Can Computers Help to Sharpen our Understanding of Ontological Arguments? (invited paper)

Christoph Benzmüller[*]
Freie Universität Berlin, Germany & University of Luxembourg, Luxembourg

David Fuenmayor
Freie Universität Berlin, Germany

Abstract

In the past decades, several emendations of Gödel's (resp. Scott's) modal ontological argument have been proposed, many of which preserve the intended conclusion (the necessary existence of God), while avoiding a controversial side result of the Gödel/Scott variant called the "modal collapse", which expresses that there are no contingent truths (everything is determined, there is no free will). In this paper we provide a summary on recent computer-supported verification studies on some of these modern variants of the ontological argument. The purpose is to provide further evidence that the interaction with the computer technology, which we have developed together with colleagues over the past years, can not only enable the formal assessment of ontological arguments, but can, in fact, help to sharpen our conceptual understanding of the notions and concepts involved. From a more abstract perspective, we claim that interreligious understanding may be fostered by means of formal argumentation and, in particular, formal logical analysis supported by modern interactive and automated theorem proving technology.

1 Introduction

Is religious rational argumentation possible? Or is religion a conversation-stopper? A positive answer to the first question requires us to find a way of acknowledging the variety of religious beliefs, while at the same time recognising that we all share, at heart, a similar assortment of concepts and are thus able to successfully understand each other (e.g. by drawing similar inferences in similar situations). Negative answers to this question often rely on the assumption that religious beliefs provide a conceptual framework through which a believer's worldview is structured, which in turn implies the futility of any interpretation effort of religious discourse, given the

[*] Funded by Volkswagen Foundation.

incommensurability between the conceptual schemes of speakers and interpreters belonging to different creeds.[1]

In this paper we argue for the possibility of effective interreligious understanding through the means of rational argumentation and, in particular, formal logic supported by modern interactive and automated theorem proving technology. Drawing upon past experience with the computer-assisted reconstruction and assessment of ontological arguments for the existence of God [11, 10, 9, 21], we aim at illustrating how it is indeed possible to use formal logic as a common ground for metaphysical and theological discussions. An interesting insight of this work concerns the conception of logic as an *ars explicandi*, in which the process of understanding (explicating) metaphysical and theological concepts take place in the very practice of argumentation (i.e. by using them to draw logical inferences). Thus, formal logic may help us to make explicit the inferential role played by the argument's concepts and thus better understand their meaning. In the context of a formalised argument (in some chosen logic), this task of conceptual explication can be carried out e.g. by direct definition or by axiomatising conceptual interrelations. The suitability of these axioms and definitions can then be evaluated by computing the logical validity of the formalised argument.[2]

We also want to highlight the prospects of using computers to systematically explore the many different inferential possibilities latent in a given argument. By utilising automated theorem proving technology we are able to engage efficiently in the speculative enterprise of coming up with new ideas (e.g. new axioms or definitions for explicating a concept) and testing them in real-time with the help of automated tools. In this paper we will illustrate how this approach has been put into practice in several case studies concerning ontological arguments for the existence of God and how this had led to interesting conceptual insights. In the following sections we therefore present and discuss some exemplary applications of computer-assisted formal methods to the analysis and critical assessment of ontological arguments for the existence of God in the spirit of computational metaphysics.[3] The arguments to be analysed draw upon Kurt Gödel's [22] modern variant of the ontological argument for the existence of God. Gödel's ontological argument [22] (Section 2) is amongst the most discussed formal proofs in modern literature. Several authors, including e.g. [30, 3, 2, 14, 24, 19], have proposed emendations with the aim of retaining its main conclusion (the necessary existence of God) while at the same time blocking the inference of the so-called modal collapse (whatever is the case is so necessarily)

[1] Terry Godlove has convincingly argued in [23] against what he calls the "framework theory" in religious studies, according to which, for believers, religious beliefs shape the interpretation of most of the objects and situations in their lives. Here Godlove relies on Donald Davidson's rejection of "the very idea of a conceptual scheme" [17].

[2] According to a well-known principle of interpretation, the principle of charity, "we make maximum sense of the words and thoughts of others when we interpret in a way that optimises agreement" [17]. We draw upon the principle of charity to presume (and foster) the logical validity of the argument we aim at analysing.

[3] Computational metaphysics is an emerging, interdisciplinary field aiming at the rigorous formalisation and deep logical assessment of philosophical arguments in an automated reasoning environment. This term was originally coined by Fitelson and Zalta in [18] as a way to describe the application of modern computer technologies to the centuries-old program of axiomatic metaphysics: applying the axiomatic method ("reasoning from first principles") to problems in philosophy, particularly in metaphysics.

[31, 32]. The modal collapse is an undesirable side-effect of the axioms postulated by Gödel (and Scott). It essentially states that there are no contingent truths and everything is determined. The modal collapse is in fact successfully avoided in the variants of the argument as proposed by Anderson [3, 2] (Section 3) and Fitting [19] (Section 4).

2 Gödel's Ontological Argument

Gödel's variant of the ontological argument is a direct descendant of Leibniz's, which in turn derives from Descartes'. All these arguments have a two-part structure: (i) prove that God's existence is possible, and (ii) prove that God's existence is necessary, if possible. The main conclusion (God's necessary existence) then follows from (i) and (ii), either by modus ponens (in non-modal contexts) or by invoking some axioms of modal logic (notably, but not necessarily as we will see, the so-called modal logic system S5).

Part I - Proving that God's existence is possible

We start out with definition D1 (Godlike) and axioms A1-A3.

D1 Being Godlike is equivalent to having all positive properties.

A1 Exactly one of a property or its negation is positive.

A2 Any property entailed by a positive property is positive.

A3 The combination of any collection of positive properties is itself positive.

From A1 and A2 follows theorem T1:

T1 Every positive property is possibly instantiated (if a property P is positive, then it is possible that some being has property P).

From D1 and A3 follows:

T2 Being Godlike is a positive property.

From T1 and T2 follows:

T3 Being Godlike is possibly instantiated.

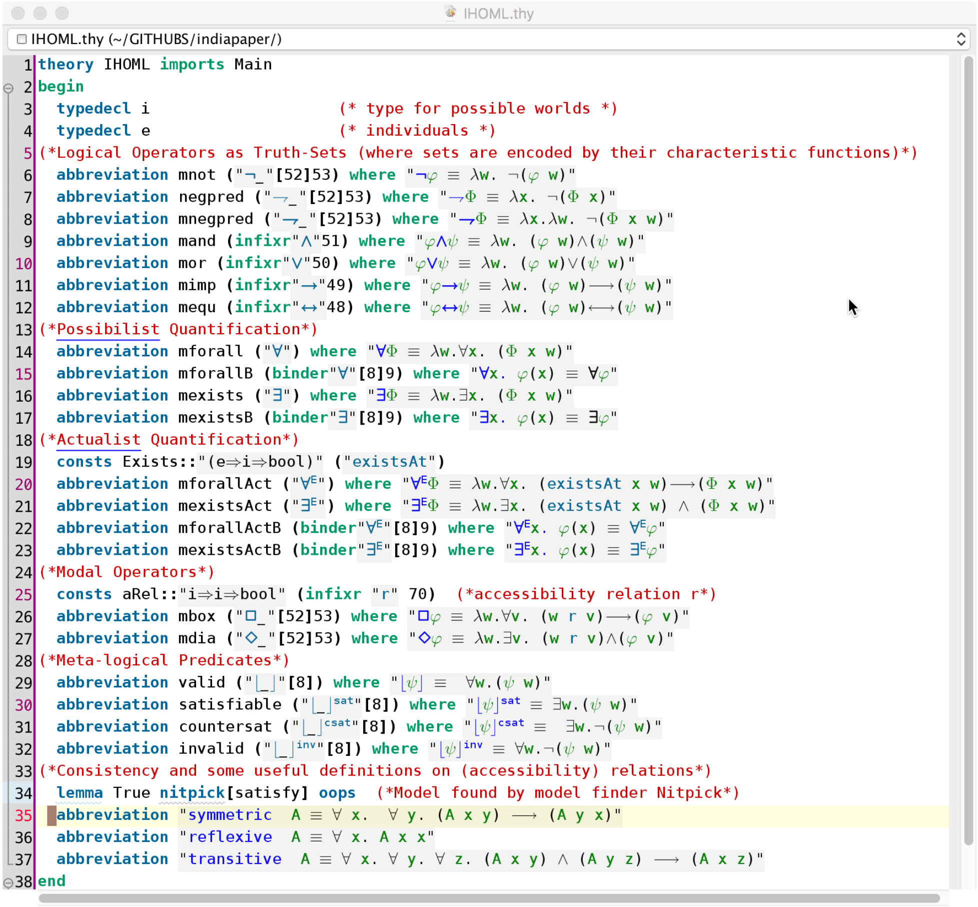
Part II - Proving that God's existence is necessary, if possible

D2 A property E is the essence of an individual x iff x has E and all of x's properties are entailed by E.[4]

A4 Positive (negative) properties are necessarily positive (negative).

From A1 and A4 (using definitions D1 and D2) follows:

[4] The underlined part in definition D2 has been added by Scott [30]. Gödel [22] originally omitted this part; more on this in Section 2.3 below.



```
theory IHOML imports Main
begin
  typedecl i                    (* type for possible worlds *)
  typedecl e                    (* individuals *)
(*Logical Operators as Truth-Sets (where sets are encoded by their characteristic functions)*)
  abbreviation mnot ("¬_"[52]53) where "¬φ ≡ λw. ¬(φ w)"
  abbreviation negpred ("¬_"[52]53) where "¬Φ ≡ λx. ¬(Φ x)"
  abbreviation mnegpred ("¬_"[52]53) where "¬Φ ≡ λx.λw. ¬(Φ x w)"
  abbreviation mand (infixr"∧"51) where "φ∧ψ ≡ λw. (φ w)∧(ψ w)"
  abbreviation mor (infixr"∨"50) where "φ∨ψ ≡ λw. (φ w)∨(ψ w)"
  abbreviation mimp (infixr"→"49) where "φ→ψ ≡ λw. (φ w)⟶(ψ w)"
  abbreviation mequ (infixr"↔"48) where "φ↔ψ ≡ λw. (φ w)⟷(ψ w)"
(*Possibilist Quantification*)
  abbreviation mforall ("∀") where "∀Φ ≡ λw.∀x. (Φ x w)"
  abbreviation mforallB (binder"∀"[8]9) where "∀x. φ(x) ≡ ∀φ"
  abbreviation mexists ("∃") where "∃Φ ≡ λw.∃x. (Φ x w)"
  abbreviation mexistsB (binder"∃"[8]9) where "∃x. φ(x) ≡ ∃φ"
(*Actualist Quantification*)
  consts Exists::"(e⇒i⇒bool)" ("existsAt")
  abbreviation mforallAct ("∀^E") where "∀^EΦ ≡ λw.∀x. (existsAt x w)⟶(Φ x w)"
  abbreviation mexistsAct ("∃^E") where "∃^EΦ ≡ λw.∃x. (existsAt x w) ∧ (Φ x w)"
  abbreviation mforallActB (binder"∀^E"[8]9) where "∀^Ex. φ(x) ≡ ∀^Eφ"
  abbreviation mexistsActB (binder"∃^E"[8]9) where "∃^Ex. φ(x) ≡ ∃^Eφ"
(*Modal Operators*)
  consts aRel::"i⇒i⇒bool" (infixr "r" 70)  (*accessibility relation r*)
  abbreviation mbox ("□_"[52]53) where "□φ ≡ λw.∀v. (w r v)⟶(φ v)"
  abbreviation mdia ("◇_"[52]53) where "◇φ ≡ λw.∃v. (w r v)∧(φ v)"
(*Meta-logical Predicates*)
  abbreviation valid ("⌊_⌋"[8]) where "⌊ψ⌋ ≡  ∀w.(ψ w)"
  abbreviation satisfiable ("⌊_⌋^sat"[8]) where "⌊ψ⌋^sat ≡ ∃w.(ψ w)"
  abbreviation countersat ("⌊_⌋^csat"[8]) where "⌊ψ⌋^csat ≡  ∃w.¬(ψ w)"
  abbreviation invalid ("⌊_⌋^inv"[8]) where "⌊ψ⌋^inv ≡ ∀w.¬(ψ w)"
(*Consistency and some useful definitions on (accessibility) relations*)
  lemma True nitpick[satisfy] oops  (*Model found by model finder Nitpick*)
  abbreviation "symmetric  A ≡ ∀ x.  ∀ y. (A x y) ⟶ (A y x)"
  abbreviation "reflexive  A ≡ ∀ x. A x x"
  abbreviation "transitive  A ≡ ∀ x. ∀ y. ∀ z. (A x y) ∧ (A y z) ⟶ (A x z)"
end
```

Figure 1: Definitions of the Modal Logic Connectives in Isabelle/HOL

T4 Being Godlike is an essential property of any Godlike individual.

D3 Necessary existence of an individual is the necessary instantiation of all its essences.

A5 Necessary existence is a positive property.

From T4 and A5 (using D1, D2, D3) follows:

T5 (The property of) being Godlike, if instantiated, is necessarily instantiated.

And finally from T3, T5 (together with some implicit modal axioms, e.g. S5) the existence of (at least a) God follows:

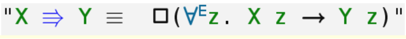
T6 Being Godlike is actually instantiated.

Figure 2 presents the formalisation of the above argument in the proof assistant Isabelle/HOL [26], which is based on classical higher-order logic [4, 6]. The definitions of the modal logic connectives are provided in the imported theory file IHOML.thy as presented in Fig. 1. Explanations and justifications for the provided definitions are available in various earlier papers, see e.g. [8, 21, 20].

2.1 The Concept of Entailment

While Leibniz provides an informal proof for the compatibility of all perfections (i.e. positive properties),[5] Gödel postulates this as a third-order axiom (A3: the conjunction/combination of any collection of positive properties is positive). As shown above, the only use of A3 is to prove that being Godlike is positive (T2). Noting this, Scott has proposed taking T2 directly as an axiom (which we do in Sections 3 and 4 as well).

Notice that in order to prove T1, Gödel assumes A2 (any property entailed by a positive property is positive). As we have discussed in previous work [20, 21], the success of this argumentation depends on the way Gödel formalises the notion of entailment in classical logic: Property x entails property Y if and only if it is necessary that every (existing) being that has property x also has property Y.



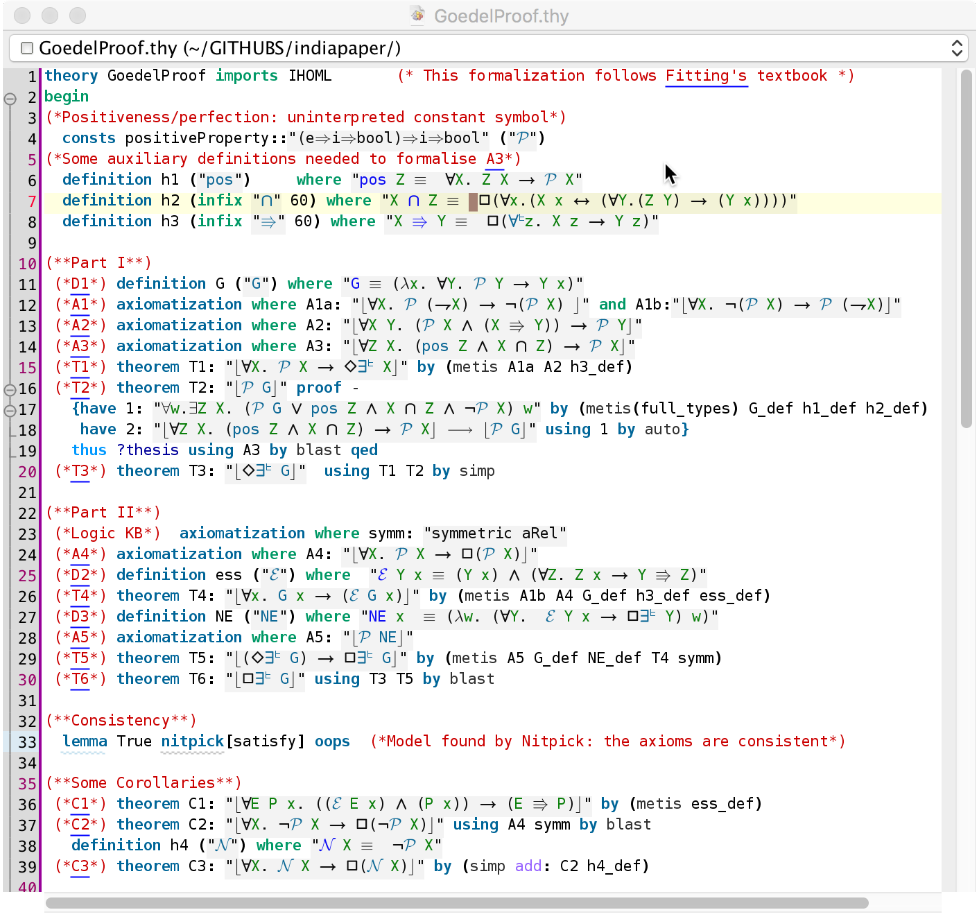
```
"X ⇛ Y ≡  □(∀ᴱz. X z → Y z)"
```

The definition of property entailment as formalised by Gödel can be criticised on the grounds that it lacks some notion of relevance, since an impossible property (like being a round square) would entail any property (like being a triangle). Moreover, when we assert that property A does not entail property B, we implicitly assume that A is possibly instantiated (since there exists at least one being that has A but does not have B). It is by virtue of these subtleties that Gödel's original formulation manages to prove T1. Since positive properties cannot entail negative ones (because of A2), positive properties need to be possibly instantiated.

Such 'question-begging' aspects, as discussed above, could only be made explicit and discussed thanks to the fact that Gödel exposed his argument using a formal language. Other, pure natural-language versions of the ontological argument may well contain similar logical peculiarities which have so far remained undetected. An alternative formalisation of the concept of entailment would eventually suitably emend this criticism of Gödel's argument.[6] However, it would eventually lead to issues (maybe even inconsistency) in other parts of the argument. We are convinced that the computer-assisted application of formal logic as exemplified in Figs. 2 and 3 constitutes a most promising way of bringing such logical and conceptual issues to the surface. This vision is shared by other researchers; see e.g. the related woks by Zalta [27, 1] and Rushby [28, 29].

[5] See e.g. Leibniz's "Two Notations for Discussion with Spinoza" [25]. For a genealogy of Gödel's argument refer to [19] p.133ff.

[6] Consider e.g. the computer-assisted formalisation of Leibniz's theory of concepts in [1], where the Leibnizian notion of concept containment is discussed.



```
theory GoedelProof imports IHOML       (* This formalization follows Fitting's textbook *)
begin
(*Positiveness/perfection: uninterpreted constant symbol*)
  consts positiveProperty::"(e⇒i⇒bool)⇒i⇒bool" ("𝒫")
(*Some auxiliary definitions needed to formalise A3*)
  definition h1 ("pos")     where "pos Z ≡  ∀X. Z X → 𝒫 X"
  definition h2 (infix "∩" 60) where "X ∩ Z ≡ □(∀x.(X x ↔ (∀Y.(Z Y) → (Y x))))"
  definition h3 (infix "⇛" 60) where "X ⇛ Y ≡  □(∀ᴱz. X z → Y z)"

(**Part I**)
 (*D1*) definition G ("G") where "G ≡ (λx. ∀Y. 𝒫 Y → Y x)"
 (*A1*) axiomatization where A1a: "⌊∀X. 𝒫 (⇁X) → ¬(𝒫 X) ⌋" and A1b:"⌊∀X. ¬(𝒫 X) → 𝒫 (⇁X)⌋"
 (*A2*) axiomatization where A2: "⌊∀X Y. (𝒫 X ∧ (X ⇛ Y)) → 𝒫 Y⌋"
 (*A3*) axiomatization where A3: "⌊∀Z X. (pos Z ∧ X ∩ Z) → 𝒫 X⌋"
 (*T1*) theorem T1: "⌊∀X. 𝒫 X → ◇∃ᴱ X⌋" by (metis A1a A2 h3_def)
 (*T2*) theorem T2: "⌊𝒫 G⌋" proof -
   {have 1: "∀w.∃Z X. (𝒫 G ∨ pos Z ∧ X ∩ Z ∧ ¬𝒫 X) w" by (metis(full_types) G_def h1_def h2_def)
    have 2: "⌊∀Z X. (pos Z ∧ X ∩ Z) → 𝒫 X⌋ ⟶ ⌊𝒫 G⌋" using 1 by auto}
   thus ?thesis using A3 by blast qed
 (*T3*) theorem T3: "⌊◇∃ᴱ G⌋"  using T1 T2 by simp

(**Part II**)
 (*Logic KB*)  axiomatization where symm: "symmetric aRel"
 (*A4*) axiomatization where A4: "⌊∀X. 𝒫 X → □(𝒫 X)⌋"
 (*D2*) definition ess ("ℰ") where  "ℰ Y x ≡ (Y x) ∧ (∀Z. Z x → Y ⇛ Z)"
 (*T4*) theorem T4: "⌊∀x. G x → (ℰ G x)⌋" by (metis A1b A4 G_def h3_def ess_def)
 (*D3*) definition NE ("NE") where "NE x  ≡ (λw. (∀Y.  ℰ Y x → □∃ᴱ Y) w)"
 (*A5*) axiomatization where A5: "⌊𝒫 NE⌋"
 (*T5*) theorem T5: "⌊(◇∃ᴱ G) → □∃ᴱ G⌋" by (metis A5 G_def NE_def T4 symm)
 (*T6*) theorem T6: "⌊□∃ᴱ G⌋" using T3 T5 by blast

(**Consistency**)
  lemma True nitpick[satisfy] oops  (*Model found by Nitpick: the axioms are consistent*)

(**Some Corollaries**)
 (*C1*) theorem C1: "⌊∀E P x. ((ℰ E x) ∧ (P x)) → (E ⇛ P)⌋" by (metis ess_def)
 (*C2*) theorem C2: "⌊∀X. ¬𝒫 X → □(¬𝒫 X)⌋" using A4 symm by blast
   definition h4 ("𝒩") where "𝒩 X ≡  ¬𝒫 X"
 (*C3*) theorem C3: "⌊∀X. 𝒩 X → □(𝒩 X)⌋" by (simp add: C2 h4_def)
```

Figure 2: Gödel's Ontological Argument in Isabelle/HOL (to be continued in Fig. 3)

2.2 Modal Logic KB, versus Modal Logic S5

By exploring the many different possibilities of combining axiom systems in our framework, we can quickly verify that the above argument is indeed already provable in the comparably weak modal logic KB; cf. also [10, 13]. Logic KB (only) assumes that transition relations between possible worlds are symmetric. Modal logic S5, in contrast, assumes that the transition relations are equivalence relations. The symmetry of transition (or accessibility) relations is postulated in Fig. 2 in line 23. Symmetry, as a semantic, meta-level condition, corresponds to what is usually called the 'B' axiom: if a statement A is the case, then A is necessarily possibly the case.[7] Note that logic KB is needed only in Part II of the argument. Part I only requires modal logic K.

Gödel's original argument was (presumably) aimed at being formalised in what logicians call nowadays an S5 modal logic, which features, amongst others, the following principle: if a statement A is possibly the case, then A is necessarily possibly the case. The argument has in fact been criticised for implying this rather strong and unintuitive assertion. Due to the use of logic KB we are here not committed to accept such controversial implications.

Consistency of the argument's assumptions is guaranteed for both logics, KB and S5. This has been confirmed by the model finder Nitpick [15]; cf. line 33 of Fig. 2 (for KB) and line 88 of Fig. 3 (for S5).

2.3 The Notion of Essence and Inconsistency

Another important result of previous work [11, 12], arrived at by the use of automated theorem provers, is as follows: Gödel's original formulation of the argument is logically inconsistent, i.e., its premises can be used to derive a logical contradiction. The problem is that Gödel originally defined essence as:

```
"ℰ Y x ≡ (∀Z. Z x → Y ⇛ Z)"
```

D2' A property Y is the essence of an individual x if all of x's properties are entailed by Y.

We can ask ourselves if, according to the original definition D2', individuals must also exemplify their essential properties (since it is not explicitly required in the definition). While a positive answer to this question may correspond to our intuitions, the definition given by Gödel in [22] omits the requirement and allows for individuals not exemplifying their essential properties. As a controversial consequence, non-exemplified (empty) properties, such as the property of being self-different, are in fact becoming essential properties of any individual according to Gödel's notion of essence. And, as shown in previous work [11, 12] with the help of automated theorem provers (not repeated here due to space restrictions), this controversial consequence in fact renders Gödel's theory inconsistent. As discussed in detail

[7] Automated theorem provers can in fact prove this well known correspondence within our framework; this is, however, not shown here due to space restrictions, cf. [8, 7, 5] for further information on this aspect.

more detail in [12], this inconsistency holds already in base logic K, and thus also in logics KB (as used above) and S5, which are both extensions of K.

The intuitive additional requirement in D2 was added by Dana Scott [30]. When Scott's definition D2 is used (as we do above), then consistency can be shown with the help of automated model finders; cf. line 33 in Fig. 2. Scott's modification of definition D2 has thus brought consistency back to this argument.

An interesting corollary of this definition of essence is that essences characterise individuals completely (since they entail all of their properties); cf. line 36 in Fig. 2.

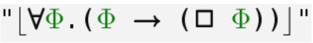
2.4 Modal Collapse

The second part of Gödel's argument features some philosophically controversial, implicit commitments, such as modal collapse (everything that is the case is necessarily the case), which directly attacks our self-understanding as agents having a free will; and the assertion that whatever is possibly necessarily the case is thereby actually the case.

The modal collapse, originally detected by Sobel [31, 32], has been confirmed as a side-effect of Gödel's argument in various previous experiments with automated theorem provers; see e.g. [10, 13, 21].

The modal collapse is independent of whether definition D2 or D2' is applied, or whether logic KB or S5 is used. For Gödel's axiom and definitions from Fig. 2 the modal collapse is proved in lines 36-42 of Fig. 3.

The modal collapse formula

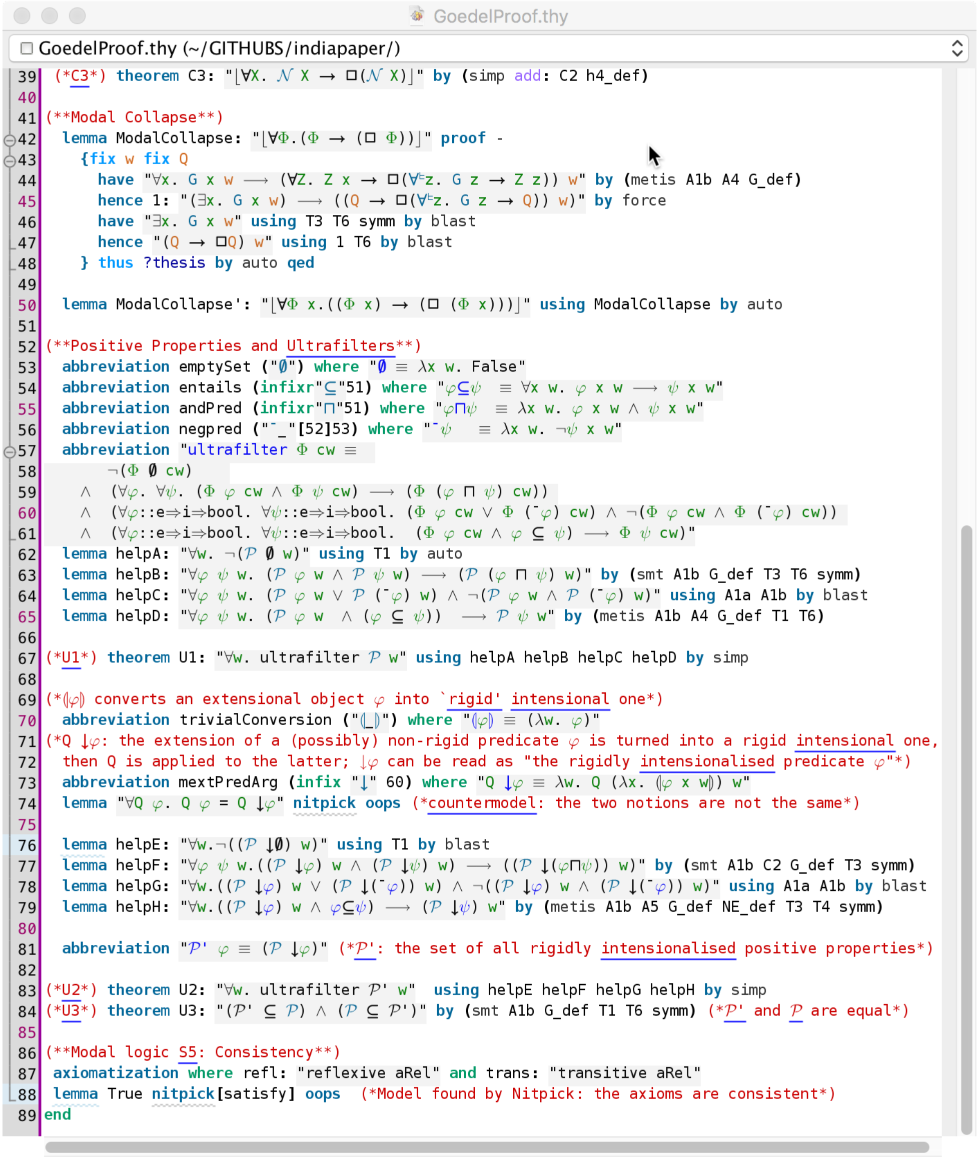


```
"⌊∀Φ.(Φ → (□ Φ))⌋"
```

reads as: "everything that is the case is so necessarily". This entails the following statement as a corollary: "for every property P, any individual having property P has P necessarily". Modal collapse is not only a quite unintuitive assertion, but also has many profound metaphysical repercussions. As an example, if it is the case that it is sunny in Berlin on August 16, 2018, then this is so necessarily, i.e., it is impossible that it could have been otherwise (and the same would apply to any other state of affairs). Other corollaries also seem intuitively quite controversial: if I happen to own a blue car, then my owning a blue car is a necessary fact, i.e., I could never have owned just a red car instead; somehow, owning a blue car is an essential trait of mine, it is part of my identity. Modal collapse implies that this happens for every single property we happen to exemplify (like owning a blue car, being brown eyed, being married to x, having done such and such, etc.). In particular, modal collapse implies that we have no agency whatsoever to choose our actions: anything we (seem to) have done so far has merely been imposed by providence upon us (simply because things could not have been otherwise). This is the reason why modal collapse has been associated with lack of free will.

2.5 Positive Properties, Negative Properties and Axiom A4

A positive property can be seen as Gödel's counterpart to what Leibniz calls a perfection, which, being an important concept in Leibnizian ethics, can be expected to



```
 (*C3*) theorem C3: "⌊∀X. 𝒩 X → □(𝒩 X)⌋" by (simp add: C2 h4_def)

(**Modal Collapse**)
  lemma ModalCollapse: "⌊∀Φ.(Φ → (□ Φ))⌋" proof -
    {fix w fix Q
      have "∀x. G x w ⟶ (∀Z. Z x → □(∀ᴱz. G z → Z z)) w" by (metis A1b A4 G_def)
      hence 1: "(∃x. G x w) ⟶ ((Q → □(∀ᴱz. G z → Q)) w)" by force
      have "∃x. G x w" using T3 T6 symm by blast
      hence "(Q → □Q) w" using 1 T6 by blast
    } thus ?thesis by auto qed

  lemma ModalCollapse': "⌊∀Φ x.((Φ x) → (□ (Φ x)))⌋" using ModalCollapse by auto

(**Positive Properties and Ultrafilters**)
  abbreviation emptySet ("∅") where "∅ ≡ λx w. False"
  abbreviation entails (infixr"⊆"51) where "φ⊆ψ  ≡ ∀x w. φ x w ⟶ ψ x w"
  abbreviation andPred (infixr"⊓"51) where "φ⊓ψ  ≡ λx w. φ x w ∧ ψ x w"
  abbreviation negpred ("¯_"[52]53) where "¯ψ   ≡ λx w. ¬ψ x w"
  abbreviation "ultrafilter Φ cw ≡
        ¬(Φ ∅ cw)
    ∧  (∀φ. ∀ψ. (Φ φ cw ∧ Φ ψ cw) ⟶ (Φ (φ ⊓ ψ) cw))
    ∧  (∀φ::e⇒i⇒bool. ∀ψ::e⇒i⇒bool. (Φ φ cw ∨ Φ (¯φ) cw) ∧ ¬(Φ φ cw ∧ Φ (¯φ) cw))
    ∧  (∀φ::e⇒i⇒bool. ∀ψ::e⇒i⇒bool.  (Φ φ cw ∧ φ ⊆ ψ) ⟶ Φ ψ cw)"
  lemma helpA: "∀w. ¬(𝒫 ∅ w)" using T1 by auto
  lemma helpB: "∀φ ψ w. (𝒫 φ w ∧ 𝒫 ψ w) ⟶ (𝒫 (φ ⊓ ψ) w)" by (smt A1b G_def T3 T6 symm)
  lemma helpC: "∀φ ψ w. (𝒫 φ w ∨ 𝒫 (¯φ) w) ∧ ¬(𝒫 φ w ∧ 𝒫 (¯φ) w)" using A1a A1b by blast
  lemma helpD: "∀φ ψ w. (𝒫 φ w  ∧ (φ ⊆ ψ))  ⟶ 𝒫 ψ w" by (metis A1b A4 G_def T1 T6)

(*U1*) theorem U1: "∀w. ultrafilter 𝒫 w" using helpA helpB helpC helpD by simp

(*⦇φ⦈ converts an extensional object φ into `rigid' intensional one*)
  abbreviation trivialConversion ("⦇_⦈") where "⦇φ⦈ ≡ (λw. φ)"
(*Q ↓φ: the extension of a (possibly) non-rigid predicate φ is turned into a rigid intensional one,
  then Q is applied to the latter; ↓φ can be read as "the rigidly intensionalised predicate φ"*)
  abbreviation mextPredArg (infix "↓" 60) where "Q ↓φ ≡ λw. Q (λx. ⦇φ x w⦈) w"
  lemma "∀Q φ. Q φ = Q ↓φ" nitpick oops (*countermodel: the two notions are not the same*)

  lemma helpE: "∀w.¬((𝒫 ↓∅) w)" using T1 by blast
  lemma helpF: "∀φ ψ w.((𝒫 ↓φ) w ∧ (𝒫 ↓ψ) w) ⟶ ((𝒫 ↓(φ⊓ψ)) w)" by (smt A1b C2 G_def T3 symm)
  lemma helpG: "∀w.((𝒫 ↓φ) w ∨ (𝒫 ↓(¯φ)) w) ∧ ¬((𝒫 ↓φ) w ∧ (𝒫 ↓(¯φ)) w)" using A1a A1b by blast
  lemma helpH: "∀w.((𝒫 ↓φ) w ∧ φ⊆ψ) ⟶ (𝒫 ↓ψ) w" by (metis A1b A5 G_def NE_def T3 T4 symm)

  abbreviation "𝒫' φ ≡ (𝒫 ↓φ)" (*𝒫': the set of all rigidly intensionalised positive properties*)

(*U2*) theorem U2: "∀w. ultrafilter 𝒫' w"  using helpE helpF helpG helpH by simp
(*U3*) theorem U3: "(𝒫' ⊆ 𝒫) ∧ (𝒫 ⊆ 𝒫')" by (smt A1b G_def T1 T6 symm) (*𝒫' and 𝒫 are equal*)

(**Modal logic S5: Consistency**)
 axiomatization where refl: "reflexive aRel" and trans: "transitive aRel"
 lemma True nitpick[satisfy] oops  (*Model found by Nitpick: the axioms are consistent*)
end
```

Figure 3: Gödel’s Ontological Argument in Isabelle/HOL (continued from Fig. 2)

play an important moral role in his argument. After all, God, being infinitely good (i.e. perfect), is defined by both Leibniz and Gödel as having all positive properties (perfections).

Premise A4 in Gödel's argument (in logic KB) in fact implies the following: "Non-positive properties are necessarily non-positive"; cf. line 37 in Fig. 2. After some moral paraphrasing, we could thus read A4 and this corollary as: "Virtues are necessarily virtuous" and "Non-virtues are necessarily non-virtuous".

Assuming as an implicit premise that non-positive properties are negative, i.e., that a non-virtue is a vice:

D4 X is a negative property := X is not a positive property,

we obtain as a corollary that "Negative properties are necessarily negative"; cf. corollary C3 in line 39 in Fig. 3. Modulo some moral paraphrasing, this can also be read as "Moral vices are necessarily vicious".

The statements A4 and C3 are equivalent to saying that the (second-order) predicates "is a positive property" and "is a negative property" (being a moral virtue or vice) apply to the same properties in all conceivable situations.[8] As an example, consider the moral virtue of honesty: Can we imagine an alternative 'world' in which being honest is not a virtue? In such a 'world', being dishonest would be a virtue, while being honest would be a vice (assuming Gödel's premise A1). Many different intuitions concerning this question are legitimate and have indeed been defended in the history of philosophy. For instance, (second-order) essentialist positions would answer this question negatively: being a virtue is an essential property of, say, honesty.[9] Others would argue, for instance, that, had the world developed in a quite radically different way (or considering a fundamentally different cultural context), many of the properties we call virtues would not be that virtuous (consider honesty this time in some kind of law-of-the-jungle scenario).

2.6 Positive Properties and Ultrafilters

Gödel's notion of positive properties can be linked to the set theoretical concept of an ultrafilter. Again, automated theorem provers can be employed to explore some interesting results in this regard. Providing such a link to a mathematically well investigated structure may indeed help to further sharpen the understanding of the concept of positive properties.

An ultrafilter U is a special filter[10] on a set X, so that for each subset A of X we have that either A or its complement $X \setminus A$ is an element of U. Ultrafilters are in this sense maximal, they cannot be further refined. In the specific context given here, we are interested in ultrafilters on the powerset of individuals. This is because the notion of positive properties applies to (world-dependent) predicates, i.e. sets, of

[8] Or, in logical jargon, they *denote rigidly*, i.e. the predicates denote the same set of individuals (or in this case concepts) in all possible worlds.

[9] Consider, as an illustration, Plato's Laches, where an explication of the concept of courage is being sought by the participants of the dialogue: several different tentative accounts are rejected because they imply some trait that undermines their pre-understanding of courage as a moral virtue.

[10] A filter, in turn, is a special subset of a partially ordered set. For example, the powerset of some set, partially ordered by set inclusion, is a filter.

individuals.[11] For this special case an ultrafilter U can be defined as follows. Given a set X, an ultrafilter U (on the powerset of X) is a set of subsets of X such that:

- $\emptyset$ (the empty set) is not an element of U.
- If A and B are subsets of X, the set A is a subset of B, and A is an element of U, then B is also an element of U.
- If A and B are elements of U, then so is the intersection of A and B.
- If A is a subset of X, then either A or its relative complement $X \setminus A$ is an element of U.

Note that the properties/predicates A and B in our given context are world-dependent (intensional) concepts, i.e. they are technically defined as binary predicates operating on individuals and possible worlds. When reading these binary predicates as characteristic functions, then, in the light of the above definition, they correspond to sets A and B consisting of pairs of individuals and possible worlds.

This short exhibition of ultrafilters on powersets of pairs of individuals and possible worlds should motivate and explain the definition of an ultrafilter as presented in lines 53 to 61 in Fig. 3 (where sets of pairs are encoded as binary predicates).

The interesting question now is, how Gödel's concept of positive properties relates to this specific notion of an ultrafilter? In Fig. 3 this question is answered in lines 62 to 67. First, in lines 62 to 65, the corresponding ultrafilter properties from above are verified for the notion of positive properties by automated theorem provers. Then in line 67 these results are combined to show that Gödel's notion of positive properties actually constitutes an ultrafilter.

The respective exploration is continued in lines 69 to 84 in Fig. 3. In lines 69 to 73 we introduce the $\downarrow$ operation. This operation is used to "rigidly intensionalise" (possibly) non-rigid intensional predicates/properties φ, by rigidly expanding their extension in the current world. In other words, the meaning of an intensional predicate in a given possible world is conveyed to all other possible worlds. As is verified by the construction of a countermodel in line 74 of Fig. 3, this generally leads to respectively modified, i.e. different, concepts. However, when we accordingly restrict Gödel's notion of positive properties to "rigidly intensionalised" properties only, see line 81, then we can prove that both notions actually coincide. It is thus obvious that the modified notion still constitutes an ultrafilter, see line 83.

3 Anderson's Variant

Anthony Anderson [3, 2] (see also the discussion in [9]) has carried out some reasonable modifications to Gödel's axioms in order to avoid the modal collapse; he also justified these modifications from a theological perspective. His central modification has to do with Gödel's first axiom A1, which postulates the restriction that a property (e.g. honesty) is positive if and only if its negation (e.g. dishonesty) is

[11] In our HOL framework sets (e.g. $\{x \mid x \geq 2\}$) are identified with their characteristic predicates (i.e. $\lambda x.x \geq 2$).

non-positive. Anderson notices that this requirement can be broken down into two different ones:

A1a The negation of a positive property is non-positive (negative) [if a property is positive, then its negation is not positive]

A1b Non-positive (negative) properties are negations of positive properties [if the negation of a property is not positive, then the property is positive]

This splitting of axiom A1 is analogous to what we also employed in line 12 of Fig. 2. However, there we postulated the contrapositives of the above formulations:

```
A1a: "⌊∀X. 𝒫 (⇁X) → ¬(𝒫 X) ⌋" and A1b:"⌊∀X. ¬(𝒫 X) → 𝒫 (⇁X)⌋"
```

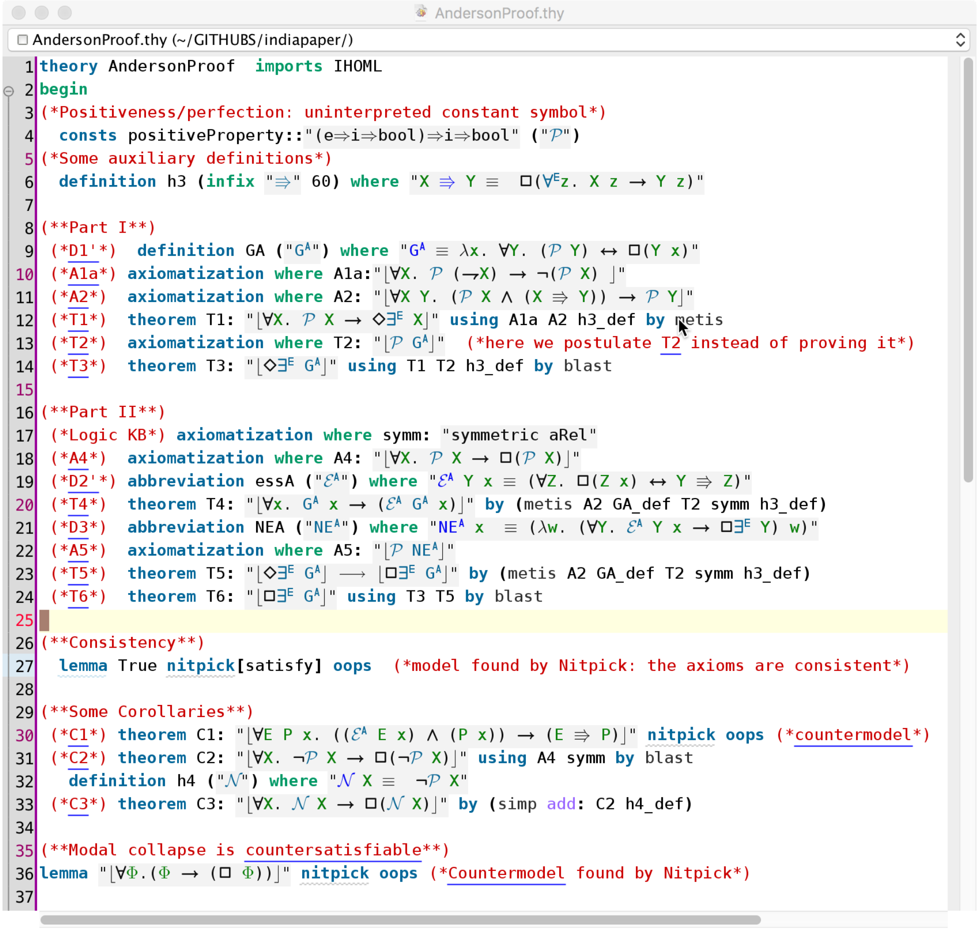
3.1 Indifferent Properties

Anderson argues for keeping axiom A1a while getting rid of axiom A1b. He refers to the work of Chisholm and Sosa's [16], in particular, their concept of "intrinsically good" states of affairs. They propose, amongst other principles, that: "If a state of affairs p is such that p is better than any state of affairs that is indifferent, then p is better than its negation." [16, p. 246]. Anderson notices that by interpreting positive properties in the "moral aesthetic sense" (as arguably intended by Gödel [32]) we can relate *mutatis mutandis* states of affairs consisting of exemplifications of positive properties with "good" states of affairs (i.e. "better" than their negations and better than other indifferent states of affairs). Conversely, we can easily argue for indifferent properties by analogy to the fact that there exist indifferent states of affairs (e.g. an object having a certain color, or a person having some neutral character trait), i.e., states of affairs which are not preferable to their negations and vice versa.

Thus, by dropping axiom A1b from his variant of Gödel's argument, Anderson makes room for "indifferent" properties, that is, properties which are neither positive nor the negation of a positive property. Importantly, his variant still keeps axiom A1a (see line 10 in Fig. 4), hence the relation between positive properties and their negations becomes one-sided: negating a positive property (e.g. honesty) gives us always a non-positive (negative) property (e.g. dishonesty). However, negating a non-positive (negative) property (e.g. laziness) does not always give us a positive property. Furthermore, we get completely 'indifferent' properties if we reject the idea of defining non-positive properties as negative.

3.2 Essence and Godlikeness

Unsurprisingly, eliminating axiom A1b has a negative effect on the argument's validity. In order to render the argument logically valid again, Anderson has to propose some modifications in the formalisation of the other sentences. In particular, he proposes a different definition of essence (which he calls essence*, and which we here call essA, resp. $\mathcal{E}^A$):



```
 1 theory AndersonProof  imports IHOML
 2 begin
 3 (*Positiveness/perfection: uninterpreted constant symbol*)
 4   consts positiveProperty::"(e⇒i⇒bool)⇒i⇒bool" ("𝒫")
 5 (*Some auxiliary definitions*)
 6   definition h3 (infix "⇛" 60) where "X ⇛ Y ≡  □(∀ᴱz. X z ⟶ Y z)"
 7
 8 (**Part I**)
 9  (*D1'*)  definition GA ("Gᴬ") where "Gᴬ ≡ λx. ∀Y. (𝒫 Y) ↔ □(Y x)"
10  (*A1a*) axiomatization where A1a:"⌊∀X. 𝒫 (⇁X) → ¬(𝒫 X) ⌋"
11  (*A2*)  axiomatization where A2: "⌊∀X Y. (𝒫 X ∧ (X ⇛ Y)) → 𝒫 Y⌋"
12  (*T1*)  theorem T1: "⌊∀X. 𝒫 X → ◇∃ᴱ X⌋" using A1a A2 h3_def by metis
13  (*T2*)  axiomatization where T2: "⌊𝒫 Gᴬ⌋"  (*here we postulate T2 instead of proving it*)
14  (*T3*)  theorem T3: "⌊◇∃ᴱ Gᴬ⌋" using T1 T2 h3_def by blast
15
16 (**Part II**)
17  (*Logic KB*) axiomatization where symm: "symmetric aRel"
18  (*A4*)  axiomatization where A4: "⌊∀X. 𝒫 X → □(𝒫 X)⌋"
19  (*D2'*) abbreviation essA ("ℰᴬ") where "ℰᴬ Y x ≡ (∀Z. □(Z x) ↔ Y ⇛ Z)"
20  (*T4*)  theorem T4: "⌊∀x. Gᴬ x → (ℰᴬ Gᴬ x)⌋" by (metis A2 GA_def T2 symm h3_def)
21  (*D3*)  abbreviation NEA ("NEᴬ") where "NEᴬ x  ≡ (λw. (∀Y. ℰᴬ Y x → □∃ᴱ Y) w)"
22  (*A5*)  axiomatization where A5: "⌊𝒫 NEᴬ⌋"
23  (*T5*)  theorem T5: "⌊◇∃ᴱ Gᴬ⌋ ⟶ ⌊□∃ᴱ Gᴬ⌋" by (metis A2 GA_def T2 symm h3_def)
24  (*T6*)  theorem T6: "⌊□∃ᴱ Gᴬ⌋" using T3 T5 by blast
25
26 (**Consistency**)
27   lemma True nitpick[satisfy] oops  (*model found by Nitpick: the axioms are consistent*)
28
29 (**Some Corollaries**)
30  (*C1*) theorem C1: "⌊∀E P x. ((ℰᴬ E x) ∧ (P x)) → (E ⇛ P)⌋" nitpick oops (*countermodel*)
31  (*C2*) theorem C2: "⌊∀X. ¬𝒫 X → □(¬𝒫 X)⌋" using A4 symm by blast
32    definition h4 ("𝒩") where "𝒩 X ≡  ¬𝒫 X"
33  (*C3*) theorem C3: "⌊∀X. 𝒩 X → □(𝒩 X)⌋" by (simp add: C2 h4_def)
34
35 (**Modal collapse is countersatisfiable**)
36 lemma "⌊∀Φ.(Φ → (□ Φ))⌋" nitpick oops (*Countermodel found by Nitpick*)
37
```

Figure 4: Anderson's Emendation of Gödel's Ontological Argument (to be continued in Fig. 5)

D2' A property E is an essence ($\mathcal{E}^A$) of an individual x if and only if all of x's necessary/essential properties are entailed by E and (conversely) all properties entailed by E are necessary/essential properties of x (cf. line 19 in Fig. 4).

The full verification of Anderson's emendation of Gödel's ontological argument in Isabelle/HOL can be found in Fig. 4. Note that the modal collapse is indeed not implied; cf. line 36, where Nitpick reports a countermodel.

Let us compare, for the sake of illustration, Anderson's definition above with Gödel's original one:

D2 A property E is an essence of an individual x if and only if all of x's properties are entailed by E.

As we can see, Anderson's definition D2' is closer to Gödel's in the sense that it does not include Scott's emendation (essences are actually exemplified by their individuals). On the other hand, Anderson's essences no longer characterise individuals completely: an Anderson essence only entails an individual's necessary/essential properties (not contingent/accidental ones). This is illustrated in Fig. 4 in line 30, where a countermodel is reported by the model finder Nitpick to corollary C1.

The above modification in the definition of essence is not the only one needed in order to validate the argument. There is in fact an additional change in nothing less than the definition of being Godlike.

D1' Being Godlike is equivalent to having all and only the positive properties as necessary/essential properties (cf. line 9 in Fig. 4).

Compare this again with Gödel's variant:

D1 Being Godlike is equivalent to having all positive properties.

Anderson argues for his choice in the following way [3]: "Having only positive properties is, I think, too much to ask. Of an indifferent property and its negation God must have one. But having all and only the positive properties as essential properties is plausibly definitive of divinity."

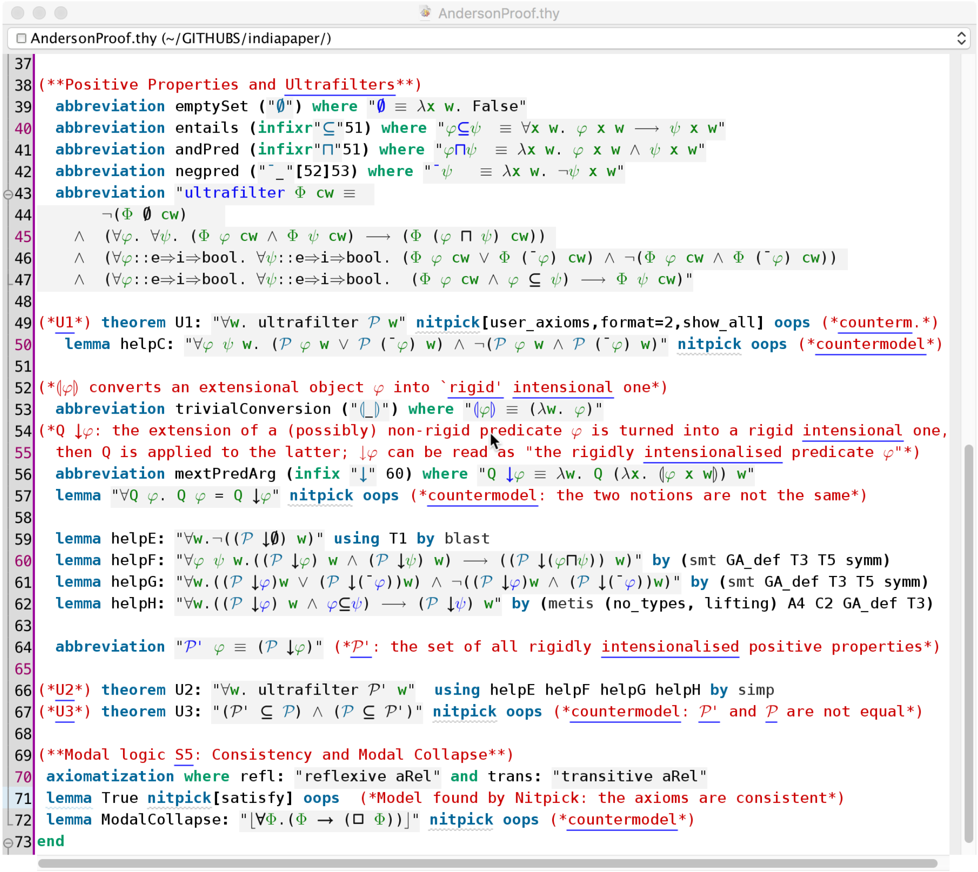
Together, both amended definitions (essence and Godlikeness) suffice to fill the logical vacuum left after weakening axiom A1 (by dropping A1b), thus rendering the argument logically valid again. This, however, comes at the cost of introducing some vagueness in the conception of Godlikeness, since there may exist different individuals exemplifying the property of being Godlike, which only differ in 'indifferent' properties. Such logically informed considerations can provide a starting point for the analysis of further theological concepts like monotheism.

3.3 Modal Logic KB versus Modal Logic S5

We demonstrate here (cf. line 17 in Fig. 4) that Anderson's variant is valid already in modal logic KB, and consequently also in the stronger modal logic S5. The modal collapse is avoided in KB, cf. line 36 in Fig. 4, and also in S5, cf. line 72 in Fig. 5, where respective countermodels are reported by the model finder Nitpick.

3.4 Positive Properties and Ultrafilters

By studying the relation of Anderson's modified notion of positive properties $\mathcal{P}$ to ultrafilters, analogous to what we did for Gödel's variant before, we gain some interesting further insights. Anderson's positive properties, in contrast to Gödel's, no longer constitute an ultrafilter, since the lemma/criterion helpC is now invalidated. This can be seen in lines 49 and 50 in Fig. 5, where countermodels are generated by the model finder Nitpick. However, the set of all rigidly intensionalised extensions of positive properties (in the given state of affairs), encoded as $\mathcal{P}'$, still constitutes an ultrafilter, just as it did in Gödel's variant of the argument; cf. line 66 in Fig. 5. In Anderson's argument, $\mathcal{P}$ and $\mathcal{P}'$ are thus not identical (cf. line 67), and only $\mathcal{P}'$ constitues an ultrafilter.



```

(**Positive Properties and Ultrafilters**)
  abbreviation emptySet ("∅") where "∅ ≡ λx w. False"
  abbreviation entails (infixr"⊆"51) where "φ⊆ψ  ≡ ∀x w. φ x w ⟶ ψ x w"
  abbreviation andPred (infixr"⊓"51) where "φ⊓ψ  ≡ λx w. φ x w ∧ ψ x w"
  abbreviation negpred ("¯_"[52]53) where "¯ψ   ≡ λx w. ¬ψ x w"
  abbreviation "ultrafilter Φ cw ≡
        ¬(Φ ∅ cw)
    ∧  (∀φ. ∀ψ. (Φ φ cw ∧ Φ ψ cw) ⟶ (Φ (φ ⊓ ψ) cw))
    ∧  (∀φ::e⇒i⇒bool. ∀ψ::e⇒i⇒bool. (Φ φ cw ∨ Φ (¯φ) cw) ∧ ¬(Φ φ cw ∧ Φ (¯φ) cw))
    ∧  (∀φ::e⇒i⇒bool. ∀ψ::e⇒i⇒bool.  (Φ φ cw ∧ φ ⊆ ψ) ⟶ Φ ψ cw)"

(*U1*) theorem U1: "∀w. ultrafilter 𝒫 w" nitpick[user_axioms,format=2,show_all] oops (*counterm.*)
   lemma helpC: "∀φ ψ w. (𝒫 φ w ∨ 𝒫 (¯φ) w) ∧ ¬(𝒫 φ w ∧ 𝒫 (¯φ) w)" nitpick oops (*countermodel*)

(*⦇φ⦈ converts an extensional object φ into `rigid' intensional one*)
  abbreviation trivialConversion ("⦇_⦈") where "⦇φ⦈ ≡ (λw. φ)"
(*Q ↓φ: the extension of a (possibly) non-rigid predicate φ is turned into a rigid intensional one,
  then Q is applied to the latter; ↓φ can be read as "the rigidly intensionalised predicate φ"*)
  abbreviation mextPredArg (infix "↓" 60) where "Q ↓φ ≡ λw. Q (λx. ⦇φ x w⦈) w"
  lemma "∀Q φ. Q φ = Q ↓φ" nitpick oops (*countermodel: the two notions are not the same*)

  lemma helpE: "∀w.¬((𝒫 ↓∅) w)" using T1 by blast
  lemma helpF: "∀φ ψ w.((𝒫 ↓φ) w ∧ (𝒫 ↓ψ) w) ⟶ ((𝒫 ↓(φ⊓ψ)) w)" by (smt GA_def T3 T5 symm)
  lemma helpG: "∀w.((𝒫 ↓φ)w ∨ (𝒫 ↓(¯φ))w) ∧ ¬((𝒫 ↓φ)w ∧ (𝒫 ↓(¯φ))w)" by (smt GA_def T3 T5 symm)
  lemma helpH: "∀w.((𝒫 ↓φ) w ∧ φ⊆ψ) ⟶ (𝒫 ↓ψ) w" by (metis (no_types, lifting) A4 C2 GA_def T3)

  abbreviation "𝒫' φ ≡ (𝒫 ↓φ)" (*𝒫': the set of all rigidly intensionalised positive properties*)

(*U2*) theorem U2: "∀w. ultrafilter 𝒫' w"  using helpE helpF helpG helpH by simp
(*U3*) theorem U3: "(𝒫' ⊆ 𝒫) ∧ (𝒫 ⊆ 𝒫')" nitpick oops (*countermodel: 𝒫' and 𝒫 are not equal*)

(**Modal logic S5: Consistency and Modal Collapse**)
 axiomatization where refl: "reflexive aRel" and trans: "transitive aRel"
 lemma True nitpick[satisfy] oops  (*Model found by Nitpick: the axioms are consistent*)
 lemma ModalCollapse: "⌊∀Φ.(Φ → (□ Φ))⌋" nitpick oops (*countermodel*)
end
```

Figure 5: Anderson’s Emendation of Gödel’s Ontological Argument (continued from Fig. 4).

4 Fitting's Variant

Another variant of Gödel's argument that avoids the *modal collapse* has been put forward by Melvin Fitting [19]. He addresses and studies a subtle ambiguity present in Gödel's formal argument, which may seem quite surprising given that it has been constructed using a formal language. This ambiguity (which in fact has already been hinted at in Fig. 3, when we introduced $\mathcal{P}'$), concerns the question whether Gödel's notion of positive properties is in fact intended to apply to extensions or intensions of properties. To better study this difference, Fitting (further) formalises Scott's emendation in an intensional (higher-order) type theory supporting a proper encoding of both alternatives. To this end Fitting's logic of formalisation (in contrast to Gödel's) makes a crucial distinction between two kinds of types: intensional and extensional types.[12] Terms with an intensional type refer to concepts like "the first person on the moon" or "being honest", whereas terms with an extensional type refer to individuals or pluralities (i.e. sets of individuals).[13] The subject(s) which exemplify intensional concepts (Neil Armstrong or the set of all honest people) may differ from one circumstance or possible world to the other; that is, it is easy to conceive an alternative world in which the term "the first person on the moon" refers to someone else than Neil Armstrong (perhaps some Russian cosmonaut). In contrast, the objects referred to by extensional terms are always constituted by the same objects, their size is fixed and their designated subjects are the same in all conceivable circumstances, they are thus said to "designate rigidly". Arguably, proper names and enumerations of individuals belong to the category of extensional terms. Consider for example an alternative world in which John Smith happened to visit a different school in third grade as he actually did. If we want to defend the metaphysical view that in that world that person is still 'our' John Smith (although he exemplifies some different properties) we would be committed to the view that the term "John Smith" designates the person John Smith rigidly.

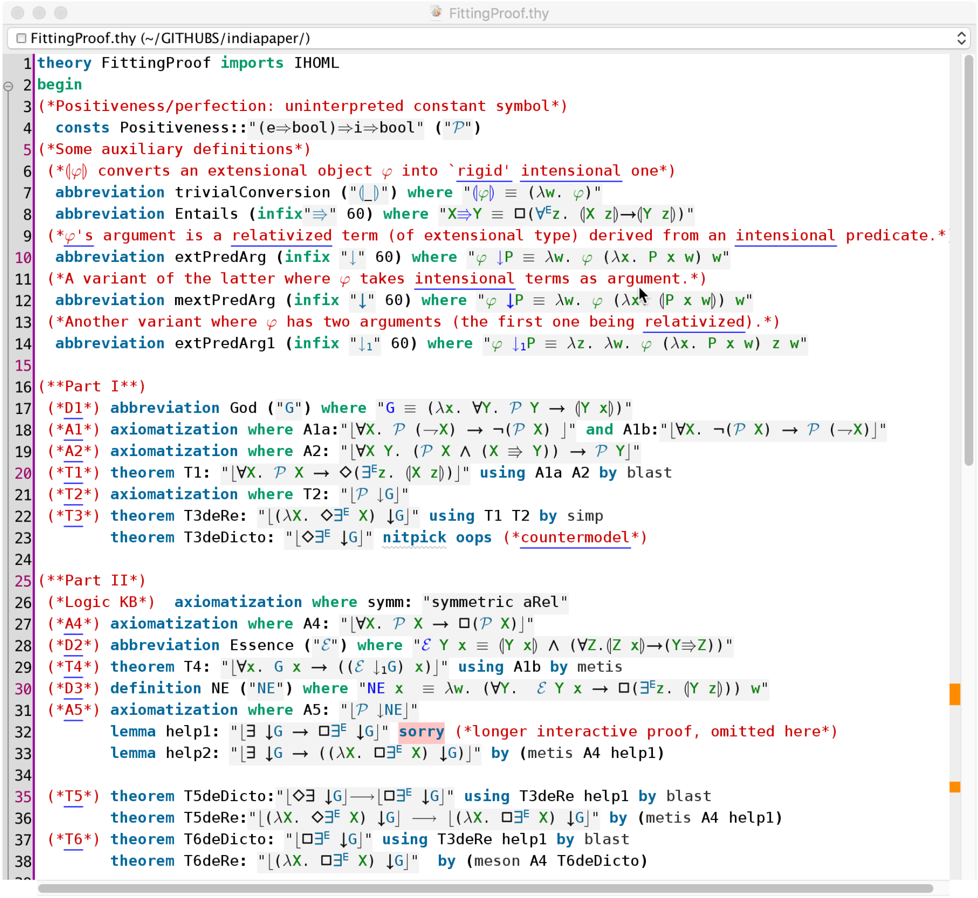
4.1 Positiveness and Essence

In Gödel's variant, positiveness and essence are said to be second-order concepts or properties since they apply to first-order concepts. However, in Fitting's variant they are intended to apply to extensional (and thus rigidly designating) terms.[14]

[12] Actually, it is not entirely fair to say that this ambiguity is inherent to Gödel's formalisation. It is rather a result of (further-)formalising Gödel's argument using Fitting's more complex logical theory. The distinction between intensional and extensional terms Fitting draws his analysis upon did not belong to the logician's theoretical toolbox at the time of Gödel. Drawing on this example, we could argue that no matter how exactly and unambiguously we think we have formalised an argument, it is possible that someone comes in the future and, by (further-)formalising our argument in a more complex logic, draws some quite (at least for us) unexpected conclusions from it.

[13] Fitting's logic also allows for higher-order extensional terms referring to pluralities of pluralities (or sets of sets).

[14] It is difficult to tell what may have been the position of Gödel concerning the purported rigidity of positive or essential properties. From the above discussion, we know that Gödel's axiom A4 entails the rigidity of positiveness (a second-order property). The rigidity of first-order properties is, however, far more philosophically controversial, particularly regarding character traits, since it implies that they are constitutive of the identity of an individual.



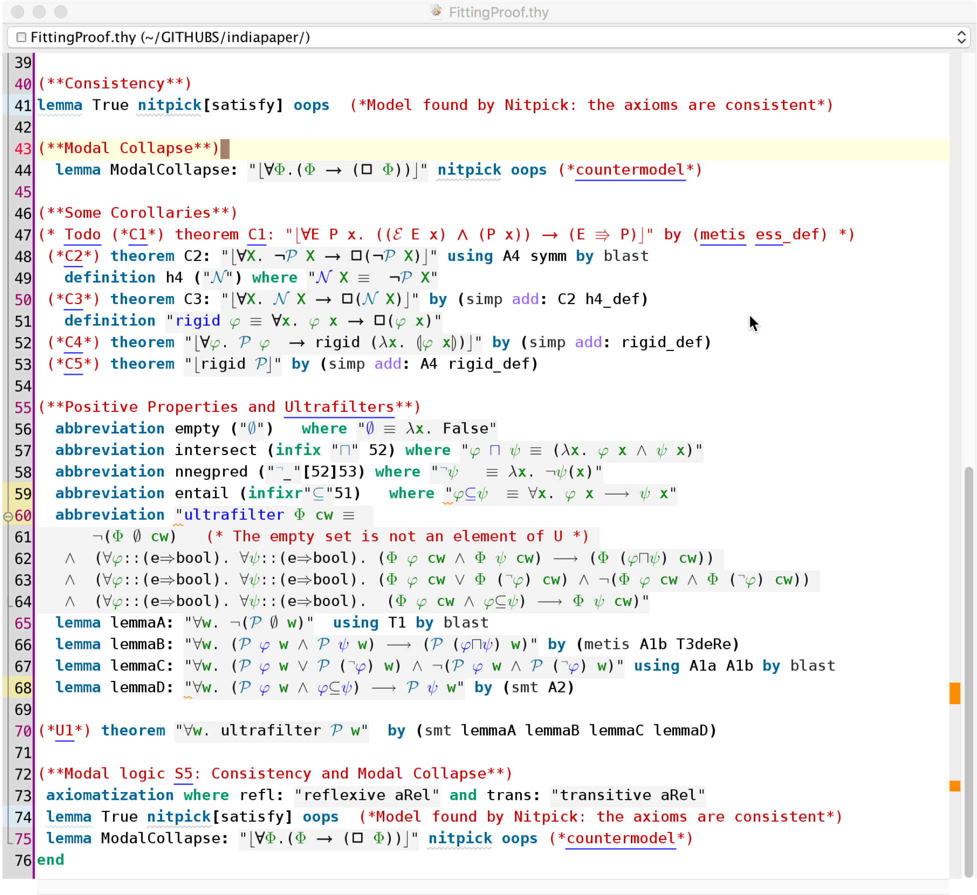
```
theory FittingProof imports IHOML
begin
(*Positiveness/perfection: uninterpreted constant symbol*)
  consts Positiveness::"(e⇒bool)⇒i⇒bool" ("𝒫")
(*Some auxiliary definitions*)
 (*⦇φ⦈ converts an extensional object φ into `rigid' intensional one*)
  abbreviation trivialConversion ("⦇_⦈") where "⦇φ⦈ ≡ (λw. φ)"
  abbreviation Entails (infix"⇛" 60) where "X⇛Y ≡ □(∀ᴱz. ⦇X z⦈→⦇Y z⦈)"
 (*φ's argument is a relativized term (of extensional type) derived from an intensional predicate.*
  abbreviation extPredArg (infix "↓" 60) where "φ ↓P ≡ λw. φ (λx. P x w) w"
 (*A variant of the latter where φ takes intensional terms as argument.*)
  abbreviation mextPredArg (infix "↓" 60) where "φ ↓P ≡ λw. φ (λx. ⦇P x w⦈) w"
 (*Another variant where φ has two arguments (the first one being relativized).*)
  abbreviation extPredArg1 (infix "↓₁" 60) where "φ ↓₁P ≡ λz. λw. φ (λx. P x w) z w"

(**Part I**)
 (*D1*) abbreviation God ("G") where "G ≡ (λx. ∀Y. 𝒫 Y → ⦇Y x⦈)"
 (*A1*) axiomatization where A1a:"⌊∀X. 𝒫 (⇁X) → ¬(𝒫 X) ⌋" and A1b:"⌊∀X. ¬(𝒫 X) → 𝒫 (⇁X)⌋"
 (*A2*) axiomatization where A2: "⌊∀X Y. (𝒫 X ∧ (X ⇛ Y)) → 𝒫 Y⌋"
 (*T1*) theorem T1: "⌊∀X. 𝒫 X → ◇(∃ᴱz. ⦇X z⦈)⌋" using A1a A2 by blast
 (*T2*) axiomatization where T2: "⌊𝒫 ↓G⌋"
 (*T3*) theorem T3deRe: "⌊(λX. ◇∃ᴱ X) ↓G⌋" using T1 T2 by simp
        theorem T3deDicto: "⌊◇∃ᴱ ↓G⌋" nitpick oops (*countermodel*)

(**Part II*)
 (*Logic KB*)  axiomatization where symm: "symmetric aRel"
 (*A4*) axiomatization where A4: "⌊∀X. 𝒫 X → □(𝒫 X)⌋"
 (*D2*) abbreviation Essence ("ℰ") where "ℰ Y x ≡ ⦇Y x⦈ ∧ (∀Z.⦇Z x⦈→(Y⇛Z))"
 (*T4*) theorem T4: "⌊∀x. G x → ((ℰ ↓₁G) x)⌋" using A1b by metis
 (*D3*) definition NE ("NE") where "NE x  ≡ λw. (∀Y.  ℰ Y x → □(∃ᴱz. ⦇Y z⦈)) w"
 (*A5*) axiomatization where A5: "⌊𝒫 ↓NE⌋"
        lemma help1: "⌊∃ ↓G → □∃ᴱ ↓G⌋" sorry (*longer interactive proof, omitted here*)
        lemma help2: "⌊∃ ↓G → ((λX. □∃ᴱ X) ↓G)⌋" by (metis A4 help1)

 (*T5*) theorem T5deDicto:"⌊◇∃ ↓G⌋⟶⌊□∃ᴱ ↓G⌋" using T3deRe help1 by blast
        theorem T5deRe:"⌊(λX. ◇∃ᴱ X) ↓G⌋ ⟶ ⌊(λX. □∃ᴱ X) ↓G⌋" by (metis A4 help1)
 (*T6*) theorem T6deDicto: "⌊□∃ᴱ ↓G⌋" using T3deRe help1 by blast
        theorem T6deRe: "⌊(λX. □∃ᴱ X) ↓G⌋"  by (meson A4 T6deDicto)
```

Figure 6: Fitting’s Emendation of Gödel’s Ontological Argument (to be continued in Fig. 7).



```

(**Consistency**)
lemma True nitpick[satisfy] oops  (*Model found by Nitpick: the axioms are consistent*)

(**Modal Collapse**)
  lemma ModalCollapse: "⌊∀Φ.(Φ → (□ Φ))⌋" nitpick oops (*countermodel*)

(**Some Corollaries**)
(* Todo (*C1*) theorem C1: "⌊∀E P x. ((ℰ E x) ∧ (P x)) → (E ⇛ P)⌋" by (metis ess_def) *)
 (*C2*) theorem C2: "⌊∀X. ¬𝒫 X → □(¬𝒫 X)⌋" using A4 symm by blast
   definition h4 ("𝒩") where "𝒩 X ≡  ¬𝒫 X"
 (*C3*) theorem C3: "⌊∀X. 𝒩 X → □(𝒩 X)⌋" by (simp add: C2 h4_def)
   definition "rigid φ ≡ ∀x. φ x → □(φ x)"
 (*C4*) theorem "⌊∀φ. 𝒫 φ  → rigid (λx. ⦇φ x⦈)⌋" by (simp add: rigid_def)
 (*C5*) theorem "⌊rigid 𝒫⌋" by (simp add: A4 rigid_def)

(**Positive Properties and Ultrafilters**)
  abbreviation empty ("∅")   where "∅ ≡ λx. False"
  abbreviation intersect (infix "⊓" 52) where "φ ⊓ ψ ≡ (λx. φ x ∧ ψ x)"
  abbreviation nnegpred ("⁻_"[52]53) where "⁻ψ   ≡ λx. ¬ψ(x)"
  abbreviation entail (infixr"⊆"51)   where "φ⊆ψ  ≡ ∀x. φ x ⟶ ψ x"
  abbreviation "ultrafilter Φ cw ≡
      ¬(Φ ∅ cw)   (* The empty set is not an element of U *)
    ∧  (∀φ::(e⇒bool). ∀ψ::(e⇒bool). (Φ φ cw ∧ Φ ψ cw) ⟶ (Φ (φ⊓ψ) cw))
    ∧  (∀φ::(e⇒bool). ∀ψ::(e⇒bool). (Φ φ cw ∨ Φ (⁻φ) cw) ∧ ¬(Φ φ cw ∧ Φ (⁻φ) cw))
    ∧  (∀φ::(e⇒bool). ∀ψ::(e⇒bool).  (Φ φ cw ∧ φ⊆ψ) ⟶ Φ ψ cw)"
  lemma lemmaA: "∀w. ¬(𝒫 ∅ w)"  using T1 by blast
  lemma lemmaB: "∀w. (𝒫 φ w ∧ 𝒫 ψ w) ⟶ (𝒫 (φ⊓ψ) w)" by (metis A1b T3deRe)
  lemma lemmaC: "∀w. (𝒫 φ w ∨ 𝒫 (⁻φ) w) ∧ ¬(𝒫 φ w ∧ 𝒫 (⁻φ) w)" using A1a A1b by blast
  lemma lemmaD: "∀w. (𝒫 φ w ∧ φ⊆ψ) ⟶ 𝒫 ψ w" by (smt A2)

(*U1*) theorem "∀w. ultrafilter 𝒫 w"  by (smt lemmaA lemmaB lemmaC lemmaD)

(**Modal logic S5: Consistency and Modal Collapse**)
 axiomatization where refl: "reflexive aRel" and trans: "transitive aRel"
 lemma True nitpick[satisfy] oops  (*Model found by Nitpick: the axioms are consistent*)
 lemma ModalCollapse: "⌊∀Φ.(Φ → (□ Φ))⌋" nitpick oops (*countermodel*)
end
```

Figure 7: Fitting's Emendation of Gödel's Ontological Argument (continued from Fig. 6).

How are we to make sense of this? One option is to say that, in his variant, Fitting takes the properties of essence and positiveness to apply to pluralities or sets of individuals (and no longer to concepts). Another option is to use the notion of a *rigid property*, a property that is exemplified by exactly the same individuals in all possible circumstances. Thus we can say that Fitting's positiveness indeed applies to *rigid* properties.

In the case of essence (as defined by Gödel in D2) it may come as surprise to affirm that only rigid properties can be essences of individuals. Such a move is, however, even more controversial in the case of positiveness. In a nutshell, it implies that anything that is said to be positive in a "moral-aesthetic sense" (like e.g. being honest), is to be treated either as a rigid property or merely as a plurality of individuals. Consequently, those individuals which are said to be, say, honest are so necessarily (given that we take being honest as something positive). In other words (by treating pluralities as rigid properties), if being honest is to be considered a positive property, then we have to accept that, for any actually honest individual x, an alternative world in which x is not honest would be inconceivable (i.e. we take honesty to be an indispensable, identity-constitutive character trait of x).

As shown above, in his effort to avoid the modal collapse ("any individual which has property P, has this property necessarily"), which has been metaphysically interpreted as implying a rejection of free will, Fitting has proposed a (further-)formalisation of Gödel's argument in a more complex logic, which leaves Gödel/Scott's original formalised argument largely unchanged up to an additional (implicit) weaker assumption: that any individual having some positive property (say, P) has this property necessarily; or, interpreted in Gödel's "morally-aesthetic sense" [22], that positive character traits (moral virtues) are identity-constitutive.[15]

4.2 Modal Logic KB versus Modal Logic S5

Also Fitting's variant is valid already in modal logic KB, and consequently also in the stronger modal logic S5. As intended, the modal collapse is avoided, in logic KB and also in S5; cf. lines 44 and 75 in Fig. 7, where countermodels are reported by the model finder Nitpick.

4.3 Positive Properties and Ultrafilters

As already discussed above, the notion of positive properties has been restricted in Fitting's argument to apply to rigid properties only. Technically, this can be seen in line 4 of Fig. 6 by noticing that constant symbol $\mathcal{P}$ now accepts argument properties of type $(e \Rightarrow bool)$, instead of accepting arguments of the richer type $(e \Rightarrow i \Rightarrow bool)$, as done in Gödel's and Anderson's variants. Thus, instead of considering pairs of individuals (terms of type e) and possible worlds (terms of type i) as arguments, the positive properties in Fitting apply to individuals only, i.e., they are independent of possible worlds. This has as a consequence that our notion of ultrafilter, which was defined on powersets of pairs of individuals and possible worlds before, can now be defined simply on powersets of individuals; cf. lines 56 to 64 of Fig. 7. And, as expected, Fitting's notion of positive properties does indeed constitute an ultrafilter on the powerset of individuals (i.e. the set of rigid properties), cf. lines 65 to 70. This relates to the results on $\mathcal{P}'$ in the argument variants of Gödel/Scott and Anderson.

5 Summary and Outlook

Formal logical argumentation supported by modern interactive and automated theorem provers may indeed help to sharpen our conceptual understanding of religious

[15] This kind of (first-order) essentialist ideas about what constitute a person's identity have a long history and have enjoyed some popularity among philosophers, particularly, in the European middle ages. In contemporary mainstream philosophy, however, they are quite unpopular. The fact that they are implied by his emendation of Gödel's argument is indeed never discussed by Fitting in his book [19], so it is not clear if he would have endorsed them.

notions and assumptions. In this paper we have demonstrated this claim by summarising and (re-)formalising the results of various previous experimental studies within a common formal representation framework. Some novel experiments have been addedIn particular, we have shown and illustrated that the ontological arguments by Gödel/Scott, Anderson and Fitting differ in the precise notion of positive properties they assume, and that these differences are well reflected when studying the detailed relations of these notions to the set-theoretic concept of an ultrafilter.

We are convinced that the technology we have employed and demonstrated in this paper can foster a deeper and significantly sharpened mathematical understanding of fundamental concepts and definitions as typically addressed in the domain of ontological arguments. However, methodologically our approach is by no means limited to this domain, and it scales for many similar applications. From the perspective of the topical spectrum of the AISSQ conference, we hope that our technology, in the future, can contribute to a better and more precise understanding of cross-religious theological concepts and foster a better intercultural understanding based on rigorous formalisation, verification, assessment and comparison.

On a more concrete basis, we have verified in this paper that consistent, abstract rational theories for Godlikeness which imply the necessary existence of God, are possible. While the Gödel/Scott variant requires us to accept the modal collapse (determinism/no free will) as a corollary, this is not the case for Anderson's and Fitting's variants. Moreover, neither of the variants by Gödel/Scott, Anderson and Fitting require the use of the (often criticised) strong modal logic S5. As we have shown, all of them are already valid in the weaker modal logic KB.

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] P. Andrews. Church's type theory. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, summer 2018 edition, 2018.

[5] C. Benzmüller. Combining and automating classical and non-classical logics in classical higher-order logic. *Annals of Mathematics and Artificial Intelligence (Special issue Computational Logics in Multi-agent Systems (CLIMA XI))*, 62(1-2):103–128, 2011.

[6] C. Benzmüller and D. Miller. Automation of higher-order logic. In D. M. Gabbay, J. H. Siekmann, and J. Woods, editors, *Handbook of the History of Logic, Volume 9 — Computational Logic*, pages 215–254. North Holland, Elsevier, 2014.

[7] C. Benzmüller and L. Paulson. Multimodal and intuitionistic logics in simple type theory. *The Logic Journal of the IGPL*, 18(6):881–892, 2010.

[8] C. Benzmüller and L. Paulson. Quantified multimodal logics in simple type theory. *Logica Universalis (Special Issue on Multimodal Logics)*, 7(1):7–20, 2013.

[9] C. Benzmüller, L. Weber, and B. Woltzenlogel Paleo. Computer-assisted analysis of the Anderson-Hájek controversy. *Logica Universalis*, 11(1):139–151, 2017.

[10] 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.

[11] C. Benzmüller and B. Woltzenlogel Paleo. The inconsistency in Gödel's ontological argument: A success story for AI in metaphysics. In S. Kambhampati, editor, *IJCAI 2016*, volume 1-3, pages 936–942. AAAI Press, 2016.

[12] C. Benzmüller and B. Woltzenlogel Paleo. An object-logic explanation for the inconsistency in Gödel's ontological theory (extended abstract, sister conferences). In M. Helmert and F. Wotawa, editors, *KI 2016: Advances in Artificial Intelligence, Proceedings*, LNCS, pages 244–250, Berlin, Germany, 2016. Springer.

[13] C. Benzmüller and B. Woltzenlogel Paleo. Experiments in Computational Metaphysics: Gödel's proof of God's existence. *Savijnanam: scientific exploration for a spiritual paradigm. Journal of the Bhaktivedanta Institute*, 9:43–57, 2017.

[14] F. Bjørdal. Understanding Gödel's ontological argument. In T. Childers, editor, *The Logica Yearbook 1998*. Filosofia, 1999.

[15] J. Blanchette and T. Nipkow. Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In *Proc. of ITP 2010*, number 6172 in LNCS, pages 131–146. Springer, 2010.

[16] R. M. Chisholm and E. Sosa. On the logic of "intrinsically better". *American Philosophical Quarterly*, 3(3):244–249, 1966.

[17] D. Davidson. On the very idea of a conceptual scheme. In *Inquiries into Truth and Interpretation*. Oxford University Press, September 2001.

[18] B. Fitelson and E. N. Zalta. Steps toward a computational metaphysics. *Journal of Philosophical Logic*, 36(2):227–247, 2007.

[19] M. Fitting. *Types, Tableaus, and Gödel's God*. Kluwer, 2002.

[20] D. Fuenmayor and C. Benzmüller. Automating emendations of the ontological argument in intensional higher-order modal logic. In *KI 2017: Advances in Artificial Intelligence 40th Annual German Conference on AI, Dortmund, Germany, September 25-29, 2017, Proceedings*, volume 10505 of *LNAI*, pages 114–127. Springer, 2017.

[21] D. Fuenmayor and C. Benzmüller. Types, Tableaus and Gödel's God in Isabelle/HOL. *Archive of Formal Proofs*, 2017. This publication is machine verified with Isabelle/HOL, but comparably mildly human reviewed.

[22] K. Gödel. *Appx.A: Notes in Kurt Gödel's Hand*, pages 144–145. In [32], 2004.

[23] T. F. Godlove, Jr. *Religion, Interpretation and Diversity of Belief: The Framework Model from Kant to Durkheim to Davidson*. Cambridge University Press, 1989.

[24] P. Hájek. A new small emendation of Gödel's ontological proof. *Studia Logica: An International Journal for Symbolic Logic*, 71(2):pp. 149–164, 2002.

[25] G. W. Leibniz. Two notations for discussion with Spinoza. In *Philosophical Papers and Letters*, pages 167–169. Springer, 1989.

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

[27] P. Oppenheimer and E. Zalta. A computationally-discovered simplification of the ontological argument. *Australasian Journal of Philosophy*, 89(2):333–349, 2011.

[28] J. Rushby. The ontological argument in PVS. In *Proc. of CAV Workshop "Fun With Formal Methods"*, St. Petersburg, Russia, 2013.

[29] J. Rushby. A mechanically assisted examination of begging the question in Anselm's Ontological Argument. In *The 2nd World Congress on Logic and Religion*, Warsaw, Poland, June 2017.

[30] D. Scott. *Appx.B: Notes in Dana Scott's Hand*, pages 145–146. In [32], 2004.

[31] J. Sobel. Gödel's ontological proof. In *On Being and Saying. Essays for Richard Cartwright*, pages 241–261. MIT Press, 1987.

[32] J. Sobel. *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge U. Press, 2004.

6 About the Authors

Christoph Benzmüller is a professor at the Freie Universität Berlin (Germany) and a visiting scholar of the University of Luxembourg (Luxembourg). Previous research stations of Christoph include Stanford University (USA), University of Cambridge (UK), Saarland University (Germany), University of Birmingham (UK) and the University of Edinburgh (UK).

Christoph received his PhD (1999) and his Habilitation (2007) in computer science from Saarland University. His PhD research was partly conducted at Carnegie Mellon University (USA). In 2012, Christoph was awarded with a Heisenberg Research Fellowship of the German National Research Foundation (DFG).

David Fuenmayor is a PhD candidate at the Department of Mathematics and Computer Science at the Freie Universität Berlin (Germany). David also carried out his undergraduate studies in Philosophy and Anthropology at this same institution (2015-18).

David's previous studies include an engineer's degree (2009) and M.Sc. (2012) in Mechatronics from the National University of Colombia and the Karlsruhe University of Applied Sciences (Germany) respectively. As an engineer, David is concerned with the critical assessment and responsible application of modern technologies within the framework of philosophy and social sciences.