primitive and defines (B:D1) the positive properties as those properties necessarily possessed by every God-like being. He then briefly indicates (B:L1) that B:D1 is logically equivalent, under modal logic **S4**, to the conjunction of D1', A2', A3' and A4'. This has been confirmed in the computer-assisted formalization: A2' and A3' can be quickly automatically derived in logic **K**. A4' can be proved in logic **KT** (i.e. assuming reflexivity of the accessibility relation). For constant domain semantics, proving D1 is possible in logic **K4**, whereas for varying domain semantics, a counter model can be found even in logic **S5**. Conversely, the proof that B:D1 is entailed by D1', A2', A3' and A4' is possible already in logic **K**. The provers also show that theorem T3' follows from B:D1, B:D2, B:D3, A:A1' and A5' already in logic **KB**. Bjordal's last paragraph briefly mentions Hájek's ideas about the superfluousness of A5' and claims that it is possible, with (unclear) additional modifications of the definitions, to eliminate A5' from his theory as well. Without any additional modification, the automated reasoners show that A5' is actually not superfluous. All these results, with the exception of the aforementioned countermodel for D1', hold for both constant and varying domain semantics.

Using our approach, the formalization and (partly) automated analysis of several variants of Gödel's ontological argument has been surprisingly straightforward. The provers not only confirmed many claimed results, but also exposed a few mistakes and novel insights in a long standing controversy.

We believe the technology employed in this work is ready to be fruitfully adopted in larger scale by philosophers.

## 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] Petr Hájek. "A New Small Emendation of Gödel's Ontological Proof". In: *Studia Logica* 71.2 (2002), pp. 149–164. DOI: 10.1023/A:1016583920890.

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

[14] 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.

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

[16] 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.

[17] 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.

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

[19] R. Magari. "Logica e Teofilia". In: *Notizie di Logica* VII.4 (1988).

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

[21] Nino B. Cocchiarella. "A Completeness Theorem in Second Order Modal Logic". In: *Theoria* 35 (1969), pp. 81–103.

## Appendix A. Scott's version of Gödel's ontological argument

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:

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

A5 Necessary existence is a positive property:

$$P(\mathrm{NE})$$

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:

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

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:

$$\mathrm{NE}_A(x) \equiv \forall\varphi[\varphi \ ess_A \ x \rightarrow \Box\exists y\varphi(y)]$$

A5' Necessary existence is a positive property:

$$P(\mathrm{NE}_A)$$

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

H:A12 The negation of a property necessarily implied by a positive property is not positive:

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

H:D1 A *God-like* being necessarily possesses those and only those properties that are necessarily implied by a positive property:

$$G_H(x) \equiv \forall\varphi[\Box\varphi(x) \leftrightarrow \exists\psi[P(\psi) \wedge \Box\forall x[\psi(x) \rightarrow \varphi(x))]]$$

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

$$P(G_H)$$

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:

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

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

$$\mathrm{NE}_A(x) \equiv \forall\varphi[\varphi\ ess_A\ x \rightarrow \Box\exists y[\varphi(y)]]$$

A5' Necessary existence is a positive property:

$$P(\mathrm{NE}_A)$$

L3 (1) The negation of a positive property is not positive:

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

(2) Positive properties are possibly exemplified:

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

(3) If a god-like being exists, then necessarily a god-like being exists:

$$\forall x[G_H(x) \rightarrow \Box G_H(x)))]$$

(4) All positive properties are necessarily implied by the property of being god-like:

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

L4 Being God-like is an essence of any God-like being:

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

T3' Necessarily, a God-like being exists:

$$\Box\exists x G_H(x)$$

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

H:A12 The negation of a property necessarily implied by a positive property is not positive:

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

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

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

H:A3 The property of being God-like and existing actually is positive:

$$P(G_A \wedge E)$$

T3' Necessarily, a God-like being exists:

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

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

H:A12 The negation of a property necessarily implied by a positive property is not positive:

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

D4 A property is positive$^{\#}$ iff it is necessarily implied by a positive property:

$$P^{\#}(\varphi) \equiv \exists\psi[P(\psi) \wedge \Box\forall x[\psi(x) \rightarrow \varphi(x)]]$$

H:D1 A *God-like* being necessarily possesses those and only those properties that are positive$^{\#}$:

$$G_H(x) \equiv \forall\varphi[P^{\#}(\varphi) \leftrightarrow \Box\varphi(x)]$$

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

$$P(G_H)$$

H: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:

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

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

$$\mathrm{NE}_A(x) \equiv \forall\varphi[\varphi\ ess_A\ x \rightarrow \Box\exists y[\varphi(y)]]$$

H:A5 Necessary existence is a positive$^{\#}$ property:

$$P^{\#}(\mathrm{NE}_A)$$

## Appendix F. Bjordal's Alternative

$G_B$ (God-like) is taken as primitive and $P_B$ (Positive) is defined.

B:D1 A property is positive iff it is necessarily possessed by every God-like being.

$$P_B(\phi) \equiv \Box\forall x(G_B(x) \rightarrow \phi(x))$$

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

$$\text{B:D1} \leftrightarrow \text{D1'} \wedge \text{A2'} \wedge \text{A3'} \wedge \text{A4'}$$

B:D2 a *maximal composite* of an individual's positive properties is a positive property possessed by the individual and necessarily implying every positive property possessed by the individual.

$$\text{MCP}(\phi, x) \equiv (\phi(x) \wedge P_B(\phi)) \wedge \forall\psi((\psi(x) \wedge P_B(\psi)) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$$

B:D3 *Necessary existence* of an individual is the necessary exemplification of all its maximal composites.

$$\text{NE}_B(x) \equiv \forall\phi(\text{MCP}(\phi, x) \rightarrow \Box\exists y\phi(y))$$

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

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

A5' Necessary existence is a positive property.

$$P_B(\text{NE}_B)$$

T3' Necessarily, a God-like being exists:

$$\Box\exists x G_B(x)$$

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