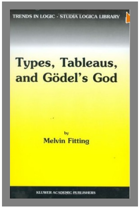
Formalization, Mechanization and Automation of Gödel's Proof of God's Existence

Christoph Benzmüller and **Bruno Woltzenlogel Paleo**

November 1, 2013



$$\dfrac{\dfrac{\text{Axiom 3}}{P(G)} \qquad \dfrac{\dfrac{\text{Theorem 1}}{\forall\varphi.[P(\varphi) \rightarrow \Diamond\exists x.\varphi(x)]}}{P(G) \rightarrow \Diamond\exists x.G(x)}\ \forall_E}{\Diamond\exists x.G(x)}\ \rightarrow_E$$


SPIEGEL ONLINE WISSENSCHAFT

Login | Registrierung

Politik | Wirtschaft | Panorama | Sport | Kultur | Netzwelt | Wissenschaft | Gesundheit | einestages | Karriere | Uni | Schule | Reise | Auto

Nachrichten > Wissenschaft > Mensch > Mathematik > Formel von Kurt Gödel: Mathematiker bestätigen Gottesbeweis

Formel von Kurt Gödel: Mathematiker bestätigen Gottesbeweis

Von Tobias Hürter

picture-alliance/ Imagno/ Wiener Stadt- und Landesbibliothek

Kurt Gödel (um das Jahr 1935): Der Mathematiker hielt seinen Gottesbeweis jahrzehntelang geheim

Ein Wesen existiert, das alle positiven Eigenschaften in sich vereint. Das bewies der legendäre Mathematiker Kurt Gödel mit einem komplizierten Formelgebilde. Zwei Wissenschaftler haben diesen Gottesbeweis nun überprüft - und für gültig befunden.

Montag, 09.09.2013 – 12:03 Uhr

Drucken | Versenden | Merken

Jetzt sind die letzten Zweifel ausgeräumt: Gott existiert tatsächlich. Ein Computer hat es mit kalter Logik bewiesen - das MacBook des Computerwissenschaftlers Christoph Benzmüller von der Freien Universität Berlin.

Germany
- Telepolis & Heise
- Spiegel Online
- FAZ
- Die Welt
- Berliner Morgenpost
- Hamburger Abendpost
- . . .

Austria
- Die Presse
- Wiener Zeitung
- ORF
- . . .

Italy
- Repubblica
- Ilsussidario
- . . .

India
- DNA India
- Delhi Daily News
- India Today
- . . .

US
- ABC News
- . . .

International
- Spiegel International
- Yahoo Finance
- United Press Intl.
- . . .

SCIENCE NEWS

HOME / SCIENCE NEWS / RESEARCHERS SAY THEY USED MACBOOK TO PROVE GOEDEL'S GOD THEOREM

Researchers say they used MacBook to prove Goedel's God theorem

Oct. 23, 2013 | 8:14 PM | 1 comments

Are we in contact with Steve Jobs? No

Do you really need a MacBook to obtain the results? No

Is Apple sending us money? No (but maybe they should)

Def: Ontological Argument/Proof

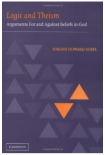
* deductive argument
* for the existence of god
* starting from premises, which are justified by pure reasoning, i.e. they do not depend on observation in the world.

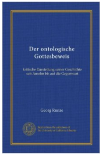
Existence of God: different types of arguments/proofs

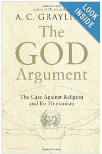
- a posteriori (use experience/observation in the world)
  - teleological
  - cosmological
  - moral
  - ...
- a priori (based on pure reasoning, independent)
  - ontological argument
    - definitional
    - modal
    - ...
  - other a priori arguments

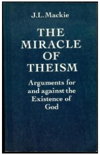
Wohl eine jede Philosophie kreist um den ontologischen Gottesbeweis

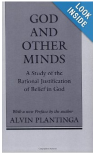
(Adorno, Th. W.: Negative Dialektik. Frankfurt a. M. 1966, p.378)

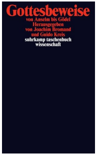
…












…

Introduction

Rich history on ontological arguments (pros and cons)

… Anselm v. C. … Th. Aquinas …… Descartes Spinoza Leibniz … Hume Kant … Hegel … Frege … Hartshorne Malcolm Lewis Plantinga Gödel …

Anselm's notion of God:

"God is that, than which nothing greater can be conceived."

Gödel's notion of God:

"A God-like being possesses all 'positive' properties."

To show by logical reasoning:

"(Necessarily) God exists."

Different Interests in Ontological Arguments:

- Philosophical: Boundaries of Metaphysics & Epistemology
  - We talk about a metaphysical concept (God),
  - but we want to draw a conclusion for the real world.
  - Necessary Existence (NE): metaphysical NE vs. logical NE vs. modal NE
- Theistic: Successful argument should convince atheists.
- Our: Can computers (theorem provers) be used
  - to formalize the definitions and axioms?
  - to verify the arguments step-by-step?
  - to fully automate (sub-)arguments?

"Computer-assisted Theoretical Philosophy"

Main challenge: No provers for *Higher-order Modal Logic* (HML)

Our solution: Embedding in *Higher-order Classical Logic* (HOL)

[BenzmüllerPaulson, Logica Universalis, 2013]

What we did (rough outline for remaining presentation!):

A: pen and paper: detailed natural deduction proof

B: formalization: in classical higher-order logic (HOL)
proof automation: theorem provers LEO-II and SATALLAX
consistency: model finder NITPICK (NITROX)

C: step-by-step verification: proof assistant COQ

D: automation & verification: proof assistant ISABELLE

Did we get new results? Yes — let's discuss later!

Gödel's Manuscript

ToDo: Show Goedel's Manuscript

Versions

A1 Either a property is positive or its negation is (never both):
$\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

A2 A property necessarily implied by a positive property is positive: $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

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:
$G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

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:
$\forall\phi[P(\phi) \rightarrow \Box\, P(\phi)]$

D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties:
$\phi\ ess\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

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:
$E(x) \leftrightarrow \forall\phi[\phi\ ess\ x \rightarrow \Box\exists y\phi(y)]$

A5 Necessary existence is a positive property: $P(E)$

T3 Necessarily, God exists: $\Box\exists x G(x)$

Proof Overview

$$\textbf{D1:}\ G(x) \equiv \forall\varphi.[P(\varphi) \to \varphi(x)]$$

$$\textbf{D2:}\ \varphi\ ess\ x \equiv \varphi(x) \wedge \forall\psi.(\psi(x) \to \Box\forall x.(\varphi(x) \to \psi(x)))$$

$$\textbf{D3:}\ E(x) \equiv \forall\varphi.[\varphi\ ess\ x \to \Box\exists y.\varphi(y)]$$

$$\dfrac{\dfrac{}{P(G)}\,\textbf{A3} \qquad \dfrac{\dfrac{}{\forall\varphi.\forall\psi.[(P(\varphi) \wedge \Box\forall x.[\varphi(x) \to \psi(x)]) \to P(\psi)]}\,\textbf{A2} \qquad \dfrac{}{\forall\varphi.[P(\neg\varphi) \to \neg P(\varphi)]}\,\textbf{A1a}}{\textbf{T1:}\ \forall\varphi.[P(\varphi) \to \Diamond\exists x.\varphi(x)]}}{\textbf{C1:}\ \Diamond\exists x.G(x)}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\forall\varphi.[\neg P(\varphi) \to P(\neg\varphi)]}\,\textbf{A1b} \qquad \dfrac{}{\forall\varphi.[P(\varphi) \to \Box\, P(\varphi)]}\,\textbf{A4}}{\textbf{T2:}\ \forall y.[G(y) \to G\ ess\ y]} \qquad \dfrac{}{P(E)}\,\textbf{A5}}{\textbf{L1:}\ \exists z.G(z) \to \Box\exists x.G(x)}}{\Diamond\exists z.G(z) \to \Diamond\Box\exists x.G(x)} \qquad \dfrac{}{\forall\xi.[\Diamond\Box\xi \to \Box\xi]}\,\textbf{S5}}{\textbf{L2:}\ \Diamond\exists z.G(z) \to \Box\exists x.G(x)}$$

$$\dfrac{\textbf{C1:}\ \Diamond\exists x.G(x) \qquad \textbf{L2:}\ \Diamond\exists z.G(z) \to \Box\exists x.G(x)}{\textbf{T3:}}$$

Natural Deduction Calculus

$$\dfrac{A \lor B \qquad \begin{array}{c}\overline{A}\\ \vdots \\ C\end{array} \qquad \begin{array}{c}\overline{B}\\ \vdots \\ C\end{array}}{C}\ \lor_E \qquad\qquad \dfrac{A \quad B}{A \land B}\ \land_I \qquad\qquad \dfrac{\begin{array}{c}\overline{A}^{\;n}\\ \vdots \\ B\end{array}}{A \longrightarrow B}\ \longrightarrow_I^n$$

$$\dfrac{A}{A \lor B}\ \lor_{I_1} \qquad\qquad \dfrac{A \land B}{A}\ \land_{E_1} \qquad\qquad \dfrac{B}{A \longrightarrow B}\ \longrightarrow_I$$

$$\dfrac{B}{A \lor B}\ \lor_{I_2} \qquad\qquad \dfrac{A \land B}{B}\ \land_{E_2} \qquad\qquad \dfrac{A \quad A \longrightarrow B}{B}\ \longrightarrow_E$$

$$\dfrac{A[\alpha]}{\forall x.A[x]}\ \forall_I \qquad \dfrac{\forall x.A[x]}{A[t]}\ \forall_E \qquad\qquad \dfrac{A[t]}{\exists x.A[x]}\ \exists_I \qquad \dfrac{\exists x.A[x]}{A[\beta]}\ \exists_E$$

$$\neg A \equiv A \longrightarrow \bot \qquad\qquad \color{red}{\dfrac{\neg\neg A}{A}\ \neg\neg_E}$$

$$\frac{\alpha : \boxed{\begin{array}{c}\vdots\\ A\end{array}}}{\Box A}\ \Box_I \qquad\qquad \frac{\Box A}{t : \boxed{\begin{array}{c}A\\ \vdots\end{array}}}\ \Box_E$$

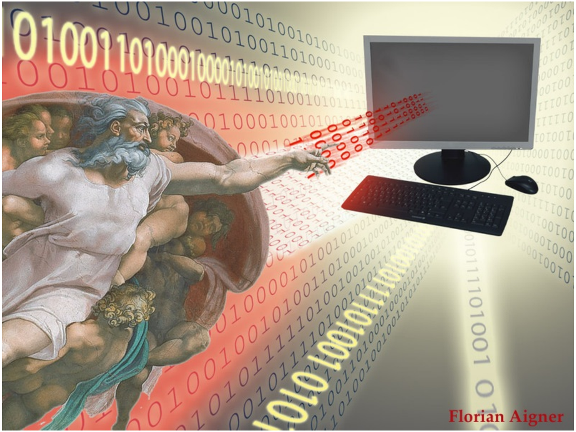
$$\frac{t : \boxed{\begin{array}{c}\vdots\\ A\end{array}}}{\Diamond A}\ \Diamond_I \qquad\qquad \frac{\Diamond A}{\beta : \boxed{\begin{array}{c}A\\ \vdots\end{array}}}\ \Diamond_E$$

$$\Diamond A \equiv \neg A$$

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\textbf{A2}}{\forall\varphi.\forall\psi.[(P(\varphi) \wedge \Box\forall x.[\varphi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]}}{\forall\psi.[(P(\rho) \wedge \Box\forall x.[\rho(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]}\;\forall_E}{(P(\rho) \wedge \Box\forall x.[\rho(x) \rightarrow \neg\rho(x)]) \rightarrow P(\neg\rho)}\;\forall_E}{(P(\rho) \wedge \Box\forall x.[\neg\rho(x)]) \rightarrow P(\neg\rho)} \qquad \dfrac{\dfrac{\textbf{A1a}}{\forall\varphi.[P(\neg\varphi) \rightarrow \neg P(\varphi)]}}{P(\neg\rho) \rightarrow \neg P(\rho)}\;\forall_E}{\dfrac{(P(\rho) \wedge \Box\forall x.[\neg\rho(x)]) \rightarrow \neg P(\rho)}{P(\rho) \rightarrow \Diamond\exists x.\rho(x)}}}{\forall\varphi.[P(\varphi) \rightarrow \Diamond\exists x.\varphi(x)]}\;\forall_I
$$

$$
\dfrac{\dfrac{\textbf{A3}}{P(G)} \qquad \dfrac{\dfrac{\textbf{T1}}{\forall\varphi.[P(\varphi) \rightarrow \Diamond\exists x.\varphi(x)]}}{P(G) \rightarrow \Diamond\exists x.G(x)}\;\forall_E}{\Diamond\exists x.G(x)}\;\rightarrow_E
$$

$$
\dfrac{
\dfrac{\psi(x)^6 \qquad \dfrac{\Pi_2}{\psi(x) \to \Box P(\psi)}}{\Box P(\psi)}\;{\to_E}
\qquad
\dfrac{\dfrac{\dfrac{\dfrac{\Box P(\psi)^7}{P(\psi)}\;\Box_E \qquad \dfrac{\Pi_3}{P(\psi) \to \forall x.(G(x) \to \psi(x))}}{\forall x.(G(x) \to \psi(x))}\;{\to_E}}{\Box\forall x.(G(x) \to \psi(x))}\;\Box_I}{\Box P(\psi) \to \Box\forall x.(G(x) \to \psi(x))}\;{\to_I^7}
}{
\dfrac{\Box\forall x.(G(x) \to \psi(x))}{\psi(x) \to \Box\forall x.(G(x) \to \psi(x))}\;{\to_I^6}
}\;{\to_E}
$$



Part B:

formalization: in classical higher-order logic (HOL)

proof automation: theorem provers Leo-II and Satallax

consistency checking: model finder Nitpick (Nitrox)

Main challenge: No provers for *Higher-order Modal Logic* (HML)

Our solution: Embedding in *Higher-order Classical Logic* (HOL)

Then use existing HOL theorem provers for reasoning in HML

[BenzmüllerPaulson, Logica Universalis, 2013]

Previous empiricial findings:

Embedding of *First-order Modal Logic* in HOL works well

[BenzmüllerOttenRaths, ECAI, 2012]

[Benzmüller, LPAR, 2013]

HML $\quad \varphi, \psi \quad ::= \quad \ldots \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \rightarrow \psi \mid \Box\varphi \mid \Diamond\varphi \mid \forall x\, \varphi \mid \exists x\, \varphi \mid \forall P\, \varphi$

- Kripke style semantics (possible world semantics)

HOL $\quad s, t \quad ::= \quad C \mid x \mid \lambda x s \mid s\, t \mid \neg s \mid s \vee t \mid \forall x\, t$

- meanwhile very well understood
- Henkin semantics vs. standard semantics
- various theorem provers do exists

  interactive: Isabelle/HOL, HOL4, Hol Light, Coq/HOL, PVS, ...

  automated: TPS, LEO-II, Satallax, Nitpick, Isabelle/HOL, ...

HML $\varphi, \psi ::= \ldots \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \rightarrow \psi \mid \Box\varphi \mid \Diamond\varphi \mid \forall x\,\varphi \mid \exists x\,\varphi \mid \forall P\,\varphi$

HOL $s, t ::= C \mid x \mid \lambda x s \mid s\,t \mid \neg s \mid s \vee t \mid \forall x\,t$

HML in HOL: HML formulas φ are mapped to HOL predicates $\varphi_{\iota\rightarrow o}$

$$\begin{array}{lcl}
\neg & = & \lambda\varphi_{\iota\rightarrow o}\lambda s_\iota \neg\varphi s \\
\wedge & = & \lambda\varphi_{\iota\rightarrow o}\lambda\psi_{\iota\rightarrow o}\lambda s_\iota(\varphi s \wedge \psi s) \\
\rightarrow & = & \lambda\varphi_{\iota\rightarrow o}\lambda\psi_{\iota\rightarrow o}\lambda s_\iota(\neg\varphi s \vee \psi s) \\
\Box & = & \lambda\varphi_{\iota\rightarrow o}\lambda s_\iota \forall u_\iota\,(\neg rsu \vee \varphi u) \\
\Diamond & = & \lambda\varphi_{\iota\rightarrow o}\lambda s_\iota \exists u_\iota\,(rsu \wedge \varphi u) \\
\forall & = & \lambda h_{\mu\rightarrow(\iota\rightarrow o)}\lambda s_\iota \forall d_\mu\, hds \\
\exists & = & \lambda h_{\mu\rightarrow(\iota\rightarrow o)}\lambda s_\iota \exists d_\mu\, hds \\
\forall & = & \lambda H_{(\mu\rightarrow(\iota\rightarrow o))\rightarrow(\iota\rightarrow o)}\lambda s_\iota \forall d_\mu\, Hds \\
 & & \\
\text{valid} & = & \lambda\varphi_{\iota\rightarrow o}\forall w_\iota \varphi w
\end{array} \qquad \text{Ax}$$

The equations in Ax are given as axioms to the HOL provers!

Example

HML formula	$\Diamond \exists x G(x)$
HML formula in HOL	valid $(\Diamond \exists x G(x))_{\iota \to o}$
expansion, $\beta\eta$-conversion	$\forall w_\iota (\Diamond \exists x G(x))_{\iota \to o}\, w$
expansion, $\beta\eta$-conversion	$\forall w_\iota \exists u_\iota (rwu \wedge (\exists x G(x))_{\iota \to o} u)$
expansion, $\beta\eta$-conversion	$\forall w_\iota \exists u_\iota (rwu \wedge \exists x Gxu)$

What are we doing?

In order to prove that φ is valid in HML,
–> we instead prove that valid $\varphi_{\iota \to o}$ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.

Proof automation and consistencey checking: Demo!

```
Terminal — bash — 125×32
MacBook-Chris %
MacBook-Chris %
MacBook-Chris % ./call_tptp.sh T3.p

Asking various HOL-ATPs in Miami remotely (thanks to Geoff Sutcliffe)

MacBook-Chris % agsyHOL---1.0 : T3.p ++++++ RESULT: SOT_7L4x_Y - agsyHOL---1.0 says Unknown - CPU = 0.00 WC = 0.02
LEO-II---1.6.0 : T3.p ++++++ RESULT: SOT_E4SCha - LEO-II---1.6.0 says Theorem - CPU = 0.03 WC = 0.09
Satallax---2.7 : T3.p ++++++ RESULT: SOT_kVZ1cB - Satallax---2.7 says Theorem - CPU = 0.00 WC = 0.14
Isabelle---2013 : T3.p ++++++ RESULT: SOT_xa0gEp - Isabelle---2013 says Theorem - CPU = 14.06 WC = 17.73 SolvedBy = auto
TPS---3.120601S1b : T3.p ++++++ RESULT: SOT_R0Egsg - TPS---3.120601S1b says Unknown - CPU = 33.56 WC = 41.57
Nitrox---2013 : T3.p ++++++ RESULT: SOT_WGY1Tx - Nitrox---2013 says Unknown - CPU = 75.55 WC = 49.24

MacBook-Chris %
MacBook-Chris % ./call_tptp.sh Consistency.p

Asking various HOL-ATPs in Miami remotely (thanks to Geoff Sutcliffe)

MacBook-Chris % agsyHOL---1.0 : Consistency.p ++++++ RESULT: SOT_ZtY_7o - agsyHOL---1.0 says Unknown - CPU = 0.00 WC = 0.00
Nitrox---2013 : Consistency.p ++++++ RESULT: SOT_HUz10C - Nitrox---2013 says Satisfiable - CPU = 6.56 WC = 8.50
TPS---3.120601S1b : Consistency.p ++++++ RESULT: SOT_fpJxTM - TPS---3.120601S1b says Unknown - CPU = 43.00 WC = 49.42
Isabelle---2013 : Consistency.p ++++++ RESULT: SOT_6Tpp9i - Isabelle---2013 says Unknown - CPU = 69.96 WC = 72.62
LEO-II---1.6.0 : Consistency.p ++++++ RESULT: SOT_dY10sj - LEO-II---1.6.0 says Timeout - CPU = 90 WC = 89.86
Satallax---2.7 : Consistency.p ++++++ RESULT: SOT_Q9WSLf - Satallax---2.7 says Timeout - CPU = 90 WC = 90.50

MacBook-Chris %
```

Provers are called remotely in Miami — no local installation needed!

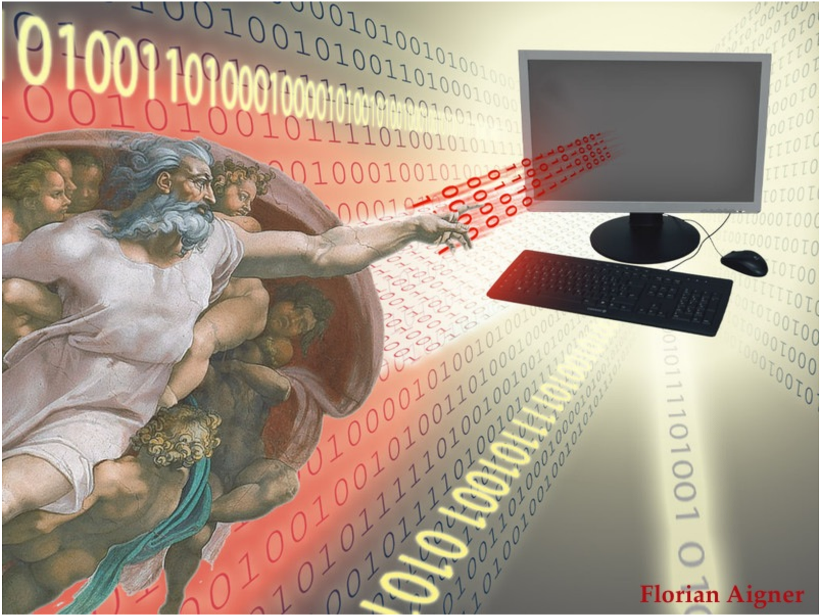
- Goal: verification of the natural deduction proof
  - Step-by-step formalization
  - Almost no automation (intentionally!)
- Interesting facts to note:
  - Embedding is transparent to the user
  - Embedding gives labeled calculus for free

todo

- K sufficient for T1, C1 and T2
- S5 not needed for T3
- KB sufficient for T3
- A simpler new proof of C1
- Gödel's original axioms (without conjunct $\phi(x)$ in D2) are inconsistent
- Scott's axioms are consistent
- For T1, only half of A1 (A1a) is needed
- For T2, the other half (A1b) is needed

Conclusion

- Infra-structure for reasoning with modal logic using existing proof assistants and higher-order automated theorem provers


Florian Aigner