Computer Modeling in Philosophy

Daniel Kirchner*, Christoph Benzmüller and Edward N. Zalta

Computer Science and Metaphysics: A Cross-Fertilization

https://doi.org/10.1515/opphil-2019-0015
Received April 28, 2019; accepted June 03, 2019

Abstract: Computational philosophy is the use of mechanized computational techniques to unearth philosophical insights that are either difficult or impossible to find using traditional philosophical methods. Computational metaphysics is computational philosophy with a focus on metaphysics. In this paper, we (a) develop results in modal metaphysics whose discovery was computer assisted, and (b) conclude that these results work not only to the obvious benefit of philosophy but also, less obviously, to the benefit of computer science, since the new computational techniques that led to these results may be more broadly applicable within computer science. The paper includes a description of our background methodology and how it evolved, and a discussion of our new results.

Keywords: Computational Metaphysics, Abstract Object Theory, Shallow Semantic Embedding, Gödel's Ontological Argument, Theorem Proving, Automated Reasoning

1 The basic computational approach to higher-order modal logic

The application of computational methods to philosophical problems was initially limited to first-order theorem provers. These are easy to use and have the virtue that they can do proof discovery. In particular, Fitelson and Zalta[1] both (a) used Prover9 to find a proof of the theorems about situation and world theory[2] and (b) found an error in a theorem about Plato's Forms that was left as an exercise in a paper by Pelletier & Zalta.[3] And, Oppenheimer and Zalta discovered,[4] using Prover9, that 1 of the 3 premises used in their reconstruction of Anselm's ontological argument[5] was sufficient to derive the conclusion. Despite these successes, it became apparent that working within a first-order theorem-proving system involved a number of technical compromises that could be solved by using a higher-order system. For example, in order to represent modal claims, second-order quantification, and schemata, etc., in Prover9, special techniques must be adopted that force formulas which are naturally expressed in higher-order systems into the less-expressive language of multi-sorted first-order logic. These techniques were discussed in the papers just

1 Fitelson & Zalta, "Steps toward a computational metaphysics".
2 Zalta, "Twenty-five basic theorems".
3 Pelletier & Zalta, "How to say goodbye to the Third Man".
4 Oppenheimer & Zalta, "A computationally-discovered simplification of the ontological argument".
5 Oppenheimer & Zalta, "On the logic of the ontological argument".

***Corresponding author: Daniel Kirchner,** Freie Universität Berlin, Berlin, Germany; E-Mail: daniel@ekpyron.org
Christoph Benzmüller, Freie Universität Berlin, Berlin, Germany / University of Luxembourg, Luxembourg; E-Mail: c.benzmueller@fu-berlin.de
Edward N. Zalta, Stanford University, Stanford, United States of America; E-Mail: zalta@stanford.edu

mentioned and outlined in some detail in a paper by Alama, Oppenheimer, and Zalta.[6] The representations of expressive, higher-order philosophical claims in first-order logic is therefore not the most natural; indeed, the complexity of the first-order representations grows as one considers philosophical claims that require greater expressivity. And this makes it more difficult to understand the proofs found by the first-order provers.

1.1 The move to higher-order systems

Ways of addressing such problems were developed in a series of papers by Benzmüller and colleagues. Using higher-order theorem provers, they brought computational techniques to the study of philosophical problems and, in the process, they (along with others) developed two methodologies:

- Using the syntactic capabilities of a higher-order theorem prover such as Isabelle/HOL (a) to represent the semantics of a target logic and (b) to define the original syntax of the target theory within the prover. We call this technique Shallow Semantic Embeddings (SSEs).[7] These SSEs suffice for the implementation of interesting modal, higher-order, and non-classical logics and for the investigation of the validity of philosophical arguments. By proving that the axioms or premises of the target system are true in the SSE, one immediately has a proof of soundness of the target system.
- Developing additional abstraction layers to represent the deductive system of philosophical theories with a reasoning system that goes beyond the deductive systems of classical modal logics.

Early papers focused on the development of SSEs. These papers show that the standard translation from propositional modal logic to first-order logic can be concisely modelled (i.e., embedded) within higher-order theorem provers, so that the modal operator $\Box$, for example, can be explicitly defined by the λ-term $\lambda\varphi.\lambda w.\forall v.(Rwv \rightarrow \varphi v)$, where R denotes the accessibility relation associated with $\Box$. Then one can construct first-order formulas involving $\Box\varphi$ and use them to represent and prove theorems. Thus, in an SSE, the target logic is internally represented using higher-order constructs in an automated reasoning environment such as Isabelle/HOL. Benzmüller and Paulson[8] developed an SSE that captures quantified extensions of modal logic (and other non-classical logics). For example, if $\forall x.\varphi x$ is shorthand in functional type theory for $\Pi(\lambda x.\varphi x)$, then $\Box\forall xPx$ would be represented as $\Box\Pi'(\lambda x.\lambda w.Pxw)$, where Π' stands for the λ-term $\lambda\Phi.\lambda w.\Pi(\lambda x.\Phi xw)$, and the $\Box$ gets resolved as described above.[9]

The SSE technique was also the starting point for a natural encoding of Gödel's modern variant of the ontological argument in second-order S5 modal logic. Various computer formalizations and assessments of recent variants of the ontological argument in higher-order theorem provers emerged

---

6 Alama, Oppenheimer & Zalta, "Automating Leibniz's theory of concepts". For example, to represent the T schema $\Box\phi \rightarrow \phi$, one begins with the intermediate representation $\forall p(\Box p \rightarrow p)$. Then one introduces two sortal predicates `Proposition(x)` and `Point(x)`, where the latter represent the possible worlds. In addition, a distinguished point `W` must be introduced, as well as a truth predicate `True(x,y)`. Then one can represent $\forall p(\Box p \rightarrow p)$ as: `all x (Proposition(x) -> (all y (Point(y) -> True(x,y)) -> True(x,W)))`.

7 This is to be contrasted with a *deep semantic embedding*, in which the syntax of the target language is represented using an inductive data structure (e.g., following the BNF of the language) and the semantics of a formula is evaluated by recursively traversing the data structure. *Shallow* semantic embeddings, by contrast, *define* the syntactic elements of the target logic while reusing as much of the infrastructure of the meta-logic as possible.

8 Benzmüller & Paulson, "Quantified multimodal logics".

9 To see how these expressions can be resolved to produce the right representation, consider the following series of reductions:

$$\begin{aligned}
\Box\forall xPx &\equiv \Box\Pi'(\lambda x.\lambda w.Pxw)\\
&\equiv \Box((\lambda\Phi.\lambda w.\Pi(\lambda x.\Phi xw))\lambda x.\lambda w.Pxw))\\
&\equiv \Box(\lambda w.\Pi(\lambda x.(\lambda x.\lambda w.Pxw)xw))\\
&\equiv \Box(\lambda w.\Pi(\lambda x.Pxw))
\end{aligned}$$

in work by Benzmüller and colleagues. Initial studies[10] investigated Gödel's and Scott's variants of the argument within the higher-order automated theorem prover (henceforth ATP) LEO-II.[11] Subsequent work deepened these assessment studies.[12] Instead of using LEO-II, these studies utilized the higher-order proof assistant Isabelle/HOL,[13] a system that is interactive and supports strong proof automation. Some of these experiments were reconstructed in the proof assistant Coq.[14] Additional follow-up work contributed similar studies[15] and includes a range of variants of the ontological argument proposed by other authors, such as Anderson, Hájek, Fitting, and Lowe.[16] Moreover ultrafilters have been used[17] to study the distinction between extensional and intensional positive properties in the variants of Scott, Anderson and Fitting. This ongoing work will be sketched in Section 2.

The other main technique (i.e., the one in the second bullet point above), was developed by Kirchner and Benzmüller to re-implement, in a higher-order system, the work by Fitelson and Zalta.[18] In order to develop a more general implementation of Abstract Object Theory (henceforth AOT or 'object theory'), it doesn't suffice to just develop an SSE for AOT. The SSE of Gödel's ontological argument relies heavily on the completeness of second-order modal logic with respect to Kripke models. Given these completeness results, the computational analysis at the SSE level accurately reflects what follows from the premises of the argument. Since such completeness results aren't available for AOT with respect to its Aczel models, some other way of investigating the proof system computationally is needed. To address this, Kirchner extended the SSE by introducing the new concept of *abstraction layers*.[19] By introducing an additional proof system (as a higher abstraction layer, on top of the semantic embedding) that involves just the axioms of the target logic, one can do automated reasoning in the target logic without generating artifactual theorems (i.e., theorems of the model that aren't theorems of the target logic), and without requiring the embedding to be complete or even provably complete. So the additional abstraction layer makes interactive and automated reasoning in Isabelle/HOL possible in a way that is independent of the model structure used for the semantic embedding itself. Whereas the SSE serves as a sound basis for implementing the abstract reasoning layer, the embedding with abstraction layers provides the infrastructure for a deeper analysis of the semantic properties of the target logic, such as completeness. We'll expand upon this theme on several occasions below.

$\equiv \quad (\lambda\varphi.\lambda w.\forall v.(Rwv \rightarrow \varphi v))(\lambda w.\Pi(\lambda x.Pxw))$

$\equiv \quad (\lambda\varphi.\lambda w.\Pi(\lambda v.Rwv \rightarrow \varphi v))(\lambda w.\Pi(\lambda x.Pxw))$

$\equiv \quad (\lambda w.\Pi(\lambda v.Rwv \rightarrow (\lambda w.\Pi(\lambda x.Pxw))v))$

$\equiv \quad (\lambda w.\Pi(\lambda v.Rwv \rightarrow \Pi(\lambda x.Pxv)))$

$\equiv \quad (\lambda w.\forall v.Rwv \rightarrow \forall x.Pxv)$

$\equiv \quad (\lambda w.\forall vx.Rwv \rightarrow Pxv)$

Thus, we end up with a representation of $\Box\forall xPx$ in functional type theory.

10 Benzmüller & Woltzenlogel Paleo, "Automating Gödel's ontological proof".

11 Benzmüller, Sultana, Paulson & Theiß, "The higher-order prover LEO-II".

12 Benzmüller & Woltzenlogel Paleo, "The inconsistency in Gödel's ontological argument". Benzmüller & Woltzenlogel Paleo, "Object-logic explanation for the inconsistency in Gödel's ontological theory".

13 Nipkow, Paulson & Wenzel, "Isabelle/HOL".

14 Bertot & Casteran, "Interactive Theorem Proving". Benzmüller & Woltzenlogel Paleo, "Interacting with modal logics in the Coq proof assistant".

15 Benzmüller, Weber & Woltzenlogel Paleo, "Analysis of the Anderson-Hájek controversy". Fuenmayor & Benzmüller, "Automating emendations of the ontological argument". Fuenmayor & Benzmüller, "A case study on computational hermeneutics". Bentert, Benzmüller, Streit & Woltzenlogel Paleo, "Analysis of an ontological proof proposed by Leibniz".

16 Anderson, "Some emendations of Gödel's ontological proof". Anderson & Gettings, "Gödel's ontological proof revisited". Hájek, "Magari and others on Gödel's ontological proof". Hájek, "Der Mathematiker und die Frage der Existenz Gottes". Hájek, "A new small emendation of Gödel's ontological proof". Fitting, "Types, Tableaus, and Gödel's God". Lowe, "A modal version of the ontological argument".

17 Benzmüller & Fuenmayor, "Can computers help to sharpen our understanding of ontological arguments?".

18 Fitelson & Zalta, "Steps toward a computational metaphysics".

19 Kirchner, "Representation and Partial Automation of the PLM in Isabelle/HOL".

Kirchner reconstructs not only AOT's fundamental theorems about possible worlds, but arrives at meta-theorems about the correspondence between AOT's syntactic possible worlds and the semantic possible worlds used in its models.[20] In particular, Kirchner shows, using the embedding of AOT in Isabelle/HOL, that for each syntactic possible world ***w*** of AOT, there exists a semantic possible world *w* in the embedding, such that all and only propositions derivably true in ***w*** (using AOT's definition of truth in possible worlds) are true in *w*, and vice-versa. The shallow semantic embedding with abstraction layers made it possible to reason both *within* the target logic itself (i.e., in the higher-level abstract reasoning layer) and *about* the target logic (i.e., using the outer logic of HOL as metalogic and the embedding as a definition of the semantics of the target logic).

To illustrate, as simply as possible, some of the technical details involved in the two basic techniques described above, we now turn to the development of an SSE for propositional modal logic, and show how an abstraction layer can be added on top of that.

1.2 Propositional S5 with abstraction layers

Our computational method can be illustrated with the simple example of a shallow semantic embedding of a propositional S5 logic with abstraction layers. In order to map modal logic to non-modal higher-order logic we use Kripke semantics. To that end we introduce a (non-empty) domain `i` for possible worlds in terms of a type declaration. In Isabelle/HOL:

typedecl `i`

Then we define a type for propositions, which are represented by functions mapping possible worlds to booleans. The right-hand side of the following in Isabelle/HOL represents the complete set of *all* (`UNIV`) functions from type `i` to type `bool`. This set is used to define a new *abstract* type ○, whose objects are represented by elements of this set.[21]

typedef ○ = *"UNIV::(i⇒bool) set" ..*

Given these definitions we then lift the already defined connectives of our meta-logic HOL to the newly introduced type ○ of propositions in the target logic:

lift_definition `S5_not :: "○ ⇒ ○" ("¬_" [54] 70)`
is *"λp. λw. ¬(p w)"*.

lift_definition `S5_impl :: "○ ⇒ ○ ⇒ ○" (`**infixl** *"→" 51)*
is *"λp. λq. λw. (p w) ⟶ (q w)"*.

The first defines the new operator `S5_not` (with the convenient syntax of a bold negation ¬) on the abstract type ○ using the given λ-function on the representation type (functions from possible worlds to booleans). In *λp. λw. ¬(p w)*, the λ-bound *p* is a function from possible worlds to booleans (type `i⇒bool`) and *w* is a possible world. The λ-term maps *p* and *w* to the negation of: *w* applied to *p*. So it defines a function of type `(i⇒bool)⇒i⇒bool`, which is exactly the representation type of the desired signature ○ ⇒ ○. The implication connective `S5_impl` with syntax → is defined in a similar manner.

Using the same mechanism the unique operators of modal logic can be defined in accordance with their Kripke semantics:

lift_definition `S5_box :: "○ ⇒ ○" ("□_" [62] 63)`
is *"λp. λw. ∀v. p v"*.

20 Kirchner, "Representation and Partial Automation of the PLM in Isabelle/HOL".

21 To introduce a new type the representation set has to be non-empty. The fact that it is non-empty here can be trivially proven, which is indicated by the two dots at the end of the line.

lift_definition `S5_dia :: "o ⇒ o" ("◊_" [62] 63)`
is "$\lambda p.\, \lambda w.\, \exists v.\, p\, v$".

To formulate statements about our newly defined target logic, we still have to define what it means for a formula of the target logic to be *valid*. There are two options for defining validity, i.e., either as truth relative to a designated actual world or as truth in all possible worlds. Thus, to define validity, we first need a definition of truth relative to a possible world:

lift_definition `S5_true_in_world :: "i ⇒ o ⇒ bool" ("[_ ⊨ _]")`
is "$\lambda p.\, \lambda w.\, p\, w$".

It turns out that this is sufficient for reasoning about validity; so we don't need to choose between the following two alternative definitions[22] of global validity:

lift_definition `S5_valid_nec :: "o ⇒ bool" ("□[_]")`
is "$\lambda p.\, \forall w.\, p\, w$" .

consts w_0 `:: i`

lift_definition `S5_valid_act :: "o ⇒ bool" ("𝒜[_]")`
is "$\lambda p.\, p\, w_0$".

What we have so far is a shallow semantical embedding of an S5 modal logic, implemented using an *abstract type* in Isabelle/HOL.[23] We can already formulate and prove statements in our target logic at this stage by initiating what Isabelle/HOL calls a *transfer* of a given statement to its counterpart with respect to the representation types, in accordance with the lifting definitions.[24] So to prove the K◊-lemma, we simply give the following command:

lemma "$[w \models \Box(p \to q) \to (\Diamond p \to \Diamond q)]$"
apply `transfer` **by** `auto`

The proof of this lemma uses the `transfer` method and is shown to be valid in the semantics, so the proof doesn't reveal which particular axioms, or axiom system, of S5 are needed to derive it in the traditional sense. In the case of propositional S5 modal logic this doesn't constitute a problem, since it is known that it is *complete* with respect to Kripke semantics. So everything that is derivable from the semantics will also be derivable from the standard axioms of S5. However, for more complex target systems like AOT, this is not the case *a priori*.

We could, at this point, show how the additional abstraction layers needed for the proof theory of AOT can be developed, but that would introduce complexity that isn't really needed for this discussion. So, instead, we shall illustrate how an abstraction layer can be added to the above SSE for propositional modal logic. For the remainder of this section, then, we proceed as if the completeness results for Kripke semantics aren't known. For the analysis of the proof theory of propositional S5 logic, the first step is to simultaneously show that the system of propositional S5 logic is *sound* with respect to our semantics and construct the basis of our abstraction layer by deriving the standard S5 axioms from the semantics:

lemma `ax_K:` "$[w \models \Box(p \to q) \to (\Box p \to \Box q)]$"
apply `transfer` **by** `auto`
lemma `ax_T:` "$[w \models \Box p \to p]$"
apply `transfer` **by** `auto`

22 It has been argued that the second option is more philosophically correct, see Zalta, "Logical and analytic truths that are not necessary".

23 For the shallow semantical embedding alone, we could have skipped the introduction of a new abstract type `o`, but instead used the representation type `i⇒bool` directly in the definitions; however, using the abstract type makes it easier to introduce the *abstraction layer* in the following paragraphs.

24 Huffman & Kuncar, "Lifting and transfer".

lemma *ax_5:* ‟[w ⊨ ◊p → □◊p]″
apply *transfer* **by** *auto*

Furthermore, we need axioms for the classical negation and implication operators, e.g.,

lemma *ax_pl_1:* ‟[w ⊨ p → (q → p)]″
apply *transfer* **by** *auto*

and so on for the other axioms of propositional logic.

Next, we need to derive the two inference rules, i.e., modus ponens and necessitation.

lemma *mp:* **assumes** "[w ⊨ p]" **and** ‟[w ⊨ p → q]″
shows ‟[w ⊨ q]″
using *assms* **apply** *transfer* **by** *auto*

lemma *necessitation:* **assumes** ‟∀v. [v ⊨ q]″
shows ‟[w ⊨ □p]″
using *assms* **apply** *transfer* **by** *auto*

Unfortunately, in our implementation we are lacking *structural induction*, i.e., induction on the complexity of a formula. For that reason, we also have to derive meta-rules for our target system from the semantics, e.g.,

lemma *deduction:* **assumes** ‟[w ⊨ p] ⇒ [w ⊨ q]″
shows ‟[w ⊨ p → q]″
using *assms* **apply** *transfer* **by** *auto*

Together, the axioms, the inference rules and our meta-rules now constitute the abstraction layer in our embedding. Subsequent reasoning can be restricted so that it doesn't use the semantic properties of the embedding (i.e., so that it won't transfer abstract types to representation types or unfold the semantic definitions). In this way, proofs can be constructed that only rely on the axioms and rules themselves.

Given a sane choice of inference rules and meta-rules, every theorem derived in this manner is guaranteed to be derivable from the axiom system. While simple propositional S5 modal logic is known to be complete with respect to its semantic representation, one can still construct abstraction layers to reproduce and analyze the deductive reasoning system of a particular formalization of S5. The abstraction layer can help a user in interactive reasoning, since it enables the same mode of reasoning as the target system with identical rules. More generally, whenever the focus of an investigation is derivability rather than semantic truth,[25] introducing abstraction layers is either necessary (if there are no completeness results) or at least helpful (even if there are completeness results), since they alleviate the need for a translation process from semantic facts to actual derivations.

A reasonable analysis of AOT is not possible without abstraction layers. For one, AOT is more expressive than propositional modal logic and uses foundations that are fundamentally different from HOL.[26] Therefore a representation of the semantics and a model structure of AOT in HOL is more complex and it becomes more difficult to reason about AOT solely by unfolding semantic definitions. Furthermore, there are as yet no results about the completeness of the canonical *Aczel models* of AOT, so there is no guarantee that theorems valid in the semantic embedding are in fact derivable using the axioms and derivation system of AOT itself. Lastly, although the original motivation for constructing an SSE of AOT was mainly to investigate the feasibility of a translation between functional and relational type theory and to gain insights about possible models of the theory, it turned out that an SSE with abstraction layers can be used as a means to analyze the effects of variations in the axiomatization of AOT itself. This led to an evolution of AOT, parts of which are described in Section 3.

25 This is generally the case when investigating an entire logical theory, rather than a logical argument that might not even specify a fixed axiom system against which it is formulated.

26 AOT is formulated in relational type theory, whereas HOL is based on functional type theory. A translation between the two is known to be challenging, e.g. see Oppenheimer & Zalta, "Relations versus functions".

The remainder of the paper is structured as follows. In Section 2, we explain how the SSE technique led to insights about Gödel's ontological argument. In Section 3, we discuss the various insights into AOT that emerged as a result of the addition of the abstraction layer to the SSE for AOT. Finally, in Section 4, we discuss how our techniques may be generalized, and how cross-pollination between computer science and philosophy works to the benefit of both disciplines.

2 Implementation of Gödel's ontological argument

This section outlines the results of a series of experiments in which the SSE approach was successfully utilized for the computer-supported assessment of modern variants of the ontological argument for the existence of God. The first series of experiments, conducted by Benzmüller and Woltzenlogel-Paleo, focused on Gödel's higher-order modal logic variant,[27] as emended by Dana Scott[28] and others; the detailed results were presented in the literature.[29] This work had a strong influence on the research mentioned above, since its success motivated the question of whether the SSE approach would eventually scale for more ambitious and larger projects in computational metaphysics. The computer-supported assessments of Gödel's version of the ontological argument and its variants revealed several novel findings some of which will be outlined below.

2.1 Inconsistency and other results about Gödel's argument

In the course of experiments,[30] the theorem prover Leo-II detected that the *unedited* version of Gödel's formulation of the argument[31] was inconsistent, and that the emendation introduced by Scott[32] while transcribing the original notes was essential to preserving consistency. The Scott version was verified for logical soundness in the interactive proof assistants Isabelle/HOL[33] and Coq.[34] In Figures 1 and 2, the axioms causing the inconsistency in Gödel's manuscript are highlighted. The inconsistency, which was missed by philosophers, is explained in detail in related publications.[35]

The problem Gödel introduced in his *scriptum*[36] is that *essence* is defined as:

- A property Y is the *essence* of an individual x iff all of x's properties are entailed by Y, i.e., iff $\forall Z(Zx \rightarrow Y \Rightarrow Z)$,

where $Y \Rightarrow Z$ means that $\Box\forall x(Yx \rightarrow Zx)$. We'll see below that this definition doesn't require that an individual x exemplify its essence, something we would intuitively expect of the notion of an essence. Scott, in contrast, added a conjunct to the definition of essence:

- A property Y is the *essence* of an individual x iff x has property Y and all of x's properties are entailed by Y.

This simple emendation by Scott preserved consistency of the axioms Gödel introduced as premises of the argument.

27 Gödel, "Appendix A. Notes in Kurt Gödel's Hand".
28 Scott, "Appendix B: Notes in Dana Scott's Hand".
29 Benzmüller & Woltzenlogel Paleo, "The inconsistency in Gödel's ontological argument". Benzmüller & Woltzenlogel Paleo, "Automating Gödel's ontological proof". Benzmüller, Weber & Woltzenlogel Paleo, "Analysis of the Anderson-Hájek controversy". Fuenmayor & Benzmüller, "Types, Tableaus and Gödel's God".
30 Benzmüller & Woltzenlogel Paleo, "Automating Gödel's ontological proof".
31 Gödel, "Appendix A. Notes in Kurt Gödel's Hand".
32 Scott, "Appendix B: Notes in Dana Scott's Hand".
33 Nipkow, Paulson & Wenzel, "Isabelle/HOL".
34 Bertot & Casteran, "Interactive Theorem Proving".
35 Benzmüller & Woltzenlogel Paleo, "Object-logic explanation for the inconsistency in Gödel's ontological theory". Benzmüller & Woltzenlogel Paleo, "The inconsistency in Gödel's ontological argument".
36 Gödel, "Appendix A. Notes in Kurt Gödel's Hand".

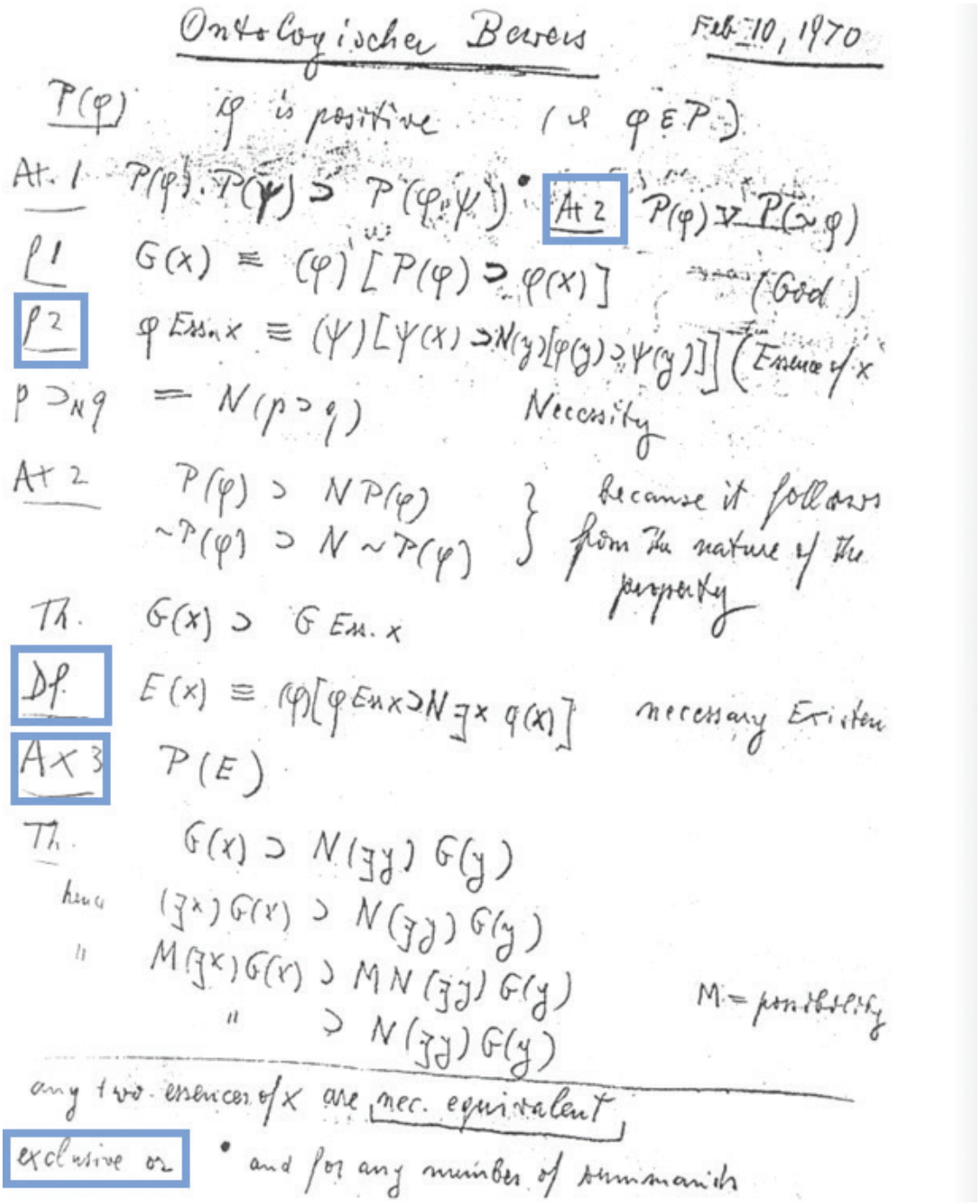

Figure 1: Page 1 of Gödel's Manuscript. The axioms causing the inconsistency in Gödel's modal logic variant of the ontological argument for the existence of God are highlighted (by us) in blue. (Unpublished works of Kurt Gödel are Copyright Institute for Advanced Study and are used with permission. All rights reserved by Institute for Advanced Study.)

The inconsistency in Gödel's original version already appears when the argument is formulated in the quantified modal logic K (with and without the Barcan formulas), and thus also appears in the stronger logics KB and S5, which are both extensions of K.[37] By proving a simple lemma, one can demonstrate how the inconsistency arises. The simple lemma is:

37 Though in S5, the Barcan formulas are derivable.

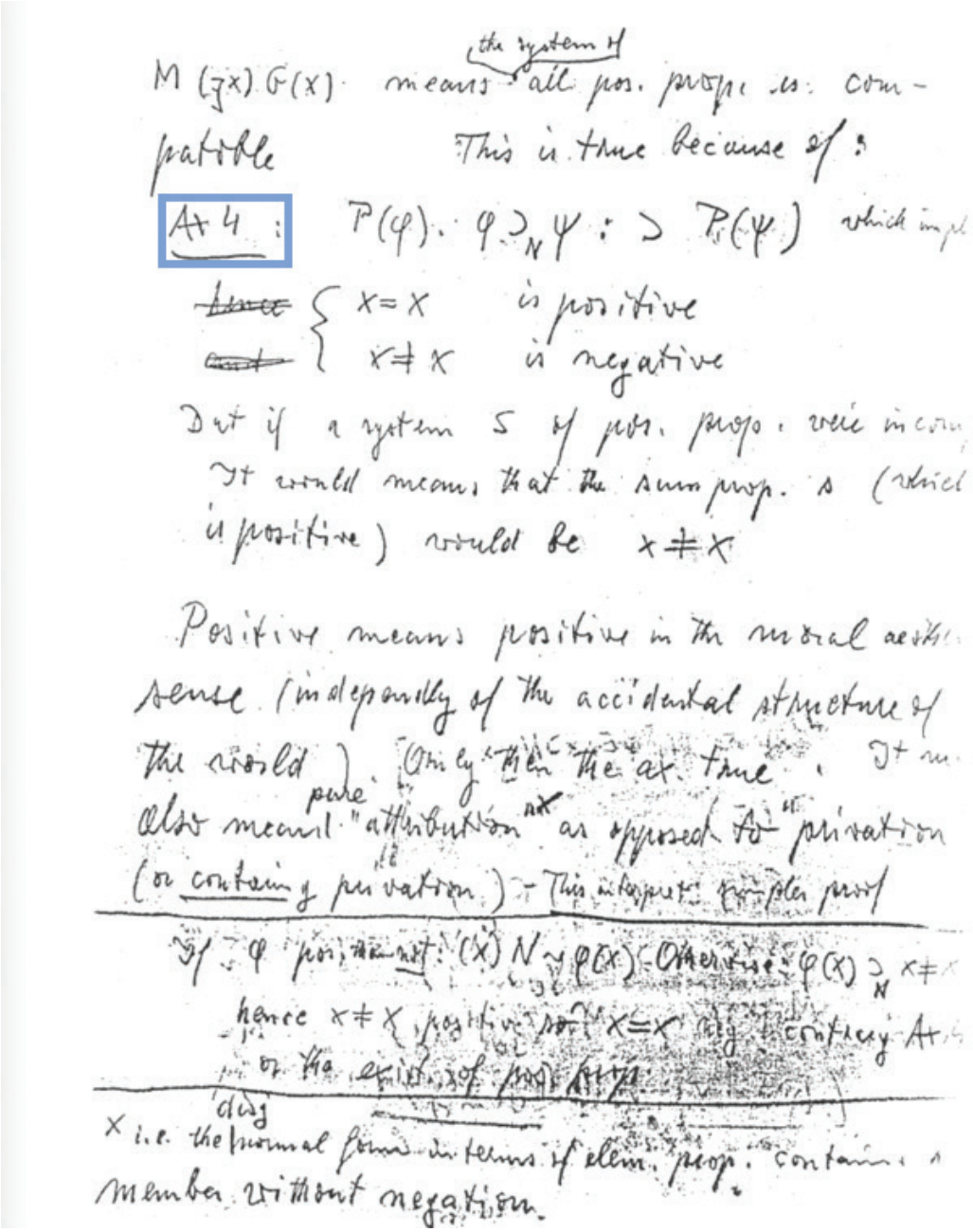

Figure 2: Page 2 of Gödel's Manuscript. The axioms causing the inconsistency in Gödel's modal logic variant of the ontological argument for the existence of God are highlighted (by us) in blue. (Unpublished works of Kurt Gödel are Copyright Institute for Advanced Study and are used with permission. All rights reserved by Institute for Advanced Study.)

EmptyEssenceLemma
An empty property (e.g., being non-self-identical) is an essence of any individual x.

This lemma, in combination with the other highlighted axioms and definitions in Figures 1 and 2, implies a contradiction.[38] The inconsistency was detected automatically by the ATP Leo-II,[39] and in

38 Benzmüller & Woltzenlogel Paleo, "Object-logic explanation for the inconsistency in Gödel's ontological theory".
39 Benzmüller, Sultana, Paulson & Theiß, "The higher-order prover LEO-II".

the course of the proof, it used the fact that an empty property obeys the above lemma to derive the contradiction.[40]

The investigation using ATPs also yielded other noteworthy results:

- it determined which axioms were otiose,
- it determined which properties of the modal operator were required for the argument, and
- it determined that the argument, even as emended by Scott, implies modal collapse, i.e., that $\varphi \equiv \Box\varphi$, so that there can only be models of the premises to the argument in which there is exactly one possible world.

The modal collapse was already noted by Sobel[41] but quickly confirmed by the ATP. One might conclude, therefore that the premises of Gödel's argument imply that everything is determined (we may even say: that there is no free will).

Further variants of Gödel's argument, in which his premises were weakened to address the above issues, were proposed by Anderson, Hájek, Fitting, and Bjørdal.[42] The modal collapse problem was the key motivation for the contributions of Anderson, Hájek, and Bjørdal (and many others), and these have also been investigated computationally.[43] Moreover, ATPs have even contributed[44] to the *clarification of an unsettled philosophical dispute* between Anderson and Hájek. In the course of this work, *different notions of quantification* (*actualist* and *possibilist*) have been utilized and combined within the SSE approach.[45]

2.2 Emendations by Anderson and Fitting

The emendations proposed by C. Anthony Anderson[46] and Melvin Fitting[47] to avoid the modal collapse are rather distinctive and merit special consideration. In order to rationally reconstruct Fitting's argument, an SSE of the richer logic underlying his argument was constructed. This same SSE was used to reconstruct Anderson's argument. By introducing the mathematical notion of an ultrafilter, the two versions of the argument can be compared. This enhanced SSE technique shows that their variations of the argument are closely related.

2.2.1 Anderson's variant

Anderson's central change was to modify a premise that governs the primitive notion of a *positive* property, which was originally governed by the axiom: Y is positive if and only if the negation of Y is non-positive (cf. axiom A2 in Figure 1 where an *exclusive or* is utilized). Anderson suggests that one direction of the biconditional should be preserved, namely, that:

- If a property is positive, then its negation is not positive.

40 It is interesting to note here that during the course of its discovery of the inconsistency, Leo-II engaged in *blind-guessing*. That is, it used a primitive substitution rule to instantiate a predicate quantifier $\forall Y$ with the λ-expression $[\lambda x\ x \neq x]$. This is a method that is *not* unification-based. See Andrews, "On connections and higher-order logic".

41 Sobel, "Gödel's ontological proof". Sobel, "Logic and Theism".

42 Anderson, "Some emendations of Gödel's ontological proof". Anderson & Gettings, "Gödel's ontological proof revisited". Hájek, "Magari and others on Gödel's ontological proof". Hájek, "Der Mathematiker und die Frage der Existenz Gottes". Hájek, "A new small emendation of Gödel's ontological proof". Fitting, "Types, Tableaus, and Gödel's God". Bjørdal, "Understanding Gödel's ontological argument".

43 Benzmüller & Woltzenlogel Paleo, "The modal collapse".

44 Benzmüller, Weber & Woltzenlogel Paleo, "Analysis of the Anderson-Hájek controversy".

45 Benzmüller & Woltzenlogel Paleo, "Higher-order modal logics".

46 Anderson, "Some emendations of Gödel's ontological proof". Anderson & Gettings, "Gödel's ontological proof revisited".

47 Fitting, "Types, Tableaus, and Gödel's God".

As expected, this has an effect on the argument's validity, and in order to render the argument logically valid again, Anderson proposes modifications to premises governing other notions of the argument — in particular, to those governing the definition of essence (which Anderson revises to *essence**) and a modified notion of Godlikeness (*Godlike**):

essence*
A property E is an essence* 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.

Godlike*
An individual x is *Godlike** if and only if all and only the necessary/essential properties of x are positive, i.e., $G^{\star}x \equiv \forall Y\,(\Box Yx \equiv P(Y))$.

These two amended definitions render the argument logically valid again. This was verified computationally.[48] However, the validity comes at the cost of introducing some vagueness in the conception of Godlikeness, since the new definition allows for there being distinct Godlike entities, which differ only by properties that are neither positive nor non-positive.

2.2.2 Fitting's variant

Fitting suggests that there is a subtle ambiguity in Gödel's argument, namely, whether the notion of a *positive property* applies to extensions or intensions of properties. In order to study the difference, Fitting formalizes Scott's emendation in an intensional type theory that makes it possible for him to encode and compare both alternatives. On Fitting's interpretation, the property of *being Godlike* would be represented by the λ-expression $[\lambda x\,\forall Y\,(\mathcal{P}Y \rightarrow Yx)]$, where $\mathcal{P}$ is the second-order property of being a positive property. On Fitting's understanding, the variable Y in the λ-expression ranges over properties whose extensions are fixed from world to world,[49] while $\mathcal{P}$ is a second-order property whose extension among the first-order properties can vary from world to world. Thus, the λ-expression that defines the *being Godlike* is a first-order property whose extension varies from world to world.

In Gödel's original version of the argument, positiveness and essence are second-order properties, but Fitting suggests that the expressions denoting the first-order properties to which positiveness and essence apply are not rigid designators; such expressions might have different extensions at different worlds. So in Fitting's variant, positiveness and essence apply only to the extensions of first-order properties, where the expressions denoting these extensions are rigid designators. If a property is positive at a world w, its extension at every world is the same as its extension at w. If we utilize the notion of a *rigid property*, that is, a property that is exemplified by exactly the same individuals in all possible circumstances, then we can say that, on Fitting's understanding, only rigid properties can be positive.

It should be noted that this technical notion of a positive property departs from the ordinary notion; for example, a property like *being honest* is something a person could have in one world but lack in another, and in those worlds where he or she has that property, it would be considered 'positive' in so far as it is contributory to a good moral character. But, on the above conception, when a property like *being honest* is designated a positive property, then for any actually honest individual x, an alternative world in which x is not honest would be inconceivable (i.e., honesty would be an indispensable, identity-constitutive character trait of x). In this sense, *being self-identical* is a prototypical positive property. By restricting the

48 Fuenmayor & Benzmüller, "Types, Tableaus, and Gödel's God".

49 Thus, the variable Y ranges over properties such that $\forall x(Yx \equiv \Box Yx)$. For example one group of such properties can be defined in terms of an actuality operator: properties of the form $[\lambda x\,\mathcal{A}\varphi]$, i.e., *being an x that is actually such that* φ, satisfy the condition just stated.

notions in Gödel's argument in this way, Fitting thus leaves Scott's variant of Gödel's argument largely unchanged but is able to prevent the modal collapse. This was confirmed computationally.[50]

2.3 Assessment and comparison using ultrafilters

These emendations proposed by Anderson and Fitting were further investigated and assessed computationally[51] by extending the SSE approach in the spirit of Fitting's book. Experiments using Isabelle/HOL that interactively call the model finder Nitpick confirm that the formula expressing modal collapse is not valid. The ATPs were still able to find proofs for the main theorem not only in S5 modal logic but even in the weaker logic KB.

In order to compare all the variant arguments by Scott, Anderson, and Fitting, the notion of an ultrafilter was formalised in Isabelle/HOL. On the technical level, ultrafilters were defined on the set of rigid properties, and on the set of non-rigid, world-dependent properties. Moreover, in these formalizations of the variants, a careful distinction was made between the original notion of a positive property ($\mathcal{P}$) that applies to (intensional) properties and a restricted notion of a positive property ($\mathcal{P}'$) that applies to the rigidified extensions of properties that would otherwise count as positive. Using these definitions, the following results were proved computationally:

- In Scott's variant both $\mathcal{P}$ and $\mathcal{P}'$ coincide, and both have ultrafilter properties.
- In Anderson's variant $\mathcal{P}$ and $\mathcal{P}'$ do not coincide, and only $\mathcal{P}'$ constitutes an ultrafilter.
- In Fitting's variant, $\mathcal{P}$ is not considered an appropriate notion and so not defined. However, $\mathcal{P}'$ is an ultrafilter.

Our computational experiments thus reveal an intriguing correspondence between the variants of the ontological argument by Anderson and Fitting, which otherwise seem quite different. The variants of Anderson and Fitting require that only the restricted notion of a positive property is an ultrafilter.

2.4 Future research

The above insights suggest an alternative approach to the argument, namely, one that starts out with semantically introducing $\mathcal{P}$ or $\mathcal{P}'$ as ultrafilters and then reconstructs variants of the formal argument on the basis of this semantics. This could lead to an alternative reconstruction in which some of the axioms of the variants described above could be derived as theorems.

The experimental setup described above also provides a basis for interesting research about how to prove that there is a *unique* object that exemplifies the property *being God*. Gödel's original premise set guarantees that there is a unique such object, but on pain of modal collapse. The emendations prevent the modal collapse but at the loss of a unique object that exemplifies *being God*. So it is important to study how various notions of equality in the context of the various logical settings described above might help one to restore uniqueness. One particular motivation is to assess whether different notions of equality do or don't yield monotheism. Formal results about this issue would be of additional interest theologically.

3 Implementation of Object Theory

Whereas the last section described the analysis of *philosophical arguments* using plain SSEs without abstraction layers, the focus of this section is the analysis of full *philosophical theories* by using SSEs

50 Fuenmayor & Benzmüller, "Types, Tableaus, and Gödel's God".

51 Benzmüller & Fuenmayor, "Can computers help to sharpen our understanding of ontological arguments?".

with abstraction layers. Section 1.2 already illustrated a simple example of this. We now examine a more complicated case, namely, the analysis of AOT.

Though AOT has been developing and evolving since its first publication,[52] the basic idea, namely, of distinguishing a new mode of predication and postulating a plenitude of abstract objects using the new mode of predication, has remained constant. In all the publications on object theory, we find a language containing the new mode of predication '*x encodes F*' ('xF'), in which F is a 1-place predicate. This new mode extends the traditional second-order modal predicate calculus, which is based on a single mode of predication, namely, $x_1,\ldots,x_n$ *exemplify* F^n ('$F^n x_1 \ldots x_n$'). The resulting language allows complex formulas built up from the two modes of predication and the system allows the two modes to be completely independent of one another (neither $xF \rightarrow Fx$ nor $Fx \rightarrow xF$ is a theorem).

Using such a language (extended to include definite descriptions and λ-expressions), the basic definitions and axioms of AOT have also remained constant. If we start with a distinguished predicate $E!$ to assert *concreteness*, then the basic definitions and axioms of AOT are:

1. Definition: Being ordinary is (defined as) being possibly concrete: $O! = [\lambda x \Diamond E!x]$
2. Axiom: Ordinary objects necessarily fail to encode properties. $O!x \rightarrow \Box\neg\exists F(xF)$
3. Definition: Being abstract is (defined as) not possibly being concrete. $A! = [\lambda x \neg\Diamond E!x]$
4. Axiom: If an object possibly encodes a property it necessarily does. $\Diamond xF \rightarrow \Box xF$
5. Comprehension Schema for Abstract Objects: Where φ is any condition on properties, there is an abstract object that encodes exactly the properties such that φ, i.e., $\exists x(A!x \,\&\, \forall F(xF \equiv \varphi))$, where φ has no free occurrences of x.

In AOT, identity is not a primitive notion, and so the above definitions and axioms are supplemented by a definition of identity for objects and a definition of identity for properties, relations and propositions. The three most important definitions are:

- Objects x and y are identical if and only if either x and y are both ordinary objects that necessarily exemplify the same properties or x and y are both abstract objects that necessarily encode the same properties.
- Properties F and G are identical if and only if they are necessarily encoded by the same objects.
- Propositions p and q are identical just in case the properties $[\lambda x\, p]$ and $[\lambda x\, q]$ are identical.

While this basis has remained stable, other parts of the theory have been developed and improved over the years. The most recent round of improvements, however, has been prompted by computational studies. Some of these improvements have not yet been published. Nevertheless, we'll describe them here.

For example, in earlier versions of object theory, λ-expressions of the form $[\lambda x_1 \ldots x_n\, \varphi]$ in which φ contained encoding subformulas were simply not well-formed. That's because certain well-known paradoxes of encoding could arise for λ-expressions like $[\lambda x\, \exists F(xF \,\&\, \neg Fx)]$.[53] But Kirchner's computational studies[54] showed that unless one is extremely careful about the formation rules, such paradoxes could arise again by constructing λ-expressions in which the matrix φ included descriptions with embedded encoding formulas (encoding formulas embedded in descriptions don't count as subformulas, and thus were allowed). A natural solution to avoid the re-emergence of paradox is to no longer assume that all λ-expressions have a denotation. Given that AOT already included a free logic to handle non-denoting descriptions, its free logic was extended to cover λ-expressions. This allowed us to suppose that λ-expressions are well-formed even if they include encoding subformulas; the paradoxical ones simply don't denote.

Other changes to object theory that have come about as a result of computational studies include:

52 Zalta, "Abstract Objects".

53 An instance of Comprehension for Abstract Objects that asserted the existence of an object that encodes just such a property would provably exemplify that property if and only if it did not.

54 Kirchner, "Representation and Partial Automation of the PLM in Isabelle/HOL".

- the notion of encoding was extended to *n*-ary encoding formulas and these allow one to define the logical notion of term existence directly by way of predication instead of by way of the notion of identity,
- the comprehension principle for propositions has been extended to cover all formulas of the language, so that even encoding formulas denote propositions, and
- the application of AOT to the theory of possible worlds, in which the latter are defined as abstract objects of a certain sort, was enhanced: the fundamental theorem for possible worlds, which asserts that a proposition is necessarily true if and only if it is true at all possible worlds, was extended to cover the new encoding propositions.

These will be explained further below, as we show how AOT was first implemented computationally and how this led to refinements both of the theory and its implementation.

3.1 Construction of an SSE of AOT in Isabelle/HOL

The first SSE of AOT that introduced abstraction layers can be found in Kirchner's work.[55] A detailed description of the structure of this SSE is beyond the scope of this paper, however, we can nevertheless illustrate some of its features and the challenges it had to overcome.

In order to construct an SSE, one has to represent the general model structure of AOT in Isabelle/HOL. The most general models of AOT are *Aczel models*,[56] an enhanced version of which we now describe. Aczel models consist of a domain of *Urelements* that is partitioned into *ordinary* Urelements and *special* Urelements. The ordinary Urelements represent AOT's ordinary objects, whereas the special Urelements will act as proxies for AOT's abstract objects and determine which properties abstract objects exemplify. In addition, a domain of semantic possible worlds (and intensional states) is assumed, and propositions are represented either as intensions (i.e., functions from possible worlds to Booleans) or more generally as hyperintensions (i.e., functions from intensional states to intensions). This way the relations of AOT can be introduced before specifying the full domain of AOT's individuals: AOT's relations can be modeled as functions from Urelements to propositions (as the latter were just represented). Since this already fixes the domain of properties, a natural way to represent AOT's abstract objects that validates their comprehension principle is to model them as sets of properties (i.e., as sets of functions from Urelements to propositions). The domain of AOT's individuals can now be represented by the union of the set of ordinary Urelements and the set of sets of properties. In order to define truth conditions for exemplification formulas involving abstract objects, a mapping σ that takes abstract objects to special Urelements is required.[57] With the help of this proxy function σ, the truth conditions of AOT's atomic formulas can be defined as follows:

- The truth conditions of an exemplification formula $F^n x_1 \ldots x_n$ are determined by the proposition obtained by applying the function used to represent F^n to the Urelements corresponding to $x_1,\ldots,x_n$ (in such a way that when x_i is an abstract object, then its Urelement is the proxy $\sigma(x_i)$). This yields a proposition, which can then be evaluated at a specific possible world (and in the hyperintensional case, at the designated 'actual' intensional state).
- An encoding formula xF is true if and only if x is an abstract object and the function representing F is contained in the set of functions representing x. An ordinary object x does not encode any properties, so all formulas of the form xF are false when x is ordinary.[58]

55 Kirchner, "Representation and Partial Automation of the PLM in Isabelle/HOL".

56 Zalta, "Natural numbers and natural cardinals as abstract objects".

57 Note that for the model to be well-founded, the function σ cannot be injective, i.e., σ must map some distinct abstract objects to the same special Urelement.

58 The truth conditions for *n*-ary encoding formulas $x_1 \ldots x_n F^n$ can be defined on the basis of monadic encoding formulas, but this requires an appeal to the semantics of complex λ-expressions, discussed below. Hence, we omit the discussion of this further development here.

In earlier formulations, AOT relied heavily on the notion of a *propositional* formula, namely, a formula free of encoding subformulas. This notion played a role in relation comprehension: only propositional formulas could be used to define new relations. However, we realized that in the modal version of AOT, encoding formulas are either necessarily true if true or necessarily false if false. This led to the realization that in the models we had constructed, all formulas could be assigned a proposition as denotation; encoding formulas could denote propositions that are necessarily equivalent to necessary truths or necessary falsehoods. As a result, the latest (unpublished) versions of AOT have been reformulated without the notion of a propositional formula and one of the consequences of this move is that comprehension for propositions can be extended to all formulas (this will be discussed further in Section 3.3).

What remains to be defined are the denotations for AOT's complex terms, namely, definite descriptions and λ-predicates. Descriptions may fail to denote and, since AOT follows Russell's analysis of definite descriptions, atomic formulas containing a non-denoting description are treated as false. Therefore, the embedding has to distinguish between the domain of individuals and the domain for individual terms. The latter domain consists of the domain of individuals plus an additional designated element that represents non-denoting terms. If there exists a unique assignment to x for which it holds that φ, the definite description $\iota x\varphi$ denotes this unique object. If there is no unique such object, $\iota x\varphi$ denotes the designated element in the domain of individual terms that represents non-denoting terms. The truth conditions of atomic formulas can now just be lifted to the new domain for terms, with the result that an atomic formula involving the designated element for non-denoting terms becomes false.

In published versions of AOT, every well-formed λ-expression was asserted to have a denotation. However, AOT now allows λ-expressions with encoding subformulas and requires that some of these (in particular, the paradoxical ones) don't denote. Only the λ-expressions that denote are governed by β-Conversion. Nevertheless, every λ-expression has to be interpreted in the model, and the mechanism for doing this is as follows, where we simplify by discussing only the 1-place case and where we suppose that an ordinary object serves as its own proxy. When the matrix φ of the λ-expression $[\lambda x\ \varphi]$ has the same truth conditions for all objects that have the same proxy, one can find a function from Urelements to propositions that, when used to represent $[\lambda x\ \varphi]$, preserves β-Conversion.[59] There is no such function when the matrix has different truth conditions for objects with the same proxy, but these are precisely the matrices for which the λ-expressions provably fail to denote. We interpret these λ-expressions in a manner similar to the interpretation of non-denoting descriptions, namely, by introducing an additional domain for relation *terms* that extends the domain of relations with a designated element for non-denoting terms. Since the condition under which $[\lambda x\ \varphi]$ cannot denote is easy to formulate, namely as $\exists x\exists y(\Box\forall F(Fx \equiv Fy) \wedge \neg(\varphi \equiv \varphi^{y}_{x}))$, such expressions can be mapped to this designated element.

Given the presence of non-denoting descriptions and λ-expressions, AOT extended its free logic for descriptions to cover all complex terms. Note that the axioms of free logic are usually stated in terms of a primitive notion of identity or a primitive notion of existence ($\downarrow$) for terms, so that, for example, the axiom for instantiating terms into universal claims can be stated as one of the following:

1. $\forall\alpha\varphi \rightarrow \left(\exists\beta(\beta = \tau) \rightarrow \varphi^{\tau}_{\alpha}\right)$
2. $\forall\alpha\varphi \rightarrow \left(\tau\downarrow \rightarrow \varphi^{\tau}_{\alpha}\right)$

Normally, these are equivalent formulations, since one usually defines $\tau\downarrow \equiv \exists\beta(\beta = \tau)$.

However, object theory now proves this standard definition as a theorem! As we saw above, it doesn't take identity as a primitive, but rather defines it. Moreover, AOT does not take term existence as primitive

59 For example, if we make use of the λ-expressions of HOL's functional type theory, then we could point to $(\lambda u.\ \exists x.\ u = |x| \wedge \varphi' x)$ as such a function, where (a) φ' is a function that represents the matrix φ of the AOT λ-expression $[\lambda x\ \varphi]$ and maps the bound variable x to a proposition, (b) the type of the bound variable u is the type of Urelements, and (c) $|x|$ is the Urelement corresponding to x.

either, but defines it as well by cases: (a) an individual term κ *exists* ('$\kappa\downarrow$') just in case $\exists F\, F\kappa$, provided F isn't free in κ, and (b) an n-place property term Π *exists* ('$\Pi\downarrow$') just in case $\exists x_1 \ldots \exists x_n\, x_1 \ldots x_n \Pi$, provided no x_i is free in Π, and (c) a 0-place proposition term Π *exists* ('$\Pi\downarrow$') just in case $[\lambda x\ \Pi]\downarrow$, provided x isn't free in Π. Thus, object theory reduces existence to predication, and indeed, given its definitions of identity, reduces identity to predication and existence.[60] Given the foregoing definitions, the claim $\tau\downarrow \equiv \exists\beta(\beta = \tau)$ becomes a theorem.

3.2 Minimal models of second-order modal AOT

Normally, the minimal model for first-order quantified modal logic (QML) contains one possible world and one individual, and in second-order QML, there have to be at least two properties (one true of everything and one false of everything). So, a question arises: what is the most natural axiom that forces the domain of possible worlds to have at least two members (so as to exclude modal collapse), and what effect does that have on the domain of properties? We've discovered that the axiom Zalta has proposed for this job in AOT, namely, the assertion that $\Diamond\exists x(E!x\ \&\ \neg\mathcal{A}E!x)$, not only forces the models to have at least 2 possible worlds, but also a minimum of 4 propositions and (given the actuality operator $\mathcal{A}$ and the comprehension principle for abstract objects) a minimum of 16 properties. Proofs of these facts are available within the system. The latter fact improved upon the original discussion in PLM, which had asserted only that there are at least 6 different properties.[61] But once properties in AOT were modeled in Isabelle/HOL as functions from Urelements and possible worlds to Booleans, it was recognized that there had to be at least 16 of those functions.

The two core axioms that need to be considered for minimal models of AOT are the modal axiom that requires the existence of a contingently nonconcrete object already mentioned above and the comprehension axiom for abstract objects:

- $\Diamond\exists x(E!x\ \&\ \neg\mathcal{A}E!x)$
- $\exists x(A!x\ \&\ \forall F(xF \equiv \varphi))$

where *being abstract* ($A!$) in the second axiom is defined as *not possibly being concrete*, i.e., where $A! = [\lambda x\ \neg\Diamond E!x]$. In particular, the first of these axioms implies:

- $\exists x(\Diamond E!x\ \&\ \neg\mathcal{A}E!x)$

while the second implies:

- $\exists x(\neg\Diamond E!x)$

---

60 To see how the latter comes about (i.e., the reduction of identity to predication and existence), note that in a system like AOT, the definition of property identity stated in the opening paragraphs of Section 3, have to be formalized using metavariables and existence clauses in the definiens, so that we have:

- $\Pi = \Pi' =_{df} \Pi\downarrow\ \&\ \Pi'\downarrow\ \&\ \Box\forall x(x\Pi \equiv x\Pi')$

The metavariables ensure that the definiendum $\Pi = \Pi'$ will be provably false when either Π or Π' is non-denoting. Otherwise one could argue, for non-denoting Π and Π', that both $x\Pi$ and $x\Pi'$ are equivalent (since both are false, given that atomic formulas with non-denoting terms are false), and since this holds for arbitrary x and can be proved without an appeal to contingencies, it follows that $\Box\forall x(x\Pi \equiv x\Pi')$. So without the existence clauses, we could prove that $\Pi = \Pi'$ for any non-denoting terms Π and Π'.

So whereas identity claims in AOT require the existence of the terms flanking the identity sign, this is not required in computational implementations of other interesting logics. For example, Scott introduces both a notion of "identity" and "existing identity", where the latter corresponds to AOT's notion of identity. See Benzmüller & Scott, "Automating free logic in HOL".

61 Originally, Zalta had proved that $E!$ and its negation, $O!$, $A!$, $[\lambda x\ E!x \rightarrow E!x]$ and the latter's negation, all of which exist by comprehension, were distinct properties.

From these consequences it follows that there are at least two distinct individuals; let's call them x_1 and x_2.[62] The proof makes it clear that x_1 is ordinary (i.e., $O!x_1 = [\lambda z \Diamond E!z]x_1$) and x_2 is abstract. But by the construction of Aczel models, the proxy Urelement of the abstract individual can't be the ordinary individual. Urelements in Aczel models determine the exemplification behaviour of individuals. So since x_1 *exemplifies* being ordinary while x_2 does not, x_1 and x_2 have to be mapped to distinct Urelements. Furthermore, the first statement $\exists x(\Diamond E!x \,\&\, \neg\mathcal{A}E!x)$, implies that there are at least two possible worlds in the Kripke semantics, namely, a non-actual world, in which $E!x_1$ holds, and the actual world, in which $E!x_1$ does not hold.

Recall that in our models, relations are represented as functions from Urelements and possible worlds to Booleans.[63] So, by a combinatorial argument from the existence of two possible worlds and two Urelements, we may derive the existence of at least $(2^2)^2 = 16$ relations in the model; each relation has a well-defined and distinct exemplification extension.

However, this doesn't yet show, within the system, that there are at least 16 distinct relations, but only that there are at least 16 distinct relations *in our models* (we don't assume *a priori* that our models are complete). However, we found a proof of the existence of at least 16 relations in AOT and this is now part of PLM.[64]

The foregoing discussion illustrates our research methodology: (1) we constructed a model for the theory and conjectured that it was complete; (2) we then analyzed the features of the model and arrived at statements formulable within the systems AOT and its representation in Isabelle/HOL that should be true given the model; (3) we investigated whether these statements are indeed derivable in AOT (or alternatively, derivable in the abstraction layer of the embedding); and (4) we then concluded either that we had derived a new theorem within these systems or that the model needed to be further refined.

3.3 An extended theory of propositions and worlds

One of the key challenges in constructing the first SSE of AOT was the fact that its syntax relied heavily on the use of the notion of a propositional formula (i.e., formulas with no encoding subformulas). Only propositional formulas were allowed in the construction of n-place complex relation terms for $n \geq 0$. φ and $[\lambda\, \varphi]$ were designated as 0-place relation terms only if φ was a propositional formula. But capturing the notion of a propositional formula in the SSE would have increased its complexity significantly. For example, it would have been necessary to define two versions of every connective and quantifier, one for non-propositional formulas and one for propositional formulas. Instead, the SSE used one type for both kinds of formulas, and thus one kind of connective and quantifier suffices.

---

62 To see this, instantiate these two existential claims using two individual variables x_1 and x_2, such that:

- $\Diamond E!x_1 \,\&\, \neg\mathcal{A}E!x_1$
- $\neg\Diamond E!x_2$

To show that $x_1 \neq x_2$, we need the principle of the substitution of identicals, which is asserted by the following axiom:

- $\alpha = \beta \rightarrow (\varphi \rightarrow \varphi')$, whenever β is substitutable for α in φ, and φ' is the result of replacing zero or more free occurrences of α in φ with occurrences of β.

Now if, for reductio, $x_1 = x_2$, then $\Diamond E!x_1 \rightarrow \Diamond E!x_2$, but since $\neg\Diamond E!x_2$, this cannot be true, hence $x_1 \neq x_2$. This already shows that there are at least *two individuals* in AOT.

63 In hyperintensional models, they additionally depend on an *intensional state*, but since there is only one intensional state in a minimal model, this dependency can be ignored; it doesn't affect the size of the model.

64 The modal axiom $\exists x(\Diamond E!x \,\&\, \neg\mathcal{A}E!x)$ of AOT requires the existence of a contingently false proposition, namely $\exists x(E!x \,\&\, \neg\mathcal{A}E!x)$. Call the false proposition q_0 and its negation $\bar{q}_0$. These propositions (in the form of propositional properties $[\lambda x\, q_0]$ and $[\lambda x\, \bar{q}_0]$) were not considered when it was thought there were at least 6 properties. It turns out that they are provably distinct from the six other properties mentioned above and that combinations (e.g., conjunctions) of these propositional properties and the six properties mentioned above in fact yield 16 properties that are provably distinct in the system and correspond to the 16 properties in the models.

However, from this it became apparent that the models used for the SSE assigned every formula a proposition, including those formulas that contained encoding subformulas. This suggested that AOT could be expanded similarly. Consequently, the comprehension principle for propositions in AOT was revised and expanded in such a way that it has become a theorem that every formula denotes a proposition. And once every formula denotes a proposition, the fundamental theorem of possible worlds becomes naturally extended to cover *all* formulas and not just propositional ones. The fundamental theorem of possible world theory asserts that for every proposition p and every world w, $\Box p \equiv \forall w\,(w \vDash p)$, where $w \vDash p$ asserts that p is true in w (where this, in turn, is cashed out as: w encodes the propositional property $[\lambda x\, p]$). In previous versions of AOT, only propositional formulas could be substituted for p, since only propositional formulas denoted propositions. But once AOT was extended (as a result of our computational investigations), every formula becomes substitutable for p, including those with encoding subformulas.

4 Generalizing the cross-fertilization

As we see it, computer science and related disciplines like philosophy that rely heavily on reasoning and argumentation, benefit from interdisciplinary studies in which computational techniques are applied. Historically, the realization that first-order theorem provers don't capture the higher-order logic of many applied systems created the impetus for the development of systems like Isabelle/HOL. In this paper, we've seen that the requirements for implementing logics and metaphysical theories has led to the development of new methodologies for creating automated reasoning environments for complex systems (e.g., those that are essentially higher-order, non-classical, or have complex terms). This is especially clear in the development of additional abstraction layers in which deductive systems are recaptured so that only the theorems of the target system, and no artifactual theorems, can be discovered computationally. Abstraction layers in turn can be used as a technical tool to analyze properties of the implementation, and in the case of AOT, the completeness of its embedding. In our particular work, not only did the interdisciplinary effort lead to improvements in the computational methodologies used for modeling, but those same methodologies led to improvements in the target metaphysical theory being implemented.

This cross-fertilization methodology can be depicted more generally in Figure 3. In this diagram, the cross-fertilization occurs primarily between the various interactions that the user can have with the front-end systems and applications. Note that Isabelle/HOL integrates state-of-the-art automated reasoning technology and benefits from the constant improvements in all the systems that it integrates.

In the lower left corner of Figure 3, the user is conducting/orchestrating experiments; in this particular case, the application in the lowest blue box is (the metaphysics of) AOT. Since AOT is based on a higher-order modal logic, the computational mechanization of this "target logic" (in the middle blue box) has served as a significant goal. However, at the start of the project, AOT's proof theory wasn't computationally implemented generally. Therefore, the task was to semantically embed the language and theory in HOL (the top blue box), which turned out to be sufficiently expressive as a meta-logic for second-order AOT. A core advantage of this meta-logical approach is that existing reasoning tools for HOL can readily be reused for interactive and automated reasoning in the embedded target logic (the black arrows). This is particularly helpful when the details of a desired language and theory in a given context are not fully determined yet; the methodology enables rapid prototyping of the different ways of formulating the language and axioms of the theory.

Our preferred proof assistant for HOL has been Isabelle/HOL. This system comes with strong user-interaction support, including a configurable user-interface, which, in our context, enables readable surface presentations of the embedded target logic. Equally important is the automation provided by the proof assistants, which include both external ATPs orchestrated by the Sledgehammer tool[65] and automated (counter-)model finding tools like Nitpick[66] and Nunchaku.[67] These systems, in turn, make calls

65 Blanchette, Böhme & Paulson "Extending Sledgehammer with SMT solvers".
66 Blanchette & Nipkow, "Nitpick".
67 Cruanes & Blanchette, "Extending Nunchaku to dependent type theory".

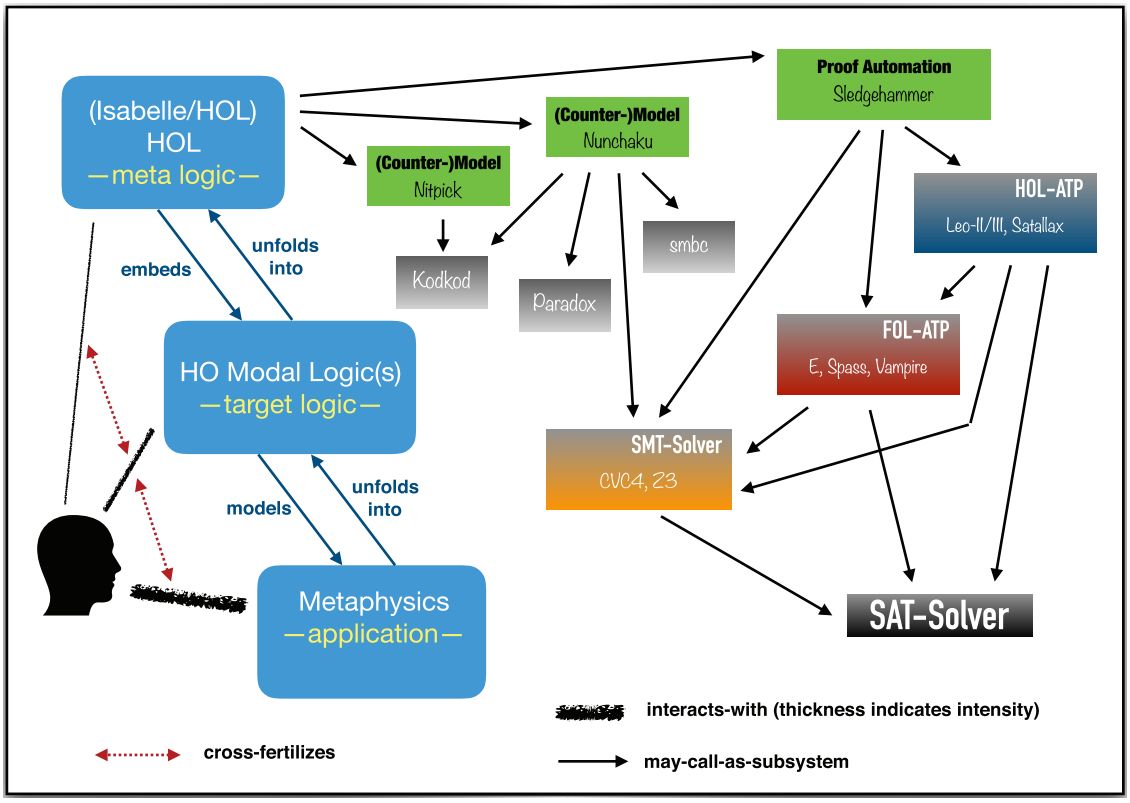

Figure 3: Our general methodology supports the reuse of state of the art theorem proving technology.

to specialist tools such as Kodkod, Paradox, smbc, and the SMT solvers CVC4[68] and Z3.[69] Other systems integrated with Sledgehammer include the first-order ATPs E,[70] Spass,[71] Vampire,[72] and the higher-order ATPs Leo-II,[73] Leo-III,[74] and Satallax.[75] If one downloads Isabelle/HOL, all of these systems are bundled with it, except for the higher-order provers like Leo-II, Leo-III and Satallax, which can be accessed via the TPTP infrastructure using remote calls. These higher-order ATPs internally collaborate in turn with first-order ATPs and SMT solvers. And all these ATPs and SMT solvers internally rely on or integrate state-of-the-art SAT technology. Thus, whenever one of the subsystems improves, the enhancements filter up to the Isabelle/HOL environment. In other words, a proof conjecture in some theory that is not automatically solvable at the present time may well become solvable as improvements to this framework accumulate.

When moving to other application domains (e.g., machine ethics), deontic logics become relevant as target logics. The overall picture stays the same. Only the two lower blue boxes on the left of the Figure change. Note that the combinations of different non-classical logics, e.g., those required for the encoding of the Gewirth principle of generic consistency,[76] can be realized and assessed as targets in this framework.

What has been described above is a generic approach to universal logical and metalogical reasoning[77] based on shallow semantic embeddings in HOL. In addition, the approach also supports the direct

68 Deters, Reynolds, King, Barrett & Tinelli, "A tour of CVC4".
69 Moura & Bjørner, "Z3: An efficient SMT solver".
70 Schulz, "System description: E".
71 Blanchette, Popescu, Wand & Weidenbach, "More SPASS with Isabelle".
72 Kovács & Voronkov, "First-order theorem proving and Vampire".
73 Benzmüller, Sultana, Paulson & Theiß, "The higher-order prover LEO-II".
74 Steen & Benzmüller, "The higher-order prover Leo-III".
75 Brown, "Satallax".
76 Fuenmayor & Benzmüller, "Formalisation and Evaluation of Alan Gewirth's Proof".
77 Benzmüller, "Universal (meta-)logical reasoning".

encoding of a target logic's proof theory of choice. The shallow semantic embedding technique and associated reasoning framework described in the previous sections scale to applications in many other areas, including, for example, mathematical foundations, artificial intelligence and machine ethics. In particular, metalogical investigations are feasible beyond what was considered possible before. In a case study in mathematics, for example, Benzmüller and Scott[78] compared different axiom systems for category theory proposed by MacLane,[79] Scott,[80] and Freyd & Scedrov.[81] This work started with an embedding of free logic in HOL, which was then utilized to encode and assess the different axiom systems. As a side result of the studies, a minor flaw in the work of Freyd and Scedrov was revealed and corrected. Applications in artificial intelligence include the verification of the dependency diagrams of systems in modal logic[82] and an elegant, higher-order encoding of common knowledge (of a group of agents) as part of a solution for the wise men puzzle, a famous riddle in artificial intelligence.[83] A normative-reasoning workbench supporting empirical studies with alternative deontic logics that are resistant to contrary-to-duty paradoxes is currently being developed,[84] and various embeddings of other logics in this area can be found elsewhere.[85] A recent extension and application of this framework[86] demonstrates that even ambitious ethical theories such as Alan Gewirth's principle of generic consistency can be formally encoded and assessed on the computer.

References

Alama, Jesse, Oppenheimer, Paul E., and Zalta, Edward N. "Automating Leibniz's Theory of Concepts." In *Automated Deduction – CADE-25 – 25th International Conference on Automated Deduction, Berlin, Germany, August 1–7, 2015, Proceedings*, edited by Amy P. Felty and Aart Middeldorp, 9195:73–97. Springer: Lecture Notes in Computer Science, 2015.

Anderson, C. Anthony, "Some Emendations of Gödel's Ontological Proof." *Faith and Philosophy* 7:3 (1990), 291–303.

Anderson, C. Anthony, and Gettings, Michael. "Gödel's Ontological Proof Revisited." In *Gödel'96: Logical Foundations of Mathematics, Computer Science, and Physics*, 6:167–172. Springer: Lecture Notes in Logic, 1996.

Andrews, Peter B. "On Connections and Higher-Order Logic." *Journal of Automated Reasoning* 5:3 (1989), 257–291.

Bentert, Matthias, Benzmüller, Christoph, Streit, David and Woltzenlogel Paleo, Bruno. "Analysis of an Ontological Proof Proposed by Leibniz." In *Death and Anti-Death, Volume 14: Four Decades After Michael Polanyi, Three Centuries After G.w. Leibniz*, edited by Charles Tandy. Ria University Press, 2016.

Benzmüller, Christoph. "Universal (Meta-)Logical Reasoning: Recent Successes." *Science of Computer Programming* 172 (2019), 48–62.

Benzmüller, Christoph, Claus, Maximilian, and Sultana, Nik. "Systematic Verification of the Modal Logic Cube in Isabelle/HOL." In *PxTP 2015*, edited by Cezary Kaliszyk and Andrei Paskevich, 186:27–41. EPTCS, 2015.

Benzmüller, Christoph, Farjami, Ali, and Parent, Xavier. "A Dyadic Deontic Logic in HOL." In *Deontic Logic and Normative Systems — 14th International Conference, Deon 2018, Utrecht, the Netherlands, 3-6 July, 2018*, edited by Jan Broersen, Cleo Condoravdi, Shyam Nair, and Gabriella Pigozzi, 33–50. College Publications, 2018.

Benzmüller, Christoph, Farjami, Ali, and Parent, Xavier. "Åqvist's Dyadic Deontic Logic E in HOL." *Journal of Applied Logics*, 2019. To appear, preprint: URL = http://orbilu.uni.lu/handle/10993/37014

Benzmüller, Christoph, and Fuenmayor, David. "Can Computers Help to Sharpen Our Understanding of Ontological Arguments?" In *Mathematics and Reality, Proceedings of the 11th All India Students' Conference on Science & Spiritual Quest (Aissq), 6-7 October, 2018, Iit Bhubaneswar, Bhubaneswar, India*, edited by Sudipto Gosh, Ramgopal Uppalari, K. Vasudeva Rao, Varun Agarwal, and Sushant Sharma, 195–226. The Bhaktivedanta Institute, Kolkata, 2018.

Benzmüller, Christoph, Parent, Xavier and van der Torre, Leendert. "A Deontic Logic Reasoning Infrastructure." In *14th Conference on Computability in Europe, Cie 2018, Kiel, Germany, July 30-August, 2018, Proceedings*, edited by Florin Manea, Russell G. Miller, and Dirk Nowotka, 10936:60–69. Springer: Lecture Notes in Computer Science, 2018.

78 Benzmüller & Scott, "Automating free logic in HOL". Benzmüller & Scott, "Axiom systems for category theory in free logic".
79 MacLane, "Groups, categories and duality".
80 Scott, "Identity and existence in intuitionistic logic".
81 Freyd & Scedrov, "Categories, Allegories".
82 Benzmüller, Claus & Sultana, "Systematic verification of the modal logic cube".
83 Benzmüller, "Universal (meta-)logical reasoning".
84 Benzmüller, Parent & van der Torre, "A deontic logic reasoning infrastructure".
85 Benzmüller, Farjami & Parent, "A dyadic deontic logic in HOL". Benzmüller, Farjami & Parent, "Åqvist's dyadic deontic logic E in HOL". Farjami, Meder, Parent & Benzmüller, "I/O logic in HOL".
86 Fuenmayor & Benzmüller, "Formalisation and Evaluation of Alan Gewirth's Proof".

Benzmüller, Christoph, and Paulson, Lawrence. "Quantified Multimodal Logics in Simple Type Theory." *Logica Universalis* 7:1 (2013), 7–20.

Benzmüller, Christoph, and Scott, Dana. "Axiom Systems for Category Theory in Free Logic." *Archive of Formal Proofs*. 2018. URL = https://www.isa-afp.org/entries/AxiomaticCategoryTheory.html

Benzmüller, Christoph, and Scott, Dana. "Automating Free Logic in HOL, with an Experimental Application in Category Theory." *Journal of Automated Reasoning.* 2019. 10.1007/s10817-018-09507-7

Benzmüller, Christoph, Sultana, Nik, Paulson, Lawrence C. and Theiß, Frank. "The Higher-Order Prover LEO-II." *Journal of Automated Reasoning* 55:4 (2015), 389–404.

Benzmüller, Christoph, Weber, Leon and Woltzenlogel Paleo, Bruno. "Computer-Assisted Analysis of the Anderson-Hájek Controversy." *Logica Universalis* 11:1 (2017), 139–151.

Benzmüller, Christoph, and Woltzenlogel Paleo, Bruno. "Automating Gödel's Ontological Proof of God's Existence with Higher-Order Automated Theorem Provers." In *ECAI 2014*, edited by Torsten Schaub, Gerhard Friedrich, and Barry O'Sullivan, 263:93–98. IOS Press: Frontiers in Artificial Intelligence and Applications, 2014.

Benzmüller, Christoph, and Woltzenlogel Paleo, Bruno. "Higher-Order Modal Logics: Automation and Applications." In *Reasoning Web 2015*, edited by Adrian Paschke and Wolfgang Faber, 32–74. Springer: Lecture Notes in Computer Science, 2015.

Benzmüller, Christoph, and Woltzenlogel Paleo, Bruno. "Interacting with Modal Logics in the Coq Proof Assistant." In *Computer Science — Theory and Applications — 10th International Computer Science Symposium in Russia, CSR 2015, Listvyanka, Russia, July 13-17, 2015, Proceedings*, edited by Lev D. Beklemishev and Daniil V. Musatov, 9139:398–411. Springer: Lecture Notes in Computer Science, 2015.

Benzmüller, Christoph, and Woltzenlogel Paleo, Bruno. "An Object-Logic Explanation for the Inconsistency in Gödel's Ontological Theory (Extended Abstract, Sister Conferences)." In *KI 2016: Advances in Artificial Intelligence, Proceedings*, edited by Malte Helmert and Franz Wotawa, 244–250. Springer: Lecture Notes in Computer Science, 2016.

Benzmüller, Christoph, and Woltzenlogel Paleo, Bruno. "The Inconsistency in Gödel's Ontological Argument: A Success Story for AI in Metaphysics." In *IJCAI 2016*, edited by Subbarao Kambhampati, 936–942. AAAI Press, 2016.

Benzmüller, Christoph, and Woltzenlogel Paleo, Bruno. "The Modal Collapse as a Collapse of the Modal Square of Opposition." In *The Square of Opposition: A Cornerstone of Thought (Collection of Papers Related to the World Congress on the Square of Opposition Iv, Vatican, 2014)*, edited by Jean-Yves Béziau and Gianfranco Basti, 307–313. Springer: Studies in Universal Logic, 2016.

Bertot, Yyes, and Casteran, Pierre. *Interactive Theorem Proving and Program Development - Coq'Art: The Calculus of Inductive Constructions*. Springer: Texts in Theoretical Computer Science, 2004.

Bjørdal, Frode. "Understanding Gödel's Ontological Argument." In *The Logica Yearbook 1998*, edited by T. Childers, 214–217. Filosofia, 1999.

Blanchette, Jasmin Christian, Böhme, Sascha and Paulson, Lawrence Charles. "Extending Sledgehammer with SMT Solvers." *Journal of Automated Reasoning* 51:1 (2013), 109–128.

Blanchette, Jasmin Christian, and Nipkow, Tobias. "Nitpick: A Counterexample Generator for Higher-Order Logic Based on a Relational Model Finder." In *Interactive Theorem Proving, First International Conference, ITP 2010, Edinburgh, Uk, July 11-14, 2010. Proceedings*, edited by Matt Kaufmann and Lawrence C. Paulson, 6172:131–146. Springer: Lecture Notes in Computer Science, 2010.

Blanchette, Jasmin Christian, Popescu, Andrei, Wand, Daniel, and Weidenbach, Christoph. "More SPASS with Isabelle – Superposition with Hard Sorts and Configurable Simplification." In *Interactive Theorem Proving — Third International Conference, ITP F2012, Princeton, Nj, Usa, August 13-15, 2012. Proceedings*, edited by Lennart Beringer and Amy P. Felty, 7406:345–360. Springer: Lecture Notes in Computer Science, 2012.

Brown, Chad E. "Satallax: An Automatic Higher-Order Prover." In *Automated Reasoning — 6th International Joint Conference, IJCAR 2012, Manchester, Uk, June 26-29, 2012. Proceedings*, edited by Bernhard Gramlich, Dale Miller, and Uli Sattler, 7364:111–117. Springer: Lecture Notes in Computer Science, 2012.

Cruanes, Simon, and Blanchette, Jasmin Christian. "Extending Nunchaku to Dependent Type Theory." In *Proceedings First International Workshop on Hammers for Type Theories, Hatt@IJCAR 2016*, 3–12. EPTCS, 2016.

de Moura, Leonardo, and Bjørner, Nikolaj. "Z3: An Efficient Smt Solver." In *Tools and Algorithms for the Construction and Analysis of Systems*, edited by C. R. Ramakrishnan and Jakob Rehof, 337–340. Springer, 2008.

Deters, Morgan, Reynolds, Andrew, King, Tim, Barrett, Clark W., and Tinelli, Cesare. "A Tour of CVC4: How It Works, and How to Use It." In *Formal Methods in Computer-Aided Design, FMCAD 2014*, edited by Koen Claessen and Viktor Kuncak, 7. IEEE, 2014.

Farjami, Ali, Meder, Paul, Parent, Xavier, and Benzmüller, Christoph. 2018. "I/O Logic in HOL." *Journal of Applied Logics*, 2019. To appear, preprint: URL = https://orbilu.uni.lu/handle/10993/37013

Fitelson, Branden, and Zalta, Edward N. "Steps Toward a Computational Metaphysics." *Journal Philosophical Logic* 36:2 (2007), 227–47.

Fitting, Melvin. *Types, Tableaus, and Gödel's God*. Kluwer, 2002.

Freyd, Peter, and Scedrov, Andre. *Categories, Allegories*. North Holland, 1990.

Fuenmayor, David, and Benzmüller, Christoph. "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, 10505:114–127. Springer: Lecture Notes in Artificial Intelligence, 2017.

Fuenmayor, David, and Benzmüller, Christoph. "Types, Tableaus and Gödel's God in Isabelle/HOL." *Archive of Formal Proofs,* 2017. URL = https://www.isa-afp.org/entries/Types_Tableaus_and_Goedels_God.html

Fuenmayor, David, and Benzmüller, Christoph. "A Case Study on Computational Hermeneutics: E. J. Lowe's Modal Ontological Argument." *Journal of Applied Logics (Special Issue on Formal Approaches to the Ontological Argument)* 5:7 (2018), 1567–1603.

Fuenmayor, David, and Benzmüller, Christoph. "Formalisation and Evaluation of Alan Gewirth's Proof for the Principle of Generic Consistency in Isabelle/HOL." *Archive of Formal Proofs*, 2018. URL = http://isa-afp.org/entries/GewirthPGCProof.html

Gödel, Kurt. "Appendix A. Notes in Kurt Gödel's Hand." In *Logic and Theism: Arguments for and Against Beliefs in God*, edited by J. H. Sobel, 144–145. Cambridge University Press, 1970.

Hájek, Petr. "Magari and Others on Gödel's Ontological Proof." In *Logic and Algebra*, edited by A. Ursini and P. Agliano, 125–135. Dekker, 1996.

Hájek, Petr. "Der Mathematiker und die Frage der Existenz Gottes." In *Kurt Gödel. Wahrheit und Beweisbarkeit*, edited by B. Buldt et al., 325–336. öbv & hpt Verlagsgesellschaft mbH, 2001.

Hájek, Petr. "A New Small Emendation of Gödel's Ontological Proof." *Studia Logica* 71:2 (2002), 149–164.

Huffman, Brian, and Kuncar, Ondrej. "Lifting and Transfer: A Modular Design for Quotients in Isabelle/HOL." In *Certified Programs and Proofs — Third International Conference, CPP 2013, Melbourne, Vic, Australia, December 11-13, 2013, Proceedings*, edited by Georges Gonthier and Michael Norrish, 8307:131–146. Springer: Lecture Notes in Computer Science, 2013.

Kirchner, Daniel. "Representation and Partial Automation of the Principia Logico-Metaphysica in Isabelle/HOL." *Archive of Formal Proofs*, 2017. URL = https://www.isa-afp.org/entries/PLM.html

Kovács, Laura, and Voronkov, Andrei. "First-Order Theorem Proving and Vampire." In *Computer Aided Verification - 25th International Conference, CAV 2013, Saint Petersburg, Russia, July 13-19, 2013. Proceedings*, edited by Natasha Sharygina and Helmut Veith, 8044:1–35. Springer: Lecture Notes in Computer Science, 2013.

Lowe, Edward Jonathan. "A Modal Version of the Ontological Argument." In *Debating Christian Theism*, edited by J. P. Moreland, K. A. Sweis, and C. V. Meister, 61–71. Oxford University Press, 2013.

MacLane, Saunders. "Groups, Categories and Duality." *Proceedings of the National Academy of Sciences* 34:6 (1948), 263–267.

Nipkow, Tobias, Paulson, Lawrence C., and Wenzel, Markus. *Isabelle/HOL — a Proof Assistant for Higher-Order Logic*. Springer: Lecture Notes in Computer Science, 2002.

Oppenheimer, Paul E., and Zalta, Edward N. "On the Logic of the Ontological Argument." *Philosophical Perspectives* 5 (1991), 509–529.

Oppenheimer, Paul E., and Zalta, Edward N. "A Computationally-Discovered Simplification of the Ontological Argument." *Australasian Journal of Philosophy* 89:2 (2011), 333–49.

Oppenheimer, Paul E., and Zalta, Edward N. "Relations Versus Functions at the Foundations of Logic: Type-Theoretic Considerations." *Journal of Logic and Computation* 21:2 (2011), 351–74.

Pelletier, Francis J., and Zalta, Edward N. "How to Say Goodbye to the Third Man." *Noûs* 34:2 (2000), 165–202.

Schulz, Stephan. "System Description: E 1.8." In *Logic for Programming, Artificial Intelligence, and Reasoning – 19th International Conference, Lpar-19, Stellenbosch, South Africa, December 14-19, 2013. Proceedings*, edited by Kenneth L. McMillan, Aart Middeldorp, and Andrei Voronkov, 8312:735–743. Springer: Lecture Notes in Computer Science, 2013.

Scott, Dana S. "Appendix B: Notes in Dana Scott's Hand." In *Logic and Theism: Arguments for and Against Beliefs in God*, edited by J. H. Sobel, 145–146. Cambridge University Press, 1972.

Scott, Dana S. "Identity and Existence in Intuitionistic Logic." In *Applications of Sheaves: Proceedings of the Research Symposium on Applications of Sheaf Theory to Logic, Algebra, and Analysis, Durham, July 9–21, 1977*, edited by Michael Fourman, Christopher Mulvey, and Dana Scott, 752:660–696. Springer: Lecture Notes in Mathematics, 1979.

Sobel, Jordan H. "Gödel's Ontological Proof." In *On Being and Saying. Essays for Richard Cartwright*, edited by Judith Jarvis Tomson, 241–61. MIT Press, 1987.

Sobel, Jordan H. *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge University Press, 2004.

Steen, Alexander, and Benzmüller, Christoph. "The Higher-Order Prover Leo-III." In *Automated Reasoning. IJCAR 2018*, edited by Didier Galmiche, Stephan Schulz, and Roberto Sebastiani, 10900:108–116. Springer: Lecture Notes in Computer Science, 2018.

Zalta, Edward N. *Abstract Objects: An Introduction to Axiomatic Metaphysics*. D. Reidel, 1983.

Zalta, Edward N. "Logical and Analytic Truths That Are Not Necessary." *The Journal of Philosophy* 85:2 (1988), 57–74.

Zalta, Edward N. "Twenty-Five Basic Theorems in Situation and World Theory." *Journal of Philosophical Logic* 22:4 (1993), 385–428.

Zalta, Edward N. "Natural Numbers and Natural Cardinals as Abstract Objects: A Partial Reconstruction of Frege's Grundgesetze in Object Theory." *Journal of Philosophical Logic* 28:6 (1999), 619–660.