

Automating Emendations of the Ontological Argument in Intensional Higher-Order Modal Logic

David Fuenmayor¹ and Christoph Benzmüller^{2,1}

¹ Freie Universität Berlin, Germany

² University of Luxembourg, Luxembourg

Abstract. A shallow semantic embedding of an intensional higher-order modal logic (IHOML) in Isabelle/HOL is presented. IHOML draws on Montague/Gallin intensional logics and has been introduced by Melvin Fitting in his textbook *Types, Tableaus and Gödel's God* in order to discuss his emendation of Gödel's ontological argument for the existence of God. Utilizing IHOML, the most interesting parts of Fitting's textbook are formalized, automated and verified in the Isabelle/HOL proof assistant. A particular focus thereby is on three variants of the ontological argument which avoid the modal collapse, which is a strongly criticized side-effect in Gödel's resp. Scott's original work.

Keywords: Automated Theorem Proving, Computational Metaphysics, Higher-Order Logic, Intensional Logic, Isabelle, Modal Logic, Ontological Argument, Semantic Embedding

1 Introduction

The first part of this paper introduces a shallow semantic embedding of an intensional higher-order modal logic (IHOML) in classical higher-order logic (Isabelle/HOL³). IHOML, as introduced by Fitting [15], is a modification of the intensional logic originally developed by Montague and later expanded by Gallin [18] by building upon Church's type theory and Kripke's possible-world semantics. Our approach builds on previous work on the semantic embedding of multi-modal logics with quantification [6], which we expand here to allow for actualist quantification, intensional terms and their related operations. From an AI perspective we contribute a highly flexible framework for automated reasoning in intensional and modal logic. IHOML, which has not been automated before, has several applications, e.g. towards the deep semantic analysis of natural language rational arguments as envisioned in the new DFG Schwerpunktprogramm RATIO (SPP 1999).

³ In this paper we work with the Isabelle/HOL proof assistant [22], which explains the chosen abbreviation. Generally, however, the work presented here can be mapped to any other system implementing Church's simple type theory [13].

In the second part, we present an exemplary, non-trivial application of this reasoning infrastructure: A study on *computational metaphysics*⁴, the computer-formalization and critical assessment of Gödel’s [19] (resp. Dana Scott’s [25]) modern variant of the ontological argument and two of its proposed emendations as discussed in [15]. Gödel’s ontological argument is amongst the most discussed formal proofs in modern literature. Several authors (e.g. [3, 2, 11, 20, 15]) have proposed emendations with the aim of retaining its essential result (the necessary existence of God) while at the same time avoiding the *modal collapse* (whatever is the case is so necessarily) [26, 27]. The modal collapse is an undesirable side-effect of the axioms postulated by Gödel (resp. Scott). It essentially states that there are no contingent truths and everything is determined.

Related work⁵ has formalized several of these variants on the computer and verified or falsified them. For example, Gödel’s axiom’s system has been shown inconsistent [9, 10], while Scott’s version has been verified [8]. Further experiments, contributing amongst others to the clarification of a related debate regarding the redundancy of some axioms in Anderson’s emendation, are presented and discussed in [7]. The enabling technique in these case studies has been shallow semantic embeddings of *extensional* higher-order modal logics in classical higher-order logic (see [6, 4] and the references therein).⁶

In contrast to the related work, Fitting’s variant is based on *intensional* higher-order modal logic. Our experiments confirm that Fitting’s argument, as presented in his textbook [15], is valid and that it avoids the modal collapse as intended. Due to lack of space, we refer the reader to our (computer-verified) paper [17] for further results. That paper has been written directly in the Isabelle/HOL proof assistant and requires some familiarity with this system and with Fitting’s textbook.

The work presented here originates from the *Computational Metaphysics* lecture course held at the FU Berlin in Summer 2016 [28].

2 Embedding of Intensional Higher-Order Modal Logic

2.1 Type Declarations

Since IHOML and Isabelle/HOL are both typed languages, we introduce a type-mapping between them. We follow as closely as possible the syntax given by

⁴ This term was originally coined by Fitelson and Zalta in [14] and describes an emerging, interdisciplinary field aiming at the rigorous formalization and deep logical assessment of philosophical arguments in an automated reasoning environment.

⁵ More loosely related work studied Anselm’s older, non-modal version of the ontological argument directly in Prover9 [23] and PVS [24].

⁶ In contrast to deep semantic embeddings, where the embedded logic is presented as an abstract datatype, our shallow semantic embeddings avoid inductive definitions and maximize the reuse of logical operations from the meta-level. In particular, tedious new binding mechanisms are avoided in our approach.

Fitting ([15] p. 86), according to which, for any extensional type τ , $\uparrow\tau$ becomes its corresponding intensional type. For instance, a set of (red) objects has the extensional type $\langle e \rangle$, whereas the concept ‘red’ has intensional type $\uparrow\langle e \rangle$.

typeddecl e — type for entities
typeddecl w — type for possible worlds
type-synonym $wo = (w \Rightarrow bool)$ — type for world-dependent formulas

Aliases for some common complex types (predicates and relations).

type-synonym $ie = (w \Rightarrow e)$ ($\uparrow e$) — individual concepts (map worlds to objects)
type-synonym $se = (e \Rightarrow bool)$ ($\langle e \rangle$) — (extensional) sets
type-synonym $ise = (e \Rightarrow wo)$ ($\uparrow\langle e \rangle$) — (intensional predicative) concepts
type-synonym $sise = (\uparrow\langle e \rangle \Rightarrow bool)$ ($\langle \uparrow\langle e \rangle \rangle$) — sets of concepts
type-synonym $isise = (\uparrow\langle e \rangle \Rightarrow wo)$ ($\uparrow\langle \uparrow\langle e \rangle \rangle$) — 2-order concepts
type-synonym $see = (e \Rightarrow e \Rightarrow bool)$ ($\langle e, e \rangle$) — (extensional) relations
type-synonym $isee = (e \Rightarrow e \Rightarrow wo)$ ($\uparrow\langle e, e \rangle$) — (intensional) relational concepts

2.2 Logical Constants as Truth-Sets

We embed modal operators as sets of worlds satisfying a corresponding formula.

abbreviation $mand :: wo \Rightarrow wo \Rightarrow wo$ (**infix** \wedge) **where** $\varphi \wedge \psi \equiv \lambda w. (\varphi w) \wedge (\psi w)$
abbreviation $mor :: wo \Rightarrow wo \Rightarrow wo$ (**infix** \vee) **where** $\varphi \vee \psi \equiv \lambda w. (\varphi w) \vee (\psi w)$
abbreviation $mimp :: wo \Rightarrow wo \Rightarrow wo$ (**infix** \rightarrow) **where** $\varphi \rightarrow \psi \equiv \lambda w. (\varphi w) \rightarrow (\psi w)$
abbreviation $mequ :: wo \Rightarrow wo \Rightarrow wo$ (**infix** \leftrightarrow) **where** $\varphi \leftrightarrow \psi \equiv \lambda w. (\varphi w) \leftrightarrow (\psi w)$
abbreviation $mnot :: wo \Rightarrow wo$ (**infix** \neg) **where** $\neg \varphi \equiv \lambda w. \neg(\varphi w)$
abbreviation $mnegpred :: \uparrow\langle e \rangle \Rightarrow \uparrow\langle e \rangle$ (**infix** \neg) **where** $\neg \Phi \equiv \lambda x. \lambda w. \neg(\Phi x w)$

Possibilist quantifiers are embedded as follows.⁷

abbreviation $mforall :: ('t \Rightarrow wo) \Rightarrow wo$ (\forall) **where** $\forall \Phi \equiv \lambda w. \forall x. (\Phi x w)$
abbreviation $mexists :: ('t \Rightarrow wo) \Rightarrow wo$ (\exists) **where** $\exists \Phi \equiv \lambda w. \exists x. (\Phi x w)$

The *actualizedAt* predicate is used to additionally embed *actualist* quantifiers by restricting the domain of quantification at every possible world. This standard technique has been referred to as *existence relativization* ([16], p. 106), highlighting the fact that this predicate can be seen as a kind of meta-logical ‘existence predicate’ telling us which individuals *actually* exist at a given world. This meta-logical concept does not appear in our object language.

consts *Actualized* :: $\uparrow\langle e \rangle$ (**infix** *actualizedAt*)
abbreviation $mforallAct :: \uparrow\langle \uparrow\langle e \rangle \rangle$ (\forall^A) — actualist variants use superscript
where $\forall^A \Phi \equiv \lambda w. \forall x. (x \text{ actualizedAt } w) \rightarrow (\Phi x w)$
abbreviation $mexistsAct :: \uparrow\langle \uparrow\langle e \rangle \rangle$ (\exists^A)
where $\exists^A \Phi \equiv \lambda w. \exists x. (x \text{ actualizedAt } w) \wedge (\Phi x w)$

Frame’s accessibility relation and modal operators.

⁷ Possibilist and actualist quantification can be seen as the semantic counterparts of the concepts of possibilism and actualism in the metaphysics of modality. They relate to natural-language expressions such as ‘there is’, ‘exists’, ‘is actual’, etc.

consts $aRel::w \Rightarrow w \Rightarrow bool$ (**infix** r)
abbreviation $mbox::wo \Rightarrow wo$ ($\Box-$) **where** $\Box\varphi \equiv \lambda w. \forall v. (w \ r \ v) \longrightarrow (\varphi \ v)$
abbreviation $mdia::wo \Rightarrow wo$ ($\Diamond-$) **where** $\Diamond\varphi \equiv \lambda w. \exists v. (w \ r \ v) \wedge (\varphi \ v)$

2.3 Equality

abbreviation $meq:: 't \Rightarrow 't \Rightarrow wo$ (**infix** \approx) — standard equality (for all types)
where $x \approx y \equiv \lambda w. x = y$
abbreviation $meqC:: \uparrow\langle e, \uparrow e \rangle$ (**infix** \approx^C) — equality for individual concepts
where $x \approx^C y \equiv \lambda w. \forall v. (x \ v) = (y \ v)$
abbreviation $meqL:: \uparrow\langle e, e \rangle$ (**infix** \approx^L) — Leibniz equality for individuals
where $x \approx^L y \equiv \lambda w. \forall \varphi. (\varphi \ x \ w) \longrightarrow (\varphi \ y \ w)$

2.4 Extension-of Operator

According to Fitting's semantics ([15], pp. 92-4), \downarrow is an unary operator applying only to intensional terms. A term of the form $\downarrow\alpha$ designates the extension of the intensional object designated by α , at some *given* world. For instance, suppose we take possible worlds as persons, we can therefore think of the concept 'red' as a function that maps each person to the set of objects that person classifies as red (its extension). We can further state that the intensional term r of type $\uparrow\langle e \rangle$ designates the concept 'red'. As can be seen, intensional terms in IHOML designate functions on possible worlds and they always do it *rigidly*. We will sometimes refer to an intensional object explicitly as 'rigid', implying that its (rigidly) designated function has the same extension in all possible worlds.⁸

Terms of the form $\downarrow\alpha$ are called *relativized* (extensional) terms; they are always derived from intensional terms and their type is extensional (in the color example $\downarrow r$ would be of type $\langle e \rangle$). Relativized terms may vary their denotation from world to world of a model, because the *extension of* an intensional term can change from world to world, i.e. they are non-rigid.

In our Isabelle/HOL embedding, we had to follow a slightly different approach; we model \downarrow as a predicate applying to formulas of the form $\Phi(\downarrow\alpha_1, \dots, \alpha_n)$. For instance, the formula $Q(\downarrow a_1)^w$ (evaluated at world w) is modeled as $\downarrow(Q, a_1)^w$, or $(Q \downarrow a_1)^w$ using infix notation, which gets further translated into $Q(a_1(w))^w$.

(a) Predicate φ takes as argument a relativized term derived from an (intensional) individual concept of type $\uparrow e$.

abbreviation $extIndArg::\uparrow\langle e \rangle \Rightarrow \uparrow e \Rightarrow wo$ (**infix** \downarrow) **where** $\varphi \downarrow c \equiv \lambda w. \varphi \ (c \ w) \ w$

(b) A variant of (a) for terms derived from predicates (types of form $\uparrow\langle t \rangle$).

abbreviation $extPredArg::('t \Rightarrow bool) \Rightarrow wo \Rightarrow ('t \Rightarrow wo) \Rightarrow wo$ (**infix** \downarrow)
where $\varphi \downarrow P \equiv \lambda w. \varphi \ (\lambda x. P \ x \ w) \ w$

⁸ The notion of *rigid designation* was introduced by Kripke in [21], where he discusses its many interesting ramifications in logic and the philosophy of language.

2.5 Verifying the Embedding

The above definitions introduce modal logic K with possibilist and actualist quantifiers, as evidenced by the following tests.⁹

abbreviation $valid::wo \Rightarrow bool$ ($\lfloor \cdot \rfloor$) **where** $\lfloor \psi \rfloor \equiv \forall w. (\psi \ w)$ — modal validity
lemma $K: \lfloor (\Box(\varphi \rightarrow \psi)) \rightarrow (\Box\varphi \rightarrow \Box\psi) \rfloor$ **by** *simp* — verifying K principle
lemma $NEC: \lfloor \varphi \rfloor \Rightarrow \lfloor \Box\varphi \rfloor$ **by** *simp* — verifying *necessitation* rule

Local consequence implies global consequence (not the other way round).¹⁰

lemma *localImpGlobalCons*: $\lfloor \varphi \rightarrow \xi \rfloor \Rightarrow \lfloor \varphi \rfloor \rightarrow \lfloor \xi \rfloor$ **by** *simp*
lemma $\lfloor \varphi \rfloor \rightarrow \lfloor \xi \rfloor \Rightarrow \lfloor \varphi \rightarrow \xi \rfloor$ **nitpick oops** — countersatisfiable

(Converse-)Barcan formulas are satisfied for possibilist, but not for actualist, quantification.

lemma $\lfloor (\forall x. \Box(\varphi \ x)) \rightarrow \Box(\forall x. (\varphi \ x)) \rfloor$ **by** *simp*
lemma $\lfloor \Box(\forall x. (\varphi \ x)) \rightarrow (\forall x. \Box(\varphi \ x)) \rfloor$ **by** *simp*
lemma $\lfloor (\forall^A x. \Box(\varphi \ x)) \rightarrow \Box(\forall^A x. (\varphi \ x)) \rfloor$ **nitpick oops** — countersatisfiable
lemma $\lfloor \Box(\forall^A x. (\varphi \ x)) \rightarrow (\forall^A x. \Box(\varphi \ x)) \rfloor$ **nitpick oops** — countersatisfiable

β -redex is valid for non-relativized (intensional or extensional) terms.

lemma $\lfloor (\lambda\alpha. \varphi \ \alpha) (\tau::\uparrow e) \leftrightarrow (\varphi \ \tau) \rfloor$ **by** *simp*
lemma $\lfloor (\lambda\alpha. \varphi \ \alpha) (\tau::e) \leftrightarrow (\varphi \ \tau) \rfloor$ **by** *simp*
lemma $\lfloor (\lambda\alpha. \Box\varphi \ \alpha) (\tau::\uparrow e) \leftrightarrow (\Box\varphi \ \tau) \rfloor$ **by** *simp*
lemma $\lfloor (\lambda\alpha. \Box\varphi \ \alpha) (\tau::e) \leftrightarrow (\Box\varphi \ \tau) \rfloor$ **by** *simp*

β -redex is valid for relativized terms as long as no modal operators occur.

lemma $\lfloor (\lambda\alpha. \varphi \ \alpha) \downarrow (\tau::\uparrow e) \leftrightarrow (\varphi \ \downarrow\tau) \rfloor$ **by** *simp*
lemma $\lfloor (\lambda\alpha. \Box\varphi \ \alpha) \downarrow (\tau::\uparrow e) \leftrightarrow (\Box\varphi \ \downarrow\tau) \rfloor$ **nitpick oops** — countersatisfiable

Modal collapse is countersatisfiable.

lemma $\lfloor \varphi \rightarrow \Box\varphi \rfloor$ **nitpick oops** — countersatisfiable

2.6 Stability, Rigid Designation, *De Dicto* and *De Re*

Intensional terms are trivially rigid. This predicate tests whether an intensional predicate is ‘rigid’ in the sense of denoting a world-independent function.

abbreviation $rigid::('t \Rightarrow wo) \Rightarrow wo$ **where** $rigid \ \tau \equiv (\lambda\beta. \Box((\lambda z. \beta \approx z) \downarrow \tau)) \downarrow \tau$

⁹ We prove theorems in Isabelle by using the keyword ‘by’ followed by the name of a proof method. Some methods used here are: *simp* (term rewriting), *blast* (tableaus), *meson* (model elimination), *metis* (ordered resolution and paramodulation), *auto* (classical reasoning and term rewriting) and *force* (exhaustive search trying different tools). In our computer-formalization and assessment of Fitting’s textbook [17], we provide further evidence that our embedded logic works as intended by verifying the book’s theorems and examples.

¹⁰ We utilize here (counter-)model finder *Nitpick* [12] for the first time. For the conjectured lemma, *Nitpick* finds a countermodel (not shown here), i.e. a model satisfying all the axioms which falsifies the given formula.

Following definitions are called ‘stability conditions’ by Fitting ([15], p. 124).

abbreviation $\text{stability}A::('t \Rightarrow wo) \Rightarrow wo$ **where** $\text{stability}A \tau \equiv \forall \alpha. (\tau \alpha) \rightarrow \Box(\tau \alpha)$

abbreviation $\text{stability}B::('t \Rightarrow wo) \Rightarrow wo$ **where** $\text{stability}B \tau \equiv \forall \alpha. \Diamond(\tau \alpha) \rightarrow (\tau \alpha)$

We prove them equivalent in *S5* logic (using *Sahlqvist correspondence*).

lemma *equivalence* $aRel \Rightarrow \lfloor \text{stability}A (\tau::\uparrow\langle e \rangle) \rfloor \rightarrow \lfloor \text{stability}B \tau \rfloor$ **by** *blast*

lemma *equivalence* $aRel \Rightarrow \lfloor \text{stability}B (\tau::\uparrow\langle e \rangle) \rfloor \rightarrow \lfloor \text{stability}A \tau \rfloor$ **by** *blast*

A term is ‘rigid’ if and only if it satisfies the stability conditions.

lemma $\lfloor \text{rigid} (\tau::\uparrow\langle e \rangle) \rfloor \longleftrightarrow \lfloor (\text{stability}A \tau \wedge \text{stability}B \tau) \rfloor$ **by** *meson*

lemma $\lfloor \text{rigid} (\tau::\uparrow\langle e \rangle) \rfloor \longleftrightarrow \lfloor (\text{stability}A \tau \wedge \text{stability}B \tau) \rfloor$ **by** *meson*

De re is equivalent to *de dicto* for non-relativized terms.¹¹

lemma $\lfloor \forall \alpha. ((\lambda \beta. \Box(\alpha \beta)) (\tau::\langle e \rangle)) \rfloor \leftrightarrow \lfloor \Box((\lambda \beta. (\alpha \beta)) \tau) \rfloor$ **by** *simp*

lemma $\lfloor \forall \alpha. ((\lambda \beta. \Box(\alpha \beta)) (\tau::\uparrow\langle e \rangle)) \rfloor \leftrightarrow \lfloor \Box((\lambda \beta. (\alpha \beta)) \tau) \rfloor$ **by** *simp*

De re is not equivalent to *de dicto* for relativized terms.

lemma $\lfloor \forall \alpha. ((\lambda \beta. \Box(\alpha \beta)) \downarrow(\tau::\uparrow\langle e \rangle)) \rfloor \leftrightarrow \lfloor \Box((\lambda \beta. (\alpha \beta)) \downarrow\tau) \rfloor$

nitpick $[card\ e=1, card\ w=2]$ **oops** — countersatisfiable

2.7 Useful Definitions for the Axiomatization of Further Logics

The best-known normal logics (*K4*, *K5*, *KB*, *K45*, *KB5*, *D*, *D4*, *D5*, *D45*, ...) can be obtained by combinations of the following axioms:

abbreviation *T* **where** $T \equiv \forall \varphi. \Box \varphi \rightarrow \varphi$

abbreviation *B* **where** $B \equiv \forall \varphi. \varphi \rightarrow \Box \Diamond \varphi$

abbreviation *D* **where** $D \equiv \forall \varphi. \Box \varphi \rightarrow \Diamond \varphi$

abbreviation *IV* **where** $IV \equiv \forall \varphi. \Box \varphi \rightarrow \Box \Box \varphi$

abbreviation *V* **where** $V \equiv \forall \varphi. \Diamond \varphi \rightarrow \Box \Diamond \varphi$

Instead of postulating combinations of the above axioms we make use of the well-known *Sahlqvist correspondence*, which links axioms to constraints on a model’s accessibility relation. We show that reflexivity, symmetry, seriality, transitivity and euclideaness imply axioms *T*, *B*, *D*, *IV*, *V* respectively.¹²

lemma *reflexive* $aRel \Rightarrow \lfloor T \rfloor$ **by** *blast*

lemma *symmetric* $aRel \Rightarrow \lfloor B \rfloor$ **by** *blast*

lemma *serial* $aRel \Rightarrow \lfloor D \rfloor$ **by** *blast*

lemma *transitive* $aRel \Rightarrow \lfloor IV \rfloor$ **by** *blast*

lemma *euclidean* $aRel \Rightarrow \lfloor V \rfloor$ **by** *blast*

lemma *preorder* $aRel \Rightarrow \lfloor T \rfloor \wedge \lfloor IV \rfloor$ **by** *blast* — S4: reflexive + transitive

lemma *equivalence* $aRel \Rightarrow \lfloor T \rfloor \wedge \lfloor V \rfloor$ **by** *blast* — S5: preorder + symmetric

¹¹ The *de dicto/de re* distinction is used regularly in the philosophy of language for disambiguation of sentences involving intensional contexts.

¹² Implication can also be proven in the reverse direction (which is not needed for our purposes). Using these definitions, we can derive axioms for the most common modal logics (see also [5]). Thereby we are free to use either the semantic constraints or the related *Sahlqvist* axioms. Here we provide both versions. In what follows we use the semantic constraints for improved performance.

3 Gödel's Ontological Argument

3.1 Part I - God's Existence is Possible

Gödel's particular version of the argument is a direct descendant of that of Leibniz, which in turn derives from one of Descartes. His argument relies on proving (T1) 'Positive properties are possibly instantiated', which together with (T2) 'God is a positive property' directly implies the conclusion. In order to prove T1, Gödel assumes (A2) 'Any property entailed by a positive property is positive'. As we will see, the success of this argumentation depends on how we formalize our notion of entailment.

abbreviation $\text{Entails}::\uparrow\langle\uparrow\langle e\rangle, \uparrow\langle e\rangle\rangle$ (infix \Rightarrow) **where** $X \Rightarrow Y \equiv \Box(\forall^A z. X\ z \rightarrow Y\ z)$
lemma $\lfloor(\lambda x\ w. x \neq x) \Rightarrow \chi\rfloor$ **by** *simp* — an impossible property entails anything
lemma $\lfloor\neg(\varphi \Rightarrow \chi) \rightarrow \Diamond \exists^A \varphi\rfloor$ **by** *auto* — possible instantiation of φ implicit

The definition of property entailment introduced by Gödel can be criticized on the grounds that it lacks some notion of relevance and is therefore exposed to the paradoxes of material implication. In particular, when we assert that property A does not entail property B , we implicitly assume that A is possibly instantiated. Conversely, an impossible property (like being a round square) entails any property (like being a triangle). It is precisely by virtue of these paradoxes that Gödel manages to prove T1.¹³

consts $\text{Positiveness}::\uparrow\langle\uparrow\langle e\rangle\rangle$ (\mathcal{P}) — positiveness applies to intensional predicates
abbreviation $\text{Existence}::\uparrow\langle e\rangle$ ($E!$) — object-language existence predicate
where $E!\ x \equiv \lambda w. (\exists^A y. y \approx x)\ w$

Gödel's axioms for the first part essentially say that (A1) either a property or its negation must be positive, (A2) positive properties are closed under entailment and (A3) also closed under conjunction.

abbreviation $\text{appliesToPositiveProps}::\uparrow\langle\uparrow\langle\uparrow\langle e\rangle\rangle\rangle$ (pos) **where**
 $\text{pos}\ Z \equiv \forall X. Z\ X \rightarrow \mathcal{P}\ X$
abbreviation $\text{intersectionOf}::\uparrow\langle\uparrow\langle e\rangle, \uparrow\langle\uparrow\langle e\rangle\rangle\rangle$ (intersec) **where**
 $\text{intersec}\ X\ Z \equiv \Box(\forall x. (X\ x \leftrightarrow (\forall Y. (Z\ Y) \rightarrow (Y\ x))))$
axiomatization where
 $A1a: \lfloor\forall X. \mathcal{P}\ (\neg X) \rightarrow \neg(\mathcal{P}\ X)\rfloor$ **and**
 $A1b: \lfloor\forall X. \neg(\mathcal{P}\ X) \rightarrow \mathcal{P}\ (\neg X)\rfloor$ **and**
 $A2: \lfloor\forall X\ Y. (\mathcal{P}\ X \wedge (X \Rightarrow Y)) \rightarrow \mathcal{P}\ Y\rfloor$ **and**
 $A3: \lfloor\forall Z\ X. (\text{pos}\ Z \wedge \text{intersec}\ X\ Z) \rightarrow \mathcal{P}\ X\rfloor$

lemma *True* **nitpick**[*satisfy*] **oops** — model found: axioms are consistent
lemma $\lfloor D\rfloor$ **using** $A1a\ A1b\ A2$ **by** *blast* — D axiom is implicitly assumed

Positive properties are possibly instantiated.

¹³ To prove T1, the fact is used that positive properties cannot *entail* negative ones (A2), from which the possible instantiation of positive properties follows. A computer-formalization of Leibniz's theory of concepts can be found in [1], where the notion of *concept containment* in contrast to ordinary *property entailment* is discussed.

theorem *T1*: $[\forall X. \mathcal{P} X \rightarrow \Diamond \exists^A X]$ **using** *A1a A2* **by** *blast*

Being Godlike is defined as having all (and only) positive properties.

abbreviation *God*:: $\uparrow\langle e \rangle (G)$ **where** $G \equiv (\lambda x. \forall Y. \mathcal{P} Y \rightarrow Y x)$

abbreviation *God-star*:: $\uparrow\langle e \rangle (G^*)$ **where** $G^* \equiv (\lambda x. \forall Y. \mathcal{P} Y \leftrightarrow Y x)$

lemma *GodDefsAreEquivalent*: $[\forall x. G x \leftrightarrow G^* x]$ **using** *A1b* **by** *force*

While Leibniz provides an informal proof for the compatibility of all perfections, Gödel postulates this as *A3* (the conjunction of *any* collection of positive properties is positive), which is a third-order axiom. As shown below, the only use of *A3* is to prove that being Godlike is positive (*T2*). Dana Scott, apparently noting this, proposed taking it directly as an axiom (see [15], p. 152).¹⁴

theorem *T2*: $[\mathcal{P} G]$ **proof** –

```
{ fix w
  have 1: ((pos P) ∧ (intersec G P)) w by simp
  have (∀ Z X. (pos Z ∧ intersec X Z) → P X) w using A3 by (rule allE)
  hence (((pos P) ∧ (intersec G P)) → P G) w using allE by (rule allE)
  hence ((pos P ∧ intersec G P) w) → P G w by simp
  hence P G w using 1 by (rule mp)
} thus ?thesis by (rule allI)
qed
```

Conclusion for the first part: Possibly God exists.

theorem *T3*: $[\Diamond \exists^A G]$ **using** *T1 T2* **by** *simp*

3.2 Part II - God's Existence is Necessary, if Possible

We show here that some additional (philosophically controversial) assumptions are needed to prove the argument's conclusion, including an *essentialist* premise and the *S5* axioms. (Gödel's resp. Scott's original version works in *extensional* HOML already for modal logic *B* [8, 9]). Further derived results like monotheism and absence of free will are also discussed.

axiomatization where *A4a*: $[\forall X. \mathcal{P} X \rightarrow \Box(\mathcal{P} X)]$

A4b was originally assumed by Gödel as an axiom. We can now prove it.

lemma *A4b*: $[\forall X. \neg(\mathcal{P} X) \rightarrow \Box \neg(\mathcal{P} X)]$ **using** *A1a A1b A4a* **by** *blast*

lemma *True nitpick[satisfy] oops* — model found: all axioms A1-4 consistent

Axiom *A4a* and its consequence *A4b* together imply that \mathcal{P} satisfies Fitting's *stability conditions* ([15], p. 124). This means \mathcal{P} designates rigidly. Note that this makes for an *essentialist* assumption which may be considered controversial by some philosophers: every property considered positive in our world (e.g. honesty) is necessarily so.

¹⁴ We provide a proof in Isabelle/Isar, a language specifically tailored for writing proofs that are both computer- and human-readable. We refer the reader to [17] for other proofs not shown in this article.

lemma $[rigid\ \mathcal{P}]$ **using** $A4a\ A4b$ **by** *blast*

Gödel defines a particular notion of essence. Y is an essence of x iff Y entails every other property x possesses.¹⁵

abbreviation $Essence::\uparrow\langle\uparrow\langle e\rangle, e\rangle\ (\mathcal{E})$ **where** $\mathcal{E}\ Y\ x \equiv Y\ x \wedge (\forall Z. Z\ x \rightarrow Y \Rightarrow Z)$

abbreviation $beingIdenticalTo::e \Rightarrow \uparrow\langle e\rangle\ (id)$ **where** $id\ x \equiv (\lambda y. y \approx x)$

Being Godlike is an essential property.

lemma $GodIsEssential: [\forall x. G\ x \rightarrow (\mathcal{E}\ G\ x)]$ **using** $A1b\ A4a$ **by** *metis*

Something can have only *one* essence.

lemma $[\forall X\ Y\ z. (\mathcal{E}\ X\ z \wedge \mathcal{E}\ Y\ z) \rightarrow (X \Rightarrow Y)]$ **by** *meson*

An essential property offers a complete characterization of an individual.

lemma $EssencesCharacterizeCompletely: [\forall X\ y. \mathcal{E}\ X\ y \rightarrow (X \Rightarrow (id\ y))]$

proof (*rule ccontr*) — Isar proof by contradiction not shown here

Gödel introduces a particular notion of *necessary existence* as the property something has, provided any essence of it is necessarily instantiated.

abbreviation $necessaryExistencePredicate::\uparrow\langle e\rangle\ (NE)$

where $NE\ x \equiv (\lambda w. (\forall Y. \mathcal{E}\ Y\ x \rightarrow \Box \exists^A Y)\ w)$

axiomatization where $A5: [\mathcal{P}\ NE]$ — necessary existence is a positive property

lemma $True$ **nitpick** $[satisfy]$ **oops** — model found: so far all axioms consistent

(Possibilist) existence of God implies its necessary (actualist) existence.

theorem $T4: [\exists\ G \rightarrow \Box \exists^A G]$ **proof** — not shown

We postulate the $S5$ axioms (via *Sahlqvist correspondence*) separately, in order to get more detailed information about their relevance in the proofs below.

axiomatization where

$ax-T$: *reflexive aRel* **and** $ax-B$: *symmetric aRel* **and** $ax-IV$: *transitive aRel*

lemma $True$ **nitpick** $[satisfy]$ **oops** — model found: axioms still consistent

Possible existence of God implies its necessary (actualist) existence (note that we only rely on axioms B and IV).

theorem $T5: [\Diamond \exists\ G] \longrightarrow [\Box \exists^A G]$ **proof** — not shown

theorem $GodExistsNecessarily: [\Box \exists^A G]$ **using** $T3\ T5$ **by** *metis*

lemma $GodExistenceIsValid: [\exists^A G]$ **using** $GodExistsNecessarily\ ax-T$ **by** *auto*

Monotheism for non-normal models (using Leibniz equality) follows directly from God having all and only positive properties, but the proof for normal models is trickier. We need to consider previous results ([15], p. 162).

lemma $Monotheism-LeibnizEq: [\forall x. G * x \rightarrow (\forall y. G * y \rightarrow x \approx^L y)]$ **by** *meson*

¹⁵ Essence is defined here (and in Fitting's variant) in the version of Scott; Gödel's original version leads to the inconsistency reported in [9, 10].

lemma *Monotheism-normal*: $[\exists x.\forall y. G y \leftrightarrow x \approx y]$ **proof** — — not shown

Fitting [15] also discusses the objection raised by Sobel [27], who argues that Gödel's axiom system is too strong since it implies that whatever is the case is so necessarily: the modal system collapses. In the context of our S5 axioms, we can formalize Sobel's argument and prove modal collapse valid ([15], pp. 163-4).

lemma *useful*: $(\forall x. \varphi x \rightarrow \psi) \Rightarrow ((\exists x. \varphi x) \rightarrow \psi)$ **by** *simp*

lemma *ModalCollapse*: $[\forall \Phi. \Phi \rightarrow \Box \Phi]$ **proof** —

```
{ fix w
  { fix Q
    have  $\forall x. G x \rightarrow (\mathcal{E} G x)$  w using GodIsEssential by (rule allE)
    hence  $\forall x. G x w \rightarrow (Q \rightarrow \Box(\forall^A z. G z \rightarrow Q))$  w by force
    hence 1:  $(\exists x. G x w) \rightarrow ((Q \rightarrow \Box(\forall^A z. G z \rightarrow Q)) w)$  by (rule useful)
    have  $\exists x. G x w$  using GodExistenceIsValid by auto
    from 1 this have  $(Q \rightarrow \Box(\forall^A z. G z \rightarrow Q)) w$  by (rule mp)
    hence  $(Q \rightarrow \Box((\exists^A z. G z) \rightarrow Q)) w$  using useful by blast
    hence  $(Q \rightarrow (\Box(\exists^A z. G z) \rightarrow \Box Q)) w$  by simp
    hence  $(Q \rightarrow \Box Q) w$  using GodExistsNecessarily by simp
  } hence  $(\forall \Phi. \Phi \rightarrow \Box \Phi) w$  by (rule allI)
} thus ?thesis by (rule allI)
qed
```

4 Fitting's Variant

In this section we consider Fitting's solution to the objections raised in his discussion of Gödel's Argument ([15], pp. 164-9), especially the problem of modal collapse, which has been metaphysically interpreted as implying a rejection of free will. In Gödel's variant, positiveness and essence were thought of as predicates applying to *intensional* properties and correspondingly formalized using intensional types for their arguments ($\uparrow\langle\langle e \rangle\rangle$ and $\uparrow\langle\langle e \rangle, e \rangle$ respectively). In this variant, Fitting chooses to reformulate these definitions using *extensional* types ($\uparrow\langle\langle e \rangle\rangle$ and $\uparrow\langle\langle e \rangle, e \rangle$) instead, and makes the corresponding adjustments to the rest of the argument (to ensure type correctness). This has some philosophical repercussions; e.g. while we could say before that honesty (as concept) was a positive property, now we can only talk of its extension at some world and say of some group of people that they are honest (necessarily honest, in fact, because \mathcal{P} has also been proven 'rigid' in this variant).¹⁶

consts *Positiveness*:: $\uparrow\langle\langle e \rangle\rangle (\mathcal{P})$

abbreviation *Entails*:: $\uparrow\langle\langle e \rangle, \langle e \rangle\rangle$ (**infix** \Rightarrow) **where** $X \Rightarrow Y \equiv \Box(\forall^A z. \langle\langle X z \rangle\rangle \rightarrow \langle\langle Y z \rangle\rangle)$

abbreviation *Essence*:: $\uparrow\langle\langle e \rangle, e \rangle (\mathcal{E})$ **where** $\mathcal{E} Y x \equiv \langle\langle Y x \rangle\rangle \wedge (\forall Z. \langle\langle Z x \rangle\rangle \rightarrow (Y \Rightarrow Z))$

Axioms and theorems remain essentially the same. Particularly (T2) $[\mathcal{P} \downarrow G]$ and (A5) $[\mathcal{P} \downarrow NE]$ work with *relativized* extensional terms now.

theorem *T1*: $[\forall X::\langle e \rangle. \mathcal{P} X \rightarrow \Diamond(\exists^A z. \langle\langle X z \rangle\rangle)]$ **using** *A1a A2* **by** *blast*

¹⁶ In what follows, the ' $\langle\langle - \rangle\rangle$ ' parentheses are used to convert an extensional object into its 'rigid' intensional counterpart (e.g. $\langle\langle \varphi \rangle\rangle \equiv \lambda w. \varphi$).

theorem *T3deRe*: $[(\lambda X. \Diamond \exists^A X) \downarrow G]$ **using** *T1 T2* **by** *simp*
lemma *GodIsEssential*: $[\forall x. G x \rightarrow ((\mathcal{E} \downarrow_1 G) x)]$ **using** *A1b* **by** *metis*

The following theorem could be formalized in two variants¹⁷ (drawing on the *de re/de dicto* distinction). We prove both of them valid and show how the argument splits, culminating in two non-equivalent versions of the conclusion, both of which are proven valid.

lemma *T4v1*: $[\exists \downarrow G \rightarrow \Box \exists^A \downarrow G]$ **proof** — not shown
lemma *T4v2*: $[\exists \downarrow G \rightarrow ((\lambda X. \Box \exists^A X) \downarrow G)]$ **using** *A4a T4v1* **by** *metis*

In contrast to Gödel's version (as presented by Fitting), the following theorems can be proven in logic *K* (the *S5* axioms are no longer needed).

lemma *T5v1*: $[\Diamond \exists \downarrow G] \rightarrow [\Box \exists^A \downarrow G]$ **using** *T4v1 T3deRe* **by** *metis*
lemma *T5v2*: $[(\lambda X. \Diamond \exists^A X) \downarrow G] \rightarrow [(\lambda X. \Box \exists^A X) \downarrow G]$ **using** *T4v2* **by** *blast*

Necessary Existence of God (*de dicto* and *de re* readings).

lemma *GodNecExists-deDicto*: $[\Box \exists^A \downarrow G]$ **using** *T3deRe T4v1* **by** *blast*
lemma *GodNecExists-deRe*: $[(\lambda X. \Box \exists^A X) \downarrow G]$ **using** *T3deRe T5v2* **by** *blast*

Modal collapse is countersatisfiable even in *S5*. Note that countermodels with a cardinality of *one* for the domain of individuals are found by *Nitpick* (the countermodel shown in Fitting's book has cardinality of *two*).

lemma *equivalence aRel* $\Rightarrow [\forall \Phi. \Phi \rightarrow \Box \Phi]$ **nitpick**[*card e=1, card w=2*] **oops**

5 Anderson's Variant

In this section, we verify Anderson's emendation of Gödel's argument [3], as presented by Fitting ([15], pp. 169-171). In the previous variants there were no 'indifferent' properties, either a property or its negation had to be positive. Anderson makes room for 'indifferent' properties by dropping axiom *A1b* ($[\forall X. \neg(\mathcal{P} X) \rightarrow \mathcal{P}(\neg X)]$). As a consequence, he changes the following definitions to ensure argument's validity.

abbreviation *God*:: $\uparrow\langle e \rangle (G)$ **where** $G \equiv \lambda x. \forall Y. (\mathcal{P} Y) \leftrightarrow \Box(Y x)$
abbreviation *Essence*:: $\uparrow\langle \uparrow\langle e \rangle, e \rangle (\mathcal{E})$ **where** $\mathcal{E} Y x \equiv (\forall Z. \Box(Z x) \leftrightarrow Y \Rightarrow Z)$

There is now the requirement that a Godlike being must have positive properties *necessarily*. For the definition of essence, Scott's addition [25], that the essence of an object actually applies to the object, is dropped. A necessity operator has been introduced instead.¹⁸

The rest of the argument is essentially similar to Gödel's (also in *S5* logic).

theorem *T1*: $[\forall X. \mathcal{P} X \rightarrow \Diamond \exists^A X]$ **using** *A1a A2* **by** *blast*
theorem *T3*: $[\Diamond \exists^A G]$ **using** *T1 T2* **by** *simp*

¹⁷ Fitting's original treatment in [15] left several details unspecified and we had to fill in the gaps by choosing appropriate formalization variants (see [17] for details).

¹⁸ Gödel's original axioms (without Scott's addition) are proven inconsistent in [9].

If g is Godlike, the property of being Godlike is its essence.¹⁹

theorem *GodIsEssential*: $[\forall x. G x \rightarrow (\mathcal{E} G x)]$ **proof** – — not shown

The necessary existence of God follows from its possible existence.

theorem *T5*: $[\Diamond \exists G] \rightarrow [\Box \exists^A G]$ **proof** – — not shown

The conclusion could be proven (with one fewer axiom, though more complex definitions) and *Nitpick* is able to find a countermodel for the *modal collapse*.

lemma *GodExistsNecessarily*: $[\Box \exists^A G]$ **using** *T3 T5* **by** *metis*

lemma *ModalCollapse*: $[\forall \Phi. \Phi \rightarrow \Box \Phi]$ **nitpick oops** — countersatisfiable

6 Conclusion

We presented a shallow semantic embedding in Isabelle/HOL for an intensional higher-order modal logic (a successor of Montague/Gallin intensional logics) and employed this logic to formalize and verify three different variants of the ontological argument: the first one by Gödel himself (resp. Scott), the second one by Fitting and the last one by Anderson.

By employing our embedding of IHOML in Isabelle/HOL, we could not only verify Fitting’s results, but also guarantee consistency of axioms. Moreover, for many theorems we could prove stronger versions and find better countermodels (i.e. with smaller cardinality) than the ones presented by Fitting. Another interesting aspect was the possibility to explore the implications of alternative formalizations of axioms and theorems which shed light on interesting philosophical issues concerning entailment, essentialism and free will.

The latest developments in automated theorem proving, in combination with the embedding approach, allow us to engage in much better experimentation during the formalization and assessment of arguments than ever before. The potential reduction (of several orders of magnitude) in the time needed for proving or disproving theorems (compared to pen-and-paper proofs), results in almost real-time feedback about the suitability of our speculations. The practical benefits of computer-supported argumentation go beyond mere quantitative aspects (easier, faster and more reliable proofs). The advantages are also qualitative, since a significantly different approach to argumentation is fostered: We can now work iteratively (by trial-and-error) on an argument by making gradual adjustments to its definitions, axioms and theorems. This allows us to continuously expose and revise the assumptions we indirectly commit ourselves to every time we opt for some particular formalization.

¹⁹ This theorem’s proof could be completely automatized for Gödel’s and Fitting’s variants. For Anderson’s version however, we had to reproduce in Isabelle/HOL the original natural-language proof given by Anderson (see [3], Theorem 2*, p. 296).

References

1. J. Alama, P. E. Oppenheimer, and E. N. Zalta. Automating Leibniz’s theory of concepts. In A. P. Felty and A. Middeldorp, editors, *Automated Deduction - CADE-25 - 25th International Conference on Automated Deduction, Berlin, Germany, August 1-7, 2015, Proceedings*, volume 9195 of *LNCS*, pages 73–97. Springer, 2015.
2. A. 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.
3. C. Anderson. Some emendations of Gödel’s ontological proof. *Faith and Philosophy*, 7(3), 1990.
4. C. Benzmüller. Universal reasoning, rational argumentation and human-machine interaction. *arXiv*, <http://arxiv.org/abs/1703.09620>, 2017.
5. C. Benzmüller, M. Claus, and N. Sultana. Systematic verification of the modal logic cube in Isabelle/HOL. In C. Kaliszyk and A. Paskevich, editors, *PxTP 2015. EPTCS*, volume 186, pages 27–41, Berlin, Germany, 2015.
6. C. Benzmüller and L. Paulson. Quantified multimodal logics in simple type theory. *Logica Universalis (Special Issue on Multimodal Logics)*, 7(1):7–20, 2013.
7. C. Benzmüller, L. Weber, and B. Woltzenlogel-Paleo. Computer-assisted analysis of the Anderson-Hájek controversy. *Logica Universalis*, 11(1):139–151, 2017.
8. C. Benzmüller and B. Woltzenlogel Paleo. Automating Gödel’s ontological proof of God’s existence with higher-order automated theorem provers. In T. Schaub, G. Friedrich, and B. O’Sullivan, editors, *ECAI 2014*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 93 – 98. IOS Press, 2014.
9. C. Benzmüller and B. Woltzenlogel Paleo. The inconsistency in Gödel’s ontological argument: A success story for AI in metaphysics. In *IJCAI 2016*, 2016.
10. C. Benzmüller and B. Woltzenlogel Paleo. An object-logic explanation for the inconsistency in Gödel’s ontological theory (extended abstract). In M. Helmert and F. Wotawa, editors, *KI 2016: Advances in Artificial Intelligence, Proceedings*, volume 9725 of *LNCS*, pages 43–50, Berlin, Germany, 2016.
11. F. Bjørdal. Understanding Gödel’s ontological argument. In T. Childers, editor, *The Logica Yearbook 1998*. Filosofia, 1999.
12. J. Blanchette and T. Nipkow. Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In *Proc. of ITP 2010*, volume 6172 of *LNCS*, pages 131–146. Springer, 2010.
13. A. Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5:56–68, 1940.
14. B. Fitelson and E. N. Zalta. Steps toward a computational metaphysics. *J. Philosophical Logic*, 36(2):227–247, 2007.
15. M. Fitting. *Types, Tableaus and Gödel’s God*. Kluwer, 2002.
16. M. Fitting and R. Mendelsohn. *First-Order Modal Logic*, volume 277 of *Synthese Library*. Kluwer, 1998.
17. D. Fuenmayor and C. Benzmüller. Types, Tableaus and Gödel’s God in Isabelle/HOL. *Archive of Formal Proofs*, 2017. Formally verified with Isabelle/HOL.
18. D. Gallin. *Intensional and Higher-Order Modal Logic*. N.-Holland, 1975.
19. K. Gödel. *Appx.A: Notes in Kurt Gödel’s Hand*, pages 144–145. In [27], 2004.
20. P. Hájek. A new small emendation of Gödel’s ontological proof. *Studia Logica*, 71(2):149–164, 2002.
21. S. Kripke. *Naming and Necessity*. Harvard University Press, 1980.

22. T. Nipkow, L. Paulson, and M. Wenzel. *Isabelle/HOL — A Proof Assistant for Higher-Order Logic*, volume 2283 of *LNCS*. Springer, 2002.
23. P. Oppenheimer and E. Zalta. A computationally-discovered simplification of the ontological argument. *Australasian Journal of Philosophy*, 89(2):333–349, 2011.
24. J. Rushby. The ontological argument in PVS. In *Proc. of CAV Workshop “Fun With Formal Methods”*, St. Petersburg, Russia, 2013.
25. D. Scott. *Appx.B: Notes in Dana Scott’s Hand*, pages 145–146. In [27], 2004.
26. J. Sobel. Gödel’s ontological proof. In *On Being and Saying. Essays for Richard Cartwright*, pages 241–261. MIT Press, 1987.
27. J. Sobel. *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge U. Press, 2004.
28. M. Wisniewski, A. Steen, and C. Benz Müller. Einsatz von Theorembeweisern in der Lehre. In A. Schwill and U. Lucke, editors, *Hochschuldidaktik der Informatik: 7. Fachtagung des GI-Fachbereichs Informatik und Ausbildung/Didaktik der Informatik; 13.-14. September 2016 an der Universität Potsdam*, *Commentarii informaticae didacticae (CID)*, Potsdam, Germany, 2016.