The Inconsistency in Gödel's Ontological Argument — A Success Story for AI in Metaphysics —

IJCAI-16 New York City, USA

Freie Universität Berlin

ANU THE AUSTRALIAN NATIONAL UNIVERSITY

Christoph Benzmüller[1] and Bruno Woltzenlogel Paleo

[1]Supported by the German Research Foundation (grant BE2501/12-1,2)

Motivation

Vision of Leibniz (1646–1716): *Calculemus!*

Quo facto, quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos Computistas. Sufficiet enim calamos in manus sumere sedereque ad abacos, et sibi mutuo …dicere: calculemus. (Leibniz, 1684)

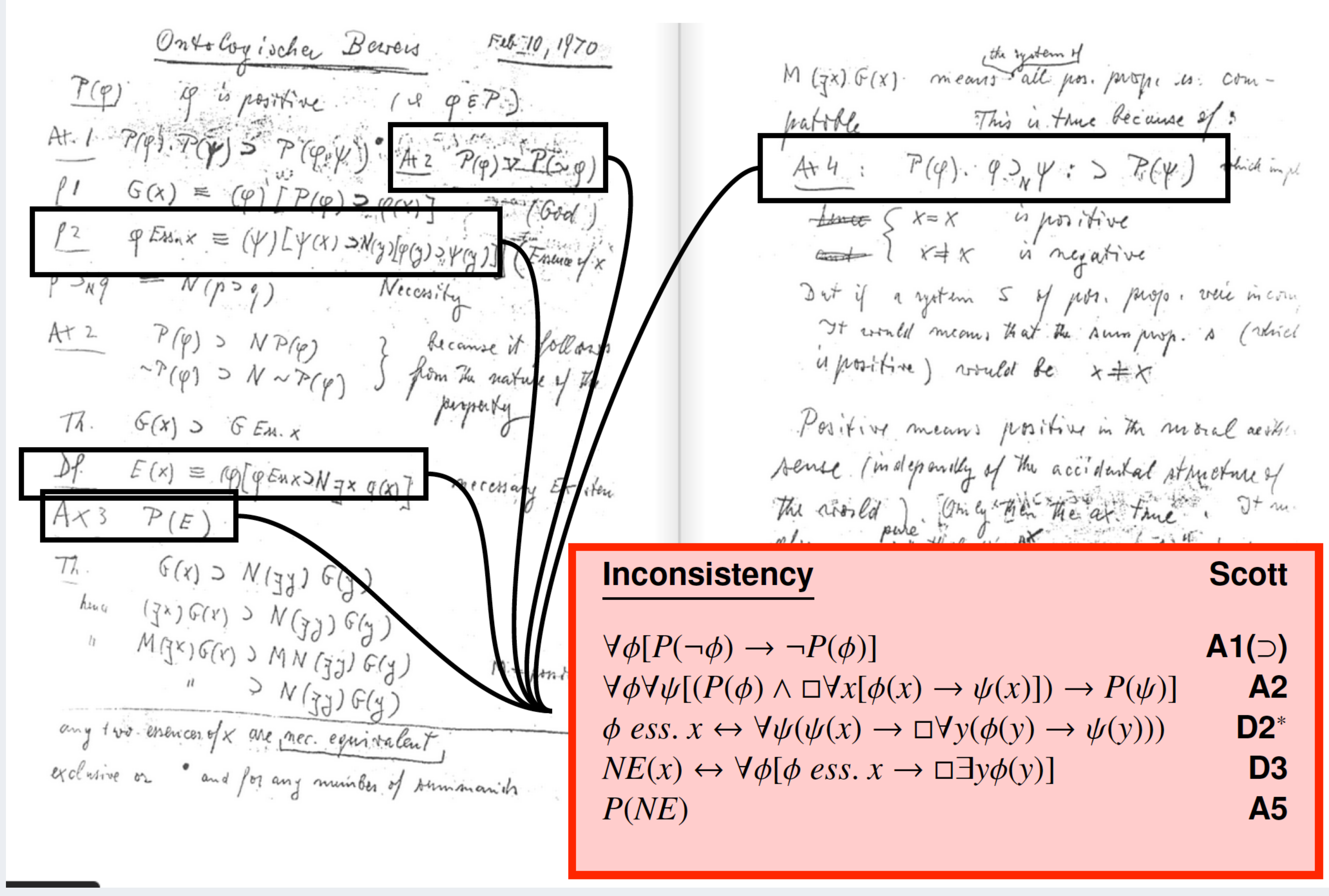
If controversies were to arise, there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and to say to each other …: Let us calculate.

(Translation by Russell)

Required:
characteristica universalis and **calculus ratiocinator**

Application: Gödel's Ontological Argument

Gödel's Manuscript: Identifying the Inconsistent Axioms



Inconsistency	**Scott**
$\forall\phi[P(\neg\phi) \to \neg P(\phi)]$	A1($\supset$)
$\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \to \psi(x)]) \to P(\psi)]$	A2
$\phi\ ess.\ x \leftrightarrow \forall\psi(\psi(x) \to \Box\forall y(\phi(y) \to \psi(y)))$	D2*
$NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \to \Box\exists y\phi(y)]$	D3
$P(NE)$	A5

Scott's and Gödel's Versions

Scott's Version of Gödel's Axioms, Definitions and Theorems

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

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

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \to \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \to \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists x G(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \to \Box P(\phi)]$

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

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \to G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \to \Box\exists y\phi(y)]$

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

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

Difference to Gödel (who omits this conjunct)

Approach: Semantic Embedding

HOL as a Universal (Meta-)Logic via Semantic Embeddings

HOL

Logic L Syntax ⟷ Logic L Semantics

Examples for L we have already studied:
Modal Logics, Conditional Logics, Intuitionistic Logics, Access Control Logics, Nominal Logics, Multivalued Logics (SIXTEEN), Logics based on Neighborhood Semantics, (Mathematical) Fuzzy Logics, Paraconsistent Logics, Free Logic …

Works also for (first-order & higher-order) quantifiers

Standard Translation of Modal Logic

Embedding HOML in HOL

Example

HOML formula	$\Diamond\exists x G(x)$
HOML formula in HOL	valid $(\Diamond\exists x G(x))_{\mu\to o}$
expansion	$(\lambda\varphi\forall w_\mu\varphi\, w)(\Diamond\exists x G(x))_{\mu\to o}$
$\beta\eta$-normalisation	$\forall w_\mu((\Diamond\exists x G(x))_{\mu\to o}\, w)$
expansion	$\forall w_\mu(((\lambda\varphi_{\mu\to o}\lambda w_\mu\exists u_\mu(rwu \wedge \varphi u))\exists x G(x))_{\mu\to o}\, w)$
$\beta\eta$-normalisation	$\forall w_\mu\exists u_\mu(rwu \wedge (\exists x G(x))_{\mu\to o} u)$
syntactic sugar	$\forall w_\mu\exists u_\mu(rwu \wedge (\exists(\lambda x G(x)))_{\mu\to o} u)$
expansion	$\forall w_\mu\exists u_\mu(rwu \wedge ((\lambda h_{\gamma\to(\mu\to o)}\lambda w_\mu\exists d_\gamma\, hdw)(\lambda x G(x)))_{\mu\to o} u)$
$\beta\eta$-normalisation	$\forall w_\mu\exists u_\mu(rwu \wedge \exists x Gxu)$

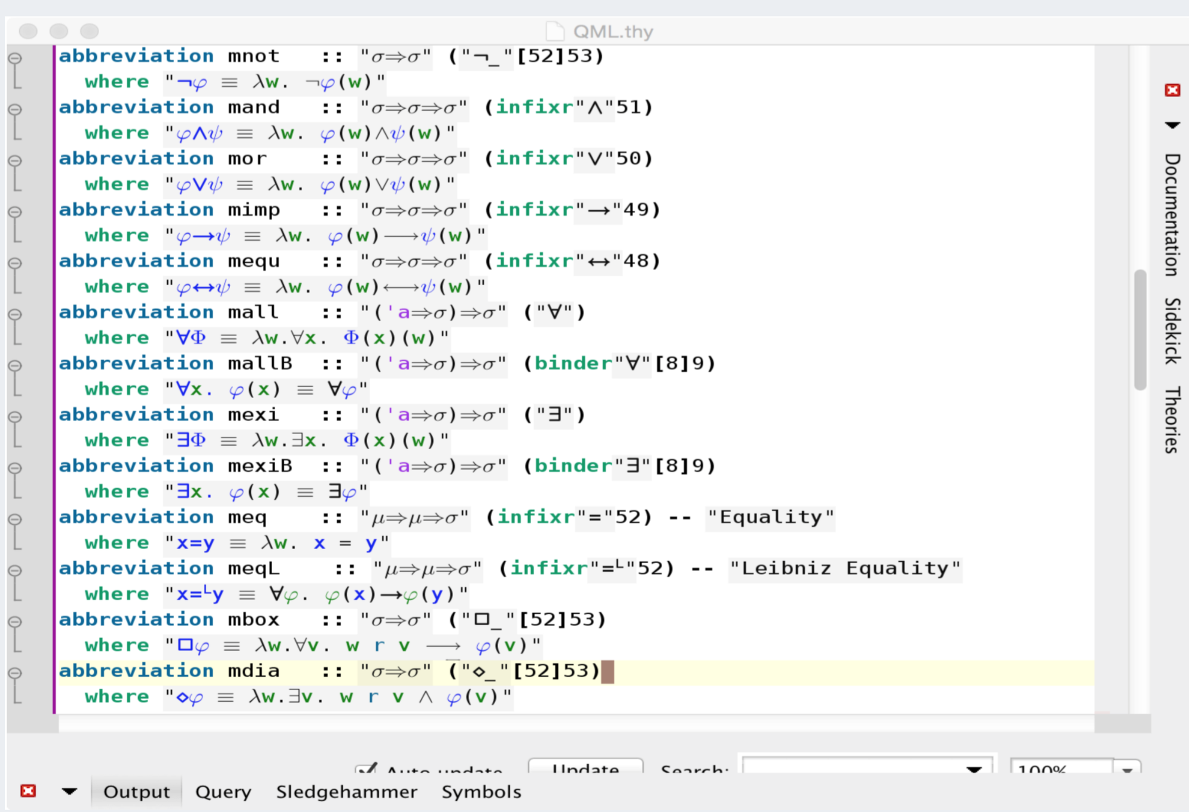
Expansion: user or prover may flexibly choose expansion depth

What are we doing?

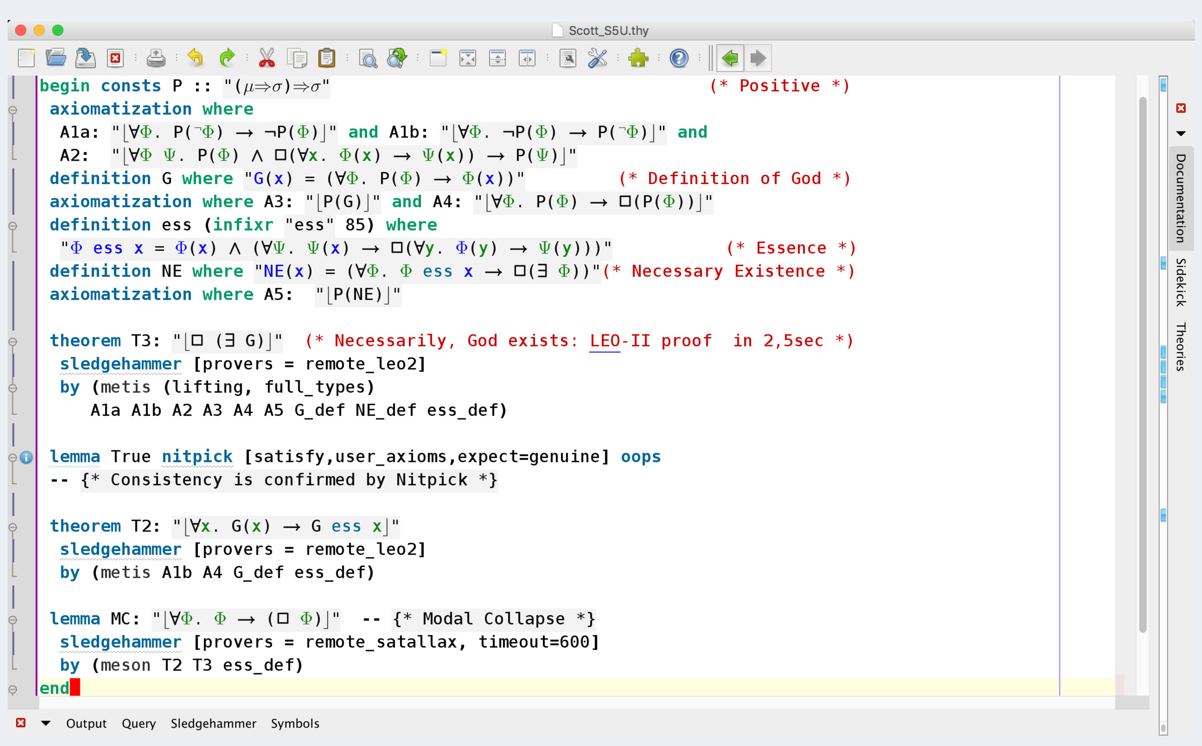
In order to prove that φ is valid in HOML,
--> we instead prove that valid $\varphi_{\mu\to o}$ can be derived from Ax in HOL.

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

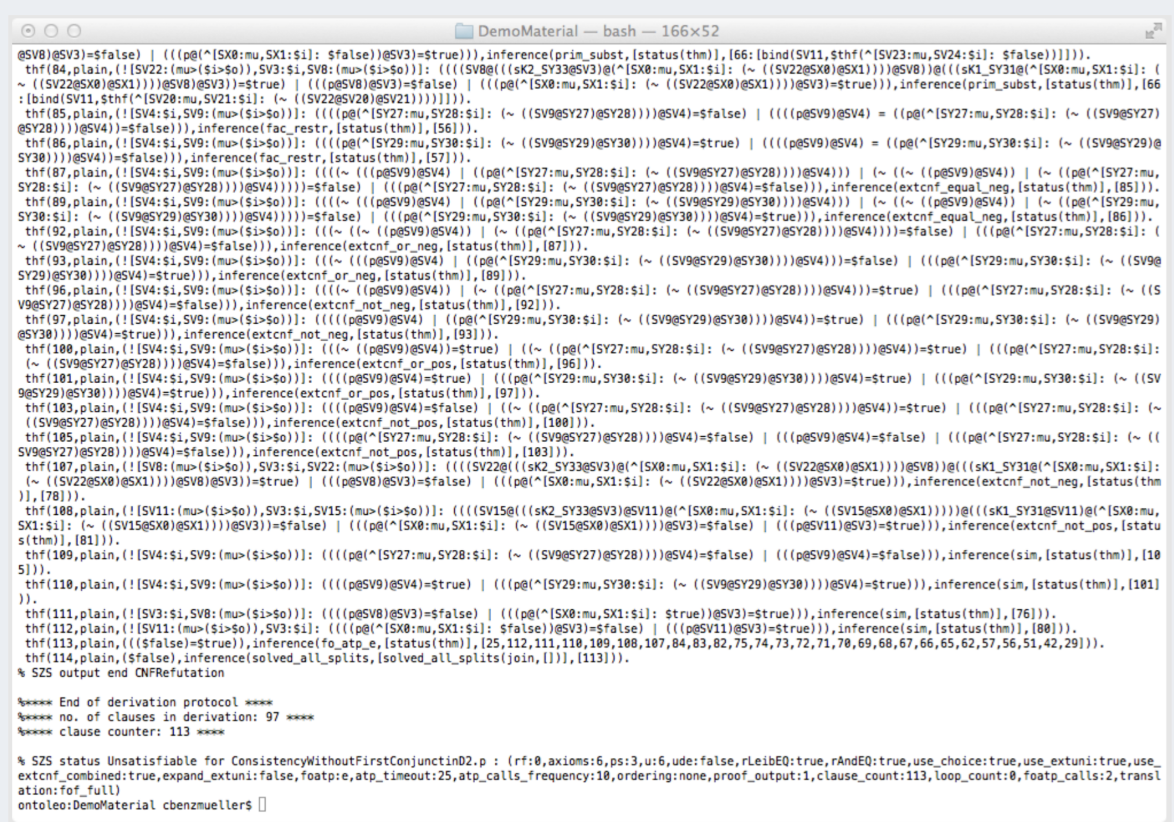
Modal Logic in Isabelle



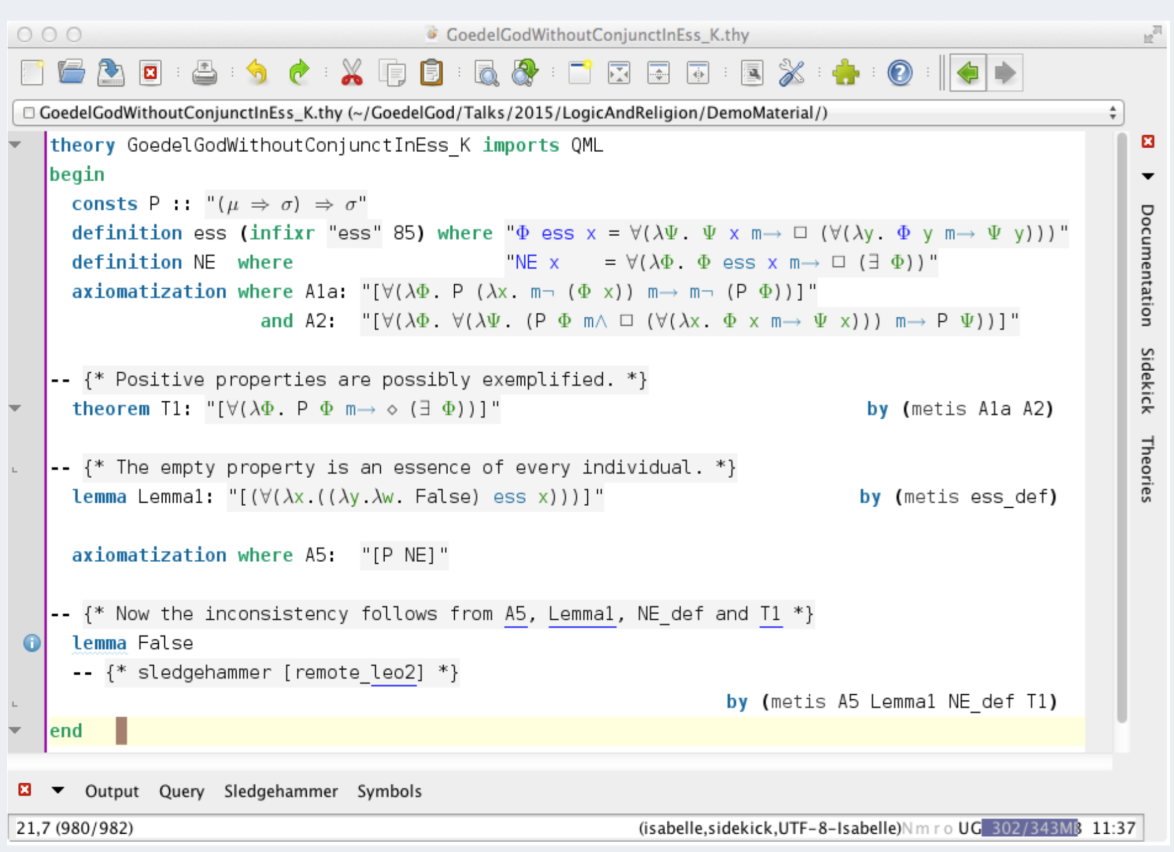
Scott's Version in Isabelle



LEO-II's Refutation of Gödel's Axioms



Refutation Reconstruction in Isabelle



Explaining the Inconsistency

Def. D2'		$\phi\ ess.\ x \leftrightarrow \cancel{\phi(x) \wedge}\ \forall\psi(\psi(x) \to \Box\forall y(\phi(y) \to \psi(y)))$
Lemma 1	The empty property is an essence of every entity.	$\forall x(\emptyset\ ess.\ x)$
Theorem 1	Positive Properties are possibly exemplified.	$\forall\phi[P(\phi) \to \Diamond\exists x\phi(x)]$
Axiom A5		$P(NE)$
▸	by T1, A5:	$\Diamond\exists x[NE(x)]$
Def. D3		$NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \to \Box\exists y\phi(y)]$
▸		$\Diamond\exists x[\forall\phi[\phi\ ess.\ x \to \Box\exists y\phi(y)]]$
▸		$\Diamond\exists x[\emptyset\ ess.\ x \to \Box\exists y[\emptyset(y)]]$
▸	by L1	$\Diamond\exists x[\top \to \Box\exists y[\emptyset(y)]]$
▸	by def. of $\emptyset$	$\Diamond\exists x[\top \to \Box\bot]$
▸		$\Diamond\exists x[\Box\bot]$
▸		$\Diamond\Box\bot$
Inconsistency		$\bot$

Summary of Results

- Inconsistency of Gödel's original axioms has now been verified
- Reason for inconsistency is finally well understood and explained
- Reconstruction of the refutation in Isabelle led to various user interface and performence improvements for the embedding of modal logic in Isabelle
- Higher-order interactive and automated reasoning technology is ready for applications in philosophy