A Case Study on Computational Hermeneutics: E. J. Lowe's Modal Ontological Argument

David Fuenmayor[1] and Christoph Benzmüller[2,1]

[1]Freie Universität Berlin, Germany
[2]University of Luxembourg, Luxembourg

November 29, 2017

Abstract

Computers may help us to understand –not just verify– arguments. Through the mechanization of a variant of St. Anselm's ontological argument by E. J. Lowe, which is a paradigmatic example of a natural-language argument with strong ties to metaphysics and religion, we offer an ideal showcase for a computer-assisted interpretive method. This method, which we name *computational hermeneutics*, has been specifically conceived for use in interactive proof assistants and aims at shedding light on the meanings of concepts and beliefs by framing their inferential roles in a given argument. By employing *automated theorem proving* technology, we are able to drastically reduce (by several orders of magnitude) the time needed to test modifications to formalized arguments. As a result, a new approach to argument's interpretation, inspired by Donald Davidson's account of *radical interpretation*, has become possible. The case study presented here relies on the use of the *Isabelle* proof assistant to quickly provide the kind of objective facts needed by a radical interpreter in order to understand an argument, which is presupposed to be valid (in the spirit of the *principle of charity*). The *computational hermeneutics* approach allows us to expose the assumptions we indirectly commit ourselves to every time we opt for some particular logical formalization and fosters the explicitation and revision of our beliefs and commitments until arriving at a state of *reflective equilibrium*: A state where our beliefs have the highest degree of coherence and acceptability.

Part I: Introductory Matter

This article is divided in three parts. In the first one, we present the philosophical motivation and theoretical underpinnings of our approach; we also outline the landscape of automated deduction. In the second part we introduce our method of *computational hermeneutics* as an iterative process of conceptual explication and working our belief-systems into coherence and acceptability. In the last part, we present our case study: the computer-assisted interpretation of E. J. Lowe's modal ontological argument, where our approach becomes exemplified.

Philosophical and Religious Arguments

Is religion a conversation-stopper? Do religious beliefs provide a conceptual framework through which a believer's world-view is structured, such that the interpretation (i.e. understanding) of a religious argument becomes a hopeless case, given the incommensurability between the conceptual schemes of the believer and of the lay interpreter? The answer to these questions boils down to finding a way to acknowledge the variety of religious belief, while recognizing that we all share, at heart, a similar assortment of concepts and are thus able to understand each other. Terry Godlove has convincingly argued in [17] 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" [11].

Davidson's criticism of what he calls "conceptual relativism" relies on the view that talk of incommensurable conceptual schemes is possible only on violating a correct understanding of interpretability, as developed in his theory of radical interpretation –especially vis-à-vis the well-known *principle of charity*. Furthermore, the kind of meaning holism implied by Davidson's account of interpretation suggests that we must share vastly more belief than not with anyone whose words and actions we are able to interpret. Thus, divergence in belief must be limited: If an interpreter is to interpret someone as asserting that Jerusalem is a holy place, she has to presume that the speaker holds true many closely related sentences; for instance, that Jerusalem is a city, that holy places are sites of pilgrimage, and, if the speaker is Christian, that Jesus is the son of God and lived in Jerusalem –and so on. Meaning holism requires us, so Godlove's thesis, to reject the notion that religions are alternative, incommensurable conceptual frameworks.

Drawing upon our experience with the computer-assisted reconstruction and assessment of ontological arguments for the existence of God [4, 5, 16, 3], we can bear witness to the previous claims. More often than not, we have been forced to consider unstated (implicit) assumptions needed for argument's validity[1] and to ponder how much we would therefore depart from the original argument. We also had to consider issues like the plausibility of our assumptions from the standpoint of the author and its compatibility with the author's purported beliefs (or what she said elsewhere).[2] Our experience also lends support to the holistic principle that beliefs have content only by virtue of their inferential relations to other beliefs. Hence we aim at showing why and how an interpretive approach drawing on content holism is especially suited for finding meaning in religious and metaphysical discourse.

We want to address the issue of understanding a certain type of arguments and the role computers can play in it, by proposing an approach named *computational hermeneutics*. We are thus urged to distinguish the kind of arguments we want to address from others that, on the one hand, rely on appeals to faith and rhetorical effects, or, on the other hand, make use of already well-defined concepts with univocal usage, like in mathematics. We have already talked of religious arguments in the spirit of St. Anselm's ontological argument as some of the arguments we are interested in; we want, nonetheless, to generalize the domain of applicability of our approach to what we call 'philosophical' arguments –for lack of a better word– since we consider many of the concepts introduced into philosophy as inexact ("explicanda" in Carnaps terminology). We want to defend the view that the process of *explicating* those philosophical concepts takes place in the very practice of argumentation through the *explicitation* of the inferential role they play in some theory or argument of our interest. In the context of a formalized argument (in some chosen logic), this task of concept explication can be carried out either by giving definitions (directly correlating to "explicata") or by axiomatizing conceptual interrelations. Both approaches will be illustrated in the case study presented in the last section.

[1] In what follows, we call an argument "valid" if and only if it is impossible for its premises to be true and its conclusion to be false –the conclusion of a valid argument will thus sometimes be referred as "valid". An argument is called "sound" if and only if it is a valid argument whose premises are actually true. Since, on the one hand, questions of soundness boil down in most cases to some kind of empirical inquiry and, on the other hand, we feel that the discussion of the many different conceptions of truth for sentences is out of the scope of this article, we want to focus on argument's validity instead of soundness, thus leaving aside the question about the truth of each individual premise or conclusion.

[2] Cf. [14]. Eder and Ramharter propose here several criteria aimed at judging the adequacy of formal reconstructions of St. Anselm's ontological argument. They also show how such reconstructions help us gain a better understanding of this argument.

Top-down and Bottom-up Approaches to Meaning

We want to talk here of the *meaning* of an expression –or argument– as that 'something', which the interpreter needs to grasp in order to *understand* it. Talk of meanings in philosophy has always been a risky business, especially when one wants to avoid the kind the ontological commitments resulting from postulating the existence, for every linguistic expression, of some obscure abstract being in need of definite identity criteria ("no entity without identity"). We will use, therefore, the word meaning in its widest possible sense, so we can talk about such blurry things as the inferential role of expressions.

We also want to acknowledge the compositional character of natural and formalized languages, so we can think of the meaning of an argument as a function of the meanings of each of its constituent sentences (premises and conclusions) and their mode of combination (logical consequence relation).[3] Accordingly, we take the meaning of each sentence as resulting from the meaning of its constituent words (concepts) and their mode of combination. We can therefore, by virtue of compositionality, conceive a *bottom-up* approach for the interpretation of an argument by starting with our pre-understanding of its main concepts and its logical structure; and then working our way up to a better understanding of its sentences and their inferential interrelations.

The bottom-up approach outlined above is the one employed in the formal verification of arguments. However, it leaves open the question of how to arrive at the meaning of words beyond our initial pre-understanding of them; and this question is central to our project, since we are interested in understanding more than mere verification. Thus, we want to complement the compositional (bottom-up) approach with a holistic (top-down) one, by proposing a computer-supported method aimed at determining the meaning of concepts from their inferential role vis-à-vis argument's validity, much in the spirit of Donald Davidson's program of *radical interpretation*.[4]

[3] It is, arguably, part of the rules of the formal argumentation game that the validity (and meaning) of an argument must depend solely on what is explicitly stated (premises, inference rules, etc.). Ideally, an argument would be analyzed as an island isolated from any external linguistic or pre-linguistic goings-on; and, for instance, when *implicit* premises are brought to our attention, they should be made *explicit* and integrated into the argument accordingly –which must always remain an *intersubjectively* accessible artifact: a product of our socio-linguistic discursive practices. In the same spirit, it is also reasonable to expect of all sentences to derive their meaning compositionally –we see no place for idioms in philosophical arguments.

[4] The connections between Davidson's truth-centered theory of meaning and theories focusing on the inferential role of expressions (e.g. [9, 20, 8]) have been much discussed in the literature. While some authors (Davidson included) see both holistic approaches as essentially different, others (e.g. [36], [21], p. 72) have come to see Davidson's theory as an instance of inferential-role semantics. We side with the latter.

Radical Interpretation and the Principle of Charity

What is the use of radical interpretation in religion? The answer is trivially stated by Davidson himself, who convincingly argues that "all understanding of the speech of another involves radical interpretation" ([10], p. 125). Furthermore, the impoverished evidential position we are faced with when interpreting religious arguments corresponds very closely to the starting situation Davidson contemplates in his thought experiments on *radical interpretation*, where he shows how an interpreter could come to understand someone's words and actions without relying on any prior understanding of them.[5] Davidson builds on the idea of taking truth as basic and extracting from it an account of translation or interpretation [12]. His project is faced from the start with two general requirements: (i) the theory of interpretation must help reveal the compositional structure of language, and (ii) the theory can be supported or verified by evidence available to the interpreter.

The first requirement (i) is addressed by noting that a theory of truth in Tarski's style (modified to apply to natural language) can be used as a theory of interpretation. This implies that, for every sentence s of an object language L, a sentence of the form: ' "s" is true in L iff p' (aka. *T-schema*) can be derived, where p is a translation of s into the metalanguage used for interpretation –note that the sentence p is being *used*, while s is only being *mentioned*. Thus, by virtue of the recursive nature of Tarski's theory, the structure of the object language becomes revealed. In *computational hermeneutics*, the object language L corresponds to the idiolect of the speaker (natural language), and the metalanguage is constituted by the formulas of our logic of formalization (e.g. modal or higher-order logic, as in our case study) plus the expression "is valid in the logic XYZ".

An instance of the *T-schema* as used in our approach may thus be: ' "There is only one God" is true iff "$\exists\, x.\ God\ x \wedge \forall\, y.\ God\ y \rightarrow y{=}x$" is valid in HOL'. Note that, in our variant, the *used* metalanguage sentence p has now the form: ' "q" is valid in HOL', where the *mentioned* sentence q corresponds to the formalization of the object sentence s. Taking q as an interpretation of s certainly helps us to clarify our understanding of s and to shed light –by virtue of compositionality– on the meanings of its individual words. We might be tempted to take q already as 'the meaning' of s, but this is not what is intended. We will illustrate in our case study how, in order to leverage our interpretive process, meanings can be better understood as the (holistic) inferential role of expressions.[6]

[5] For an interesting discussion of the relevance of Davidson's language philosophy in religious studies, we refer the reader to [18].

[6] As is well known, there is a tension between the holistic nature of inferential roles and a compositional account of meaning. In *computational hermeneutics*, by showing both approaches in action (top-down and bottom-up), we demonstrate their compatibility in

The second general requirement (ii) states that the interpreter can have access to objective evidence in order to judge the appropriateness of her interpretations (i.e. access to the events and objects in the 'external world' which cause sentences to be true, or in our case valid). This requirement implies, particularly, access to the speaker's attitudes regarding the truth or falsity of sample sentences, under specified circumstances observable by speaker and interpreter alike. In *computational hermeneutics*, this kind of objective (i.e. intersubjective) evidence is provided by the output of automated tools. Thus, the computer acts as an arguably unbiased arbiter deciding on the truth (i.e. validity) of sentences in the context of an argument.

A central concept in Davidson's account of *radical interpretation* is the *principle of charity*, which he holds as a condition for the possibility of engaging in any kind of interpretive endeavor. The *principle of charity* has been summarized by Davidson as follows: "We make maximum sense of the words and thoughts of others when we interpret in a way that optimizes agreement" [11]. Hence the principle builds on the possibility of intersubjective agreement about external facts among speaker and interpreter. The *principle of charity* can thus be invoked to make sense of a speaker's ambiguous utterances and, in our case, to presume (and foster) the validity of an argument. Davidson has argued in many places (e.g. [10]) that many basic sentences must be true at those times they are held true by a speaker; he calls this "a form of 'charity' in the sense that it assumes meanings are more or less the same when relevant verbal behaviors are the same" [13]. In *computational hermeneutics*, we take the speaker's –and interpreter's– act of holding that a conclusion follows from some set of premises (i.e. that the conclusion is *valid*) as one of the "relevant verbal behaviors" Davidson is referring to. We thus argue, following the *principle of charity*, that a condition for the possibility of interpreting a philosophical argument is to assume that its conclusions indeed follow from its premises.

The Automated Reasoning Landscape

Automated Reasoning is an umbrella term used for a wide range of technologies sharing the overall goal of mechanizing different forms of reasoning –understood as the ability to draw inferences. Born as a subfield of artificial intelligence with the aim of automatically generating mathematical proofs,[7] automated reasoning has moved to close proximity of logic and philosophy, thanks to theoretical developments in the last decades. Nevertheless, its

practice. For a theoretical treatment of the relationship between holism and compositionality, we refer the reader to [29, 27, 28].

[7] For instance, the first widely recognized AI system: *Logic Theorist*, was able to prove 38 of the first 52 theorems of Whitehead and Russell's "Principia Mathematica" back in 1956.

main field of application has mostly remained bounded to mathematics and hardware and software verification. In this respect, the field of *automated theorem proving* (ATP) has traditionally been its most developed subarea. ATP involves the design of algorithms that automate the process of construction (proof generation) and verification (proof checking) of mathematical proofs. Some extensive work has also been done in other non-deductive forms of reasoning (i.e. by analogy, induction and abduction). However, those fields remain largely underrepresented in comparison.

There have been major advances regarding the automatic generation of formal proofs during the last years, which we think make the utilization of formal methods in philosophy very promising and have even brought about some novel philosophical results (e.g. [5]). We will, on this occasion, restrain ourselves to the computer-supported interpretation of extant arguments and will thus be mostly concerned with the subfield of automated proof checking.

Proof checking can be carried out either non-interactively (for instance as a batch operation) or interactively by utilizing a proof assistant. A non-interactive proof-checking program would normally get as input some formula (string of characters in some predefined syntax) and a context (some set of formulas) and will, in positive cases, generate a listing of the formulas (in the given context) from which the input formula logically follows, together with the name of the proof method[8] used and, in some cases, a proof string (as in the case of proof generators). Some proof checking programs, called *model finders*, are specialized in searching for models and, more importantly, countermodels for a given formula. This functionality proves very useful in practice by sparing us the thankless task of proving non-theorems.

Human guidance is oftentimes required by theorem provers in order to effectively solve interesting problems. A need has been recognized for the synergistic combination of the vast memory resources and information-processing capabilities of modern computers, together with human ingenuity, by allowing people to give hints to these tools by the means of especially crafted user interfaces. The field of *interactive theorem proving* has grown out of this endeavor and its software programs are known as *proof assistants*.[9]

Automated reasoning is currently being applied to solve problems in formal logic, mathematics and computer science, software and hardware verification and many others. For instance, the Mizar Library[10] and TPTP (Thousands

[8] Some proof methods commonly employed by the Isabelle proof assistant are: term rewriting, classical reasoning, tableaus, model elimination, ordered resolution and paramodulation.

[9] A survey and system comparison of the most famous interactive proof assistants has been carried out in [35]. The results of this survey remain largely accurate to this day.

[10] Cf. [24]. Mizar proofs and their corresponding articles are published regularly in the peer-reviewed *Journal of Formalized Mathematics*.

of Problems for Theorem Provers) [34] are two of the biggest libraries of such problems being maintained and updated on a regular basis. There is also a yearly competition among automated theorem provers held at the CADE conference [30], whose problems are selected from the TPTP library.

Automated theorem provers have been used to assist in the formalization of many advanced mathematical proofs such as Erdös-Selberg's proof of the *Prime Number Theorem* (about 30,000 lines in Isabelle), the proof of the *Four Color Theorem* (60,000 lines in Coq), and the proof of the *Jordan Curve Theorem* (75,000 lines in HOL-Light) [32]. The monumental proof of Kepler's conjecture by Thomas Hales and his research team has been recently formalized and verified using the Isabelle and HOL-Light proof assistants as part of the *Flyspeck project* [19].

Isabelle [25] is the proof assistant we will use to illustrate our *computational hermeneutics* method. Isabelle offers a structured proof language called *Isar* specifically tailored for writing proofs that are both computer- and human-readable and which focuses on higher-order classical logic. The different variants of the ontological argument assessed in our case study are formalized directly in Isabelle's HOL dialect or, for the modal variants, through the technique of shallow semantic embeddings [2].

Part II: The Computational Hermeneutics Method

Why Should We Use Computers in Metaphysics?

It is easy to argue that using computers for the assessment of arguments brings us many *quantitative* advantages, since it gives us the means to construct and verify proofs easier, faster and much more reliably. Furthermore, the main task of this paper is to illustrate a central *qualitative* advantage of computer-supported argumentation: It enables a different, *holistic* approach to philosophical argumentation which forms the backbone of our *computational hermeneutics* method.

To get an idea of this, let us imagine the following scenario: A philosopher working on a formal argument wants to test a variation on one of its premises or definitions and find out if the argument still holds. Since our philosopher is working with pen and paper, she will have to follow some kind of proof procedure (e.g. natural-deduction calculus or tableaus), which, depending on her calculation skills, may take some minutes to be carried out. It seems clear that the working philosopher or logician cannot allow herself many of these experiments on such conditions.

Now compare the above scenario to using an interactive *proof assistant*, where our working philosopher can carry out such an experiment in just a few seconds and with almost no effort. In a best-case scenario, the proof assistant would automatically generate a proof (or the sketch of a countermodel), so she just needs to interpret the results and use them to inform her new conjectures. In any case, she would at least know if her speculations had the intended consequences, or not. After some minutes of work, she will have tried plenty of different variations of the argument while getting real-time feedback regarding their suitability.[11] We aim at showing how this radical *quantitative* increase in productivity does indeed entail a *qualitative* change in the way we approach formal argumentation, since it allows us to do something we were not able to do before by using pen and paper: We are set free to engage in a lot more experimentation during the assessment of formal philosophical arguments, and this freedom allows us to take things to a whole new level, as will become apparent in the discussion of our case study. (Note that we are talking of many hundreds of such trial-and-error experiments!)

The Procedure

The method presented here, which we call *computational hermeneutics*, is aimed at exploiting the computing power and usability of modern theorem provers, by drawing on both the bottom-up and the top-down approaches to meaning: We work iteratively on an argument by temporarily fixing truth-values and inferential relations among its sentences, and then, after choosing a logic for formalization, working back and forth on the formalization of its axioms and theorems by making gradual adjustments while getting automatic feedback about the suitability of our speculations. In this fashion, by engaging in a dialectic process of questions and answers –of conjectures and refutations– we work our way towards a proper understanding of an argument by circular movements between its parts and the whole (cf. hermeneutical circle). This way, we progressively get insight into the meaning of individual concepts and get in a position to assess and revise our beliefs and commitments until arriving at a state of *reflective equilibrium*: A state where our beliefs have the highest degree of coherence and acceptability.[12]

[11] The situation is obviously idealized, since, as is well known, most of theorem-proving problems are computationally complex (NP-hard) and even undecidable, so in many cases a solution will take several minutes or just never be found. Nevertheless, as work in the emerging field of *computational metaphysics* [26, 1, 33, 4, 5] suggests, the lucky situation depicted above is not rare.

[12] We have been inspired by John Rawls' notion of *reflective equilibrium* as applied in ethics and political philosophy. Loosely speaking, we can see radical interpretation, reflective equilibrium and computational hermeneutics as instances of the *hypothetico-deductive* method (aka. 'scientific method'), since the sort of mutual adjustment between theory and data involved is a familiar feature of the idealized scientific practice. For an interesting discussion of the application of the method of reflective equilibrium to logical analysis by drawing on an inferentialist account of meaning, we refer the reader to [31].

The iterative structure of the *computational hermeneutics* method can be depicted as follows.

Repeat until reflective equilibrium:

1. **Reconstruct argument in natural language**

1.1. **Add and remove sentences** such as premises and (wanted or unwanted) conclusions.

1.2. **Fix truth-values for sentences**, since we may want to validate some sentences (e.g. conclusions) while avoiding some others (e.g. contradictions, modal collapse, etc.), according to what we think the speaker's attitudes regarding such sentences are.

1.3. **Optional: establish inferential relations**, i.e. the extension of the logical consequence relation: which sentences follow logically from which others. We can, alternatively, let our automated tools find out this for themselves (after formalization in stage 3). This frequently leads to the simplification of the argument by dropping unnecessary premises.

2. **Establish a logic for formalization**, by determining the logical structure of our natural-language argument.

3. **Formalize sentences in the chosen logic**, while getting continuous feedback about the argument's validity. This stage is itself iterative, since, for every sentence, we (charitably) try several different formalizations until getting a valid argument. Here is where we take most advantage of the real-time feedback offered by our automated tools. Two main tasks are:

3.1. **Translate natural-language sentences into the target logic**, by relying either on our pre-understanding or on provided definitions of the argument's concepts.

3.2. **Bring related terms together**, either by introducing definitions or by axiomatizing new interrelations among them. These newly introduced expressions are then translated back into natural language to be integrated into the argument in step 1.1, thus being disclosed as former *implicit* premises.

4. **Reflect upon the results**, and eventually come back to stage 1, 2 or 3.

Part III: Lowe's Modal Ontological Argument

In this section we illustrate the computer-supported interpretation of a variant of St. Anselm's ontological argument for the existence of God, using *Isabelle/HOL*.[13] This argument was introduced by the philosopher E. J. Lowe in an article named "A Modal Version of the Ontological Argument" [23] and offers a paradigmatic example of a natural-language argument with strong ties to metaphysics and religion. The interpretation of this argument thus makes for an ideal showcase for *computational hermeneutics* in practice.

Lowe offers in his article a new modal variant of the ontological argument, which is specifically aimed at proving the *necessary* existence of God. In a nutshell, Lowe's argument works by first postulating the existence of "necessary abstract" beings, that is, abstract beings that exist in every possible world (e.g. numbers). He then introduces the concepts of *ontological dependence* and *metaphysical explanation* and argues that the existence of every (mind-dependent) abstract being is ultimately explained by some concrete being (e.g. a mind). By interrelating the concepts of *dependence* and *explanation*, he argues that the concrete being(s), on which each "necessary abstract" being depends for its existence, must also be necessary. This way he proves the existence of at least one "necessary concrete" being (God according to his definition).

Lowe further argues that his argument qualifies as a modal ontological argument, since it focuses on *necessary* existence, and not just existence of some kind of supreme being. His argument differs from other familiar variants of the modal ontological argument (like Gödel's) in that it does not appeal, in the first place, to the possible existence of God (whose essence contains its necessary existence) in order to use the modal *S5* axioms to deduce its necessary existence as a conclusion.[14] Lowe wants therefore to circumvent the usual criticisms to the *S5* axiom system, like implying the unintuitive assertion that *whatever is possibly necessarily the case is thereby actually the case.*

The structure of Lowe's argument is very representative of philosophical arguments. It features eight premises from which new inferences are drawn until arriving at a final conclusion: the necessary existence of God (which in this case amounts to the existence of some "necessary concrete being"). The argument's premises are reproduced verbatim below:

[13] We refer the reader to [15] for further details. That computer-verified article has been completely written in the Isabelle proof assistant and thus requires some familiarity with this system.

[14] Actually, a modal logic *KB* suffices to prove Gödel's argument (with Scott's emendation) as shown in [4].

(P1) God is, by definition, a necessary concrete being.

(P2) Some necessary abstract beings exist.

(P3) All abstract beings are dependent beings.

(P4) All dependent beings depend for their existence on independent beings.

(P5) No contingent being can explain the existence of a necessary being.

(P6) The existence of any dependent being needs to be explained.

(P7) Dependent beings of any kind cannot explain their own existence.

(P8) The existence of dependent beings can only be explained by beings on which they depend for their existence.

We will consider here only a representative subset of the argument's conclusions, which are reproduced below (exactly as presented in Lowe's article):

(C1) All abstract beings depend for their existence on concrete beings. (Follows from P3 and P4 together with definitions D3 and D4.)

(C5) In every possible world there exist concrete beings. (Follows from C1 and P2.)

(C7) The existence of necessary abstract beings needs to be explained. (Follows from P2, P3 and P6.)

(C8) The existence of necessary abstract beings can only be explained by concrete beings. (Follows from C1, P3, P7 and P8.)

(C9) The existence of necessary abstract beings is explained by one or more necessary concrete beings. (Follows from C7, C8 and P5.)

(C10) A necessary concrete being exists. (Follows from C9.)

Lowe also introduces some informal definitions which should help the reader understand the meaning of the concepts involved in his argument (necessity, concreteness, ontological dependence, metaphysical explanation, etc.). In the following discussion, we will see that most of these definitions do not bear the significance Lowe claims.

(D1) x is a necessary being := x exists in every possible world.

(D2) x is a contingent being := x exists in some but not every possible world.

(D3) x is a concrete being := x exists in space and time, or at least in time.

(D4) x is an abstract being := x does not exist in space or time.

(D5) x depends for its existence on y := necessarily, x exists only if y exists.

In the following sections we apply the *computational hermeneutics* method to the argument shown above. We compile in each section the results of a series of iterations and present them as a new variant of the original argument. We want to illustrate how the argument (as well as our understanding of it) gradually evolves as we experiment with different combinations of definitions, premises and logics for formalization.

First Iteration Series: Initial Formalization

Let us first turn to the formalization of premise P1: "God is, by definition, a necessary concrete being".[15]

In order to understand the concept of *necessariness* (i.e. being a "necessary being") employed in this argument, we have a look at the definitions D1 and D2 provided by the author. They relate the concepts of necessariness and contingency (i.e. being a "contingent being") with existence:[16]

(D1) *x is a necessary being := x exists in every possible world.*

(D2) *x is a contingent being := x exists in some but not every possible world.*

The two definitions above, aimed at explicating the concepts of necessariness and contingency by reducing them to existence and quantification over possible worlds, have a direct impact on the choice of a logic for formalization. They not only call for some kind of modal logic with possible-world semantics but also lead us to consider the complex issue of existence, since we need to restrict the domain of quantification at every world.

Regarding the choice of a modal logic for formalization, it was necessary to circumvent a technical constraint: The Isabelle proof assistant, as well as many others, does not natively support modal logics but only classical higher-order logic. As a workaround, we have used a technique known as *shallow semantic embedding*, which allows us to take advantage of the expressive power of higher-order logic in order to embed the semantics of an object language. We draw on previous work on the embedding of multimodal logics in HOL [2], which has successfully been applied to the analysis

[15] When the author says of something that it is a "necessary concrete being" we will take him to say that it is both necessary and concrete. Certainly, when we say of Tom that he is a lousy actor, we just don't mean that he is lousy and that he also acts. For the time being, we won't differentiate between predicative and attributive uses of adjectives, so we will formalize both sorts as unary predicates; since the particular linguistic issues concerning attributive adjectives don't seem to play a role in this argument. Following the *computational hermeneutics* approach (in the spirit of the *principle of charity*) we may justify adding further complexity to the argument's formalization if we later find out that it is required for its validity.

[16] Here, the concepts of necessariness and contingency are meant as properties of beings, in contrast to the concepts of necessity and possibility which are modals. We will see later how both pairs of concepts can be related in order to validate this argument.

and verification of ontological arguments (e.g. [5, 4, 3, 16]). Using this technique, we can embed a modal logic K by defining the box and diamond operators using restricted quantification over the set of 'accessible' worlds (using an *accessibility* relation R as a guard). Note that, in the following definitions, the type *wo* is declared as an abbreviation for $w{\Rightarrow}bool$, which corresponds to the type of a function mapping worlds (of type w) to boolean values. *wo* thus corresponds to the type of a world-dependent formula (i.e. its *truth set*).

consts $R{::}w{\Rightarrow}w{\Rightarrow}bool$ (**infix R**) — Accessibility relation
abbreviation *mbox* :: $wo{\Rightarrow}wo$ ($\Box$-)
  where $\Box\varphi \equiv \lambda w.\forall v.\ (w\ \mathbf{R}\ v){\longrightarrow}(\varphi\ v)$
abbreviation *mdia* :: $wo{\Rightarrow}wo$ ($\Diamond$-)
  where $\Diamond\varphi \equiv \lambda w.\exists v.\ (w\ \mathbf{R}\ v){\wedge}(\varphi\ v)$

Validity is defined as truth in *all* worlds and represented by wrapping the formula in special brackets ($\lfloor - \rfloor$).

abbreviation *valid*::$wo{\Rightarrow}bool$ ($\lfloor$-$\rfloor$) **where** $\lfloor\psi\rfloor \equiv\ \forall w.(\psi\ w)$

We verify our embedding by using Isabelle's simplifier to prove the K principle and the *necessitation* rule.

lemma *K*: $\lfloor(\Box(\varphi \to \psi)) \to (\Box\varphi \to \Box\psi)\rfloor$ **by** *simp* — Verifying K principle
lemma *NEC*: $\lfloor\varphi\rfloor \Longrightarrow \lfloor\Box\varphi\rfloor$ **by** *simp* — Verifying *necessitation* rule

Regarding existence, we need to commit ourselves to a certain position in metaphysics known as *metaphysical contingentism*, which roughly states that the existence of any entity is a contingent fact: *some* entities can exist at *some* worlds, while not existing at some others. The negation of metaphysical contingentism is known as *metaphysical necessitism*, which basically says that *all* entities must exist at *all* possible worlds. By not assuming contingentism, and therefore assuming necessitism, the whole argument would become trivial, since all beings would end up being trivially necessary (i.e. existing in all worlds).[17]

Hence we can guard our quantifiers so they range only over those entities existing (i.e. being actualized) at a given world. This approach is known as *actualist quantification* and is implemented, using the semantic embedding technique, by defining a world-dependent meta-logical 'existence' predicate (called "actualizedAt" below), which is the one used as a guard in the definition of the quantifiers. Note that the type e characterizes the domain of all beings (i.e. existing and non-existing entities), and the type *wo* characterizes sets of worlds. The term "isActualized" thus relates beings to worlds.

[17] Metaphysical contingentism looks *prima facie* like a very natural assumption to make; nevertheless an interesting philosophical debate between advocates of necessitism and contingentism has arisen during the last years, especially in the wake of Timothy Williamson's work [37].

consts *isActualized*::$e \Rightarrow wo$ (**infix** *actualizedAt*)

abbreviation *forallAct*::$(e \Rightarrow wo) \Rightarrow wo$ ($\forall^A$)
where $\forall^A \Phi \equiv \lambda w. \forall x.\ (x\ actualizedAt\ w) \longrightarrow (\Phi\ x\ w)$
abbreviation *existsAct*::$(e \Rightarrow wo) \Rightarrow wo$ ($\exists^A$)
where $\exists^A \Phi \equiv \lambda w. \exists x.\ (x\ actualizedAt\ w) \wedge (\Phi\ x\ w)$

The corresponding binder syntax is defined below.

abbreviation *mforallActB*::$(e \Rightarrow wo) \Rightarrow wo$ (**binder**$\forall^A$[*8*]*9*)
where $\forall^A x.\ (\varphi\ x) \equiv \forall^A \varphi$
abbreviation *mexistsActB*::$(e \Rightarrow wo) \Rightarrow wo$ (**binder**$\exists^A$[*8*]*9*)
where $\exists^A x.\ (\varphi\ x) \equiv \exists^A \varphi$

We use Isabelle's Nitpick tool [7] to verify that actualist quantification validates neither the Barcan formula nor its converse.[18]

lemma $\lfloor (\forall^A x.\ \Box(\varphi\ x)) \rightarrow \Box(\forall^A x.\ \varphi\ x) \rfloor$
nitpick oops — Countermodel found: formula not valid
lemma $\lfloor \Box(\forall^A x.\ \varphi\ x) \rightarrow (\forall^A x.\ \Box(\varphi\ x)) \rfloor$
nitpick oops — Countermodel found: formula not valid

Non-guarded quantifiers, in contrast, validate both the Barcan formula and its converse.

lemma $\lfloor (\forall x.\Box(\varphi\ x)) \rightarrow \Box(\forall x.(\varphi\ x)) \rfloor$
by *simp* — Proven by Isabelle's simplifier
lemma $\lfloor \Box(\forall x.(\varphi\ x)) \rightarrow (\forall x.\Box(\varphi\ x)) \rfloor$
by *simp* — Proven by Isabelle's simplifier

With actualist quantification in place we can: (i) formalize the concept of existence in the usual form (by using a restricted particular quantifier), (ii) formalize necessariness as existing necessarily, and (iii) formalize contingency as existing possibly but not necessarily.

definition *Existence*::$e \Rightarrow wo$ ($E!$) **where** $E!\ x \equiv \exists^A y.\ y \approx x$

definition *Necessary*::$e \Rightarrow wo$ **where** $Necessary\ x \equiv \Box E!\ x$
definition *Contingent*::$e \Rightarrow wo$ **where** $Contingent\ x \equiv \Diamond E!\ x \wedge \neg Necessary\ x$

Note that we have just chosen a logic for formalization: a free quantified modal logic *K* with positive semantics. The logic is *free* because the domain of quantification (for actualist quantifiers) is a proper subset of our universe of discourse, so we can refer to non-actual objects. The semantics is *positive* because we have placed no restriction regarding predication on non-actual objects, so they are also allowed to exemplify properties and relations. We are also in a position to embed stronger normal modal logics (*KB, KB5, S4,*

[18] We utilize here model finder Nitpick for the first time. For the conjectured lemma, Nitpick has found a countermodel, i.e. a model satisfying all stated axioms which falsifies the given formula. The formula is consequently non-valid (as indicated by the "oops" keyword).

S5, etc.) by restricting the accessibility relation R with additional axioms, if needed.

Having chosen our logic, we can now turn to the formalization of the concepts of abstractness and concreteness. As seen previously, Lowe has already provided us with an explication of these concepts:

(D3) *x is a concrete being := x exists in space and time, or at least in time.*

(D4) *x is an abstract being := x does not exist in space or time.*

Lowe himself acknowledges that the explication of these concepts in terms of existence "in space and time" is superfluous, since we are only interested in them being complementary.[19] Thus, we start by formalizing concreteness as a *primitive* world-dependent predicate and then derive abstractness from it, namely as its negation.

consts *Concrete*::$e \Rightarrow wo$
abbreviation *Abstract*::$e \Rightarrow wo$ **where** $Abstract\ x \equiv\ \neg(Concrete\ x)$

We can now formalize the definition of Godlikeness (P1) as follows:

abbreviation *Godlike*::$e \Rightarrow wo$ **where** $Godlike\ x \equiv Necessary\ x \wedge Concrete\ x$

We also formalize premise P2 ("Some necessary abstract beings exist") as shown below:

axiomatization where
P2: $\lfloor \exists^{A} x.\ Necessary\ x \wedge Abstract\ x \rfloor$

Let us now turn to premises P3 ("All abstract beings are dependent beings") and P4 ("All dependent beings depend for their existence on independent beings"). We have here three new concepts to be explicated: two predicates "dependent" and "independent" and a relation "depends (for its existence) on", which has been called *ontological dependence* by Lowe. Following our linguistic intuitions concerning their interrelation, we start by proposing the following formalization:

consts *dependence*::$e \Rightarrow e \Rightarrow wo$ (**infix** *dependsOn*)
definition *Dependent*::$e \Rightarrow wo$ **where** $Dependent\ x \equiv \exists^{A} y.\ x\ dependsOn\ y$
abbreviation *Independent*::$e \Rightarrow wo$ **where** $Independent\ x\ \equiv \neg(Dependent\ x)$

[19] We quote from Lowe's original article: "Observe that, according to these definitions, a being cannot be both concrete and abstract: being concrete and being abstract are mutually exclusive properties of beings. Also, all beings are either concrete or abstract ... the abstract/concrete distinction is exhaustive. Consequently, a being is concrete if and only if it is not abstract."

We have formalized ontological dependence as a *primitive* world-dependent relation and refrained from any explication (as suggested by Lowe).[20]

We have called an entity *dependent* if and only if there *actually exists* an object y such that x *depends for its existence* on it; accordingly, we have called an entity *independent* if and only if it is not dependent.

As a consequence, premises P3 ("All abstract beings are dependent beings") and P4 ("All dependent beings depend for their existence on independent beings") become formalized as follows.

axiomatization where
P3: $\lfloor\forall^A x.\ Abstract\ x \rightarrow Dependent\ x\rfloor$ **and**
P4: $\lfloor\forall^A x.\ Dependent\ x \rightarrow (\exists^A y.\ Independent\ y \wedge x\ dependsOn\ y)\rfloor$

Concerning premises P5 ("No contingent being can explain the existence of a necessary being") and P6 ("The existence of any dependent being needs to be explained"), a suitable formalization for expressions of the form: "the entity X explains the existence of Y" and "the existence of X is explained" needs to be found.[21] These expressions rely on a single binary relation, which will initially be taken as *primitive*. This relation has been called *metaphysical explanation* by Lowe.[22]

consts *explanation*::$e \Rightarrow e \Rightarrow wo$ (**infix** *explains*)
definition *Explained*::$e \Rightarrow wo$ **where** $Explained\ x \equiv \exists^A y.\ y\ explains\ x$

axiomatization where
P5: $\lfloor\neg(\exists^A x.\ \exists^A y.\ Contingent\ y \wedge Necessary\ x \wedge y\ explains\ x)\rfloor$

Premise P6, together with the last two premises: P7 ("Dependent beings of any kind cannot explain their own existence") and P8 ("The existence of dependent beings can only be explained by beings on which they depend for their existence"), were introduced by Lowe in order to relate the concept of

[20] An explication of this concept has been suggested by Lowe in definition D5 ("x depends for its existence on y := necessarily, x exists only if y exists"). Concerning this alleged definition, he has written in a footnote to the same article: "Note, however, that the two definitions (D5) and (D6) presented below are not in fact formally called upon in the version of the ontological argument that I am now developing, so that in the remainder of this chapter the notion of existential dependence may, for all intents and purposes, be taken as primitive. There is an advantage in this, inasmuch as finding a perfectly apt definition of existential dependence is no easy task, as I explain in 'Ontological Dependence.'" Lowe refers hereby to his article on ontological dependence in the *Stanford Encyclopedia of Philosophy* [22] for further discussion.

[21] Note that we have omitted the expressions "can" and "needs to", since they seem to play here only a rhetorical role. As in the case of attributive adjectives discussed above, we first aim at the simplest workable formalization; however, we are willing to later improve on this formalization in order to foster argument's validity, in accordance to the *principle of charity*.

[22] This concept is closely related to what has been called *metaphysical grounding* in contemporary literature.

metaphysical explanation to *ontological dependence*.[23]

axiomatization where
P6: $\lfloor \forall x.\ Dependent\ x \rightarrow Explained\ x \rfloor$ **and**
P7: $\lfloor \forall x.\ Dependent\ x \rightarrow \neg(x\ explains\ x) \rfloor$ **and**
P8: $\lfloor \forall x\ y.\ y\ explains\ x \rightarrow x\ dependsOn\ y \rfloor$

Although the last three premises seem to couple very tightly the concepts of (metaphysical) explanation and (ontological) dependence, both concepts are not meant by the author to be equivalent.[24] We have used Nitpick to test this claim. Since a countermodel has been found, we have proven that the (inverse) equivalence of metaphysical explanation and ontological dependence is not implied by the axioms.

lemma $\lfloor \forall x\ y.\ x\ explains\ y \leftrightarrow y\ dependsOn\ x \rfloor$ **nitpick**[*user-axioms*] **oops**

For any being, however, having its existence "explained" is equivalent to its existence being "dependent" (on some other being). This follows already from premises P6 and P8, as shown above by Isabelle's prover.

lemma $\lfloor \forall x.\ Explained\ x \leftrightarrow Dependent\ x \rfloor$
using *P6 P8 Dependent-def Explained-def* **by** *auto*

The Nitpick model finder is also useful to check axioms consistency at any stage during the formalization of an argument. We instruct Nitpick to generate a model satisfying some tautological sentence (here we use a trivial 'True' proposition) while taking into account all previously defined axioms. Nitpicks output is a text-based description of the model found, as shown.

lemma *True* **nitpick**[*satisfy*, *user-axioms*] **oops**

In this case, Nitpick was able to find a model satisfying the given tautology; this means that all axioms defined so far are consistent. The model found has a cardinality of two for the set of individual objects and a single world.

We can also use model finders to perform 'sanity checks': We instruct Nitpick to find a countermodel for some specifically tailored formula which we want to make sure is not valid, because of its implausibility from the point of view of the author (as we interpret him). We check below, for instance, that our axioms are not too strong as to imply *metaphysical necessitism* (i.e. all beings necessarily exist) or *modal collapse*. Since both would trivially validate the argument.

[23] Note that we use non-restricted quantifiers for the formalization of the last three premises in order to test the argument's validity under the strongest assumptions. As before, we turn a blind eye to the modal expression "can".

[24] Lowe says: "Existence-explanation is not simply the inverse of existential dependence. If x depends for its existence on y, this only means that x cannot exist without y existing. This is not at all the same as saying that x exists because y exists, or that x exists in virtue of the fact that y exists."

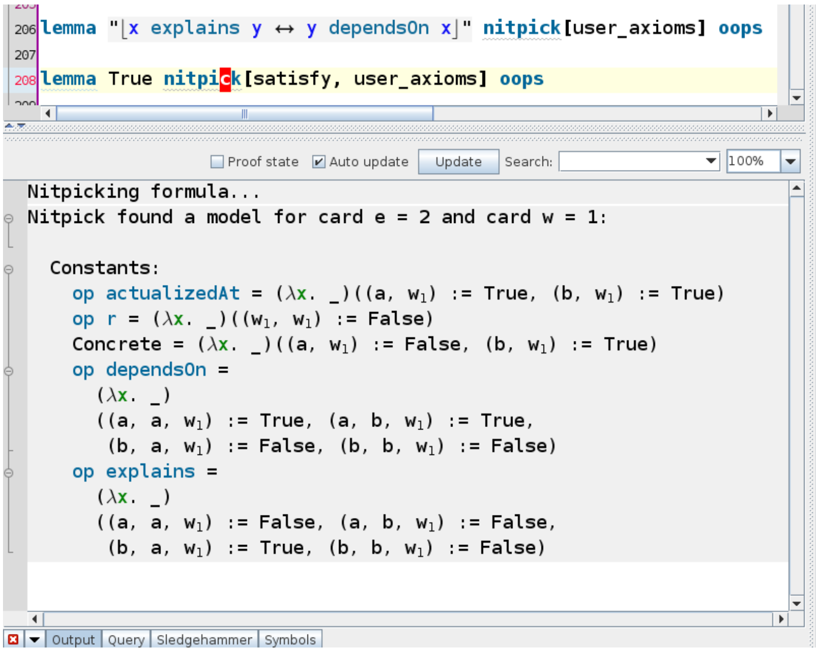

Figure 1: A screenshot of the Isabelle proof assistant showing a textual representation of the countermodel found by Nitpick for the purported inverse equivalence between ontological dependence and metaphysical explanation.

lemma $\lfloor \forall x.\ E!\ x \rfloor$
  nitpick[*user-axioms*] **oops** — Countermodel found: necessitism is not valid
lemma $\lfloor \varphi \rightarrow \Box\varphi \rfloor$
  nitpick[*user-axioms*] **oops** — Countermodel found: modal collapse is not valid

Model finders like Nitpick are only able to verify consistency (by finding a model) or non-validity (by finding a countermodel) for a given formula. When it comes to verifying validity or invalidity, we are reliant on automated theorem provers. Isabelle is equipped with various different provers (with and without proof generation) tailored for specific kinds of problems and thus employing different approaches, strategies and heuristics. We have used Isabelle's *Sledgehammer* tool [6] which successively applies several provers (feeding them with different combinations of axioms and theorems) until finding the most optimized proof for a given formula, all within seconds. By using Sledgehammer we can verify the validity of our partial conclusions (C1, C5 and C7) and even find the premises they rely upon.[25]

[25] We prove theorems in Isabelle by using the keyword "by" followed by the name of a proof method (i.e. some computer-implemented algorithm). Some methods commonly used in Isabelle are: *simp* (term rewriting), *blast* (tableaus), *meson* (model elimination), *metis* (ordered resolution and paramodulation) and *auto* (classical reasoning and term rewriting).

(C1) *All abstract beings depend for their existence on concrete beings.*

theorem *C1*: $\lfloor \forall^A x.\ Abstract\ x \rightarrow (\exists\ y.\ Concrete\ y \wedge x\ dependsOn\ y)\rfloor$
using *P3 P4* **by** *blast*

(C5) *In every possible world there exist concrete beings.*

theorem *C5*: $\lfloor \exists^A x.\ Concrete\ x\rfloor$
using *P2 P3 P4* **by** *blast*

(C7) *The existence of necessary abstract beings needs to be explained.*

theorem *C7*: $\lfloor \forall^A x.\ (Necessary\ x \wedge Abstract\ x) \rightarrow Explained\ x\rfloor$
using *P3 P6* **by** *simp*

The last three conclusions are shown by Nitpick to be non-valid even in the stronger *S5* logic. *S5* can be easily introduced by postulating that the accessibility relation R is an equivalence relation. This exploits the *Sahlqvist correspondence* which relates modal axioms to constraints on a model's accessibility relation: reflexivity, symmetry, seriality, transitivity and euclideanness imply axioms T, B, D, IV, V respectively (and also the other way round).

axiomatization where
S5: *equivalence R* — We assume *T*: $\Box\varphi\rightarrow\varphi$, *B*: $\varphi\rightarrow\Box\Diamond\varphi$ and *4*: $\Box\varphi\rightarrow\Box\Box\varphi$

(C8) *The existence of necessary abstract beings can only be explained by concrete beings.*

lemma *C8*: $\lfloor \forall^A x.(Necessary\ x \wedge Abstract\ x)\rightarrow(\forall^A y.\ y\ explains\ x\rightarrow Concrete\ y)\rfloor$
nitpick[*user-axioms*] **oops**

(C9) *The existence of necessary abstract beings is explained by one or more necessary concrete (Godlike) beings.*

lemma *C9*: $\lfloor \forall^A x.(Necessary\ x \wedge Abstract\ x)\rightarrow(\exists^A y.\ y\ explains\ x \wedge Godlike\ y)\rfloor$
nitpick[*user-axioms*] **oops**

(C10) *A necessary concrete (Godlike) being exists.*

theorem *C10*: $\lfloor \exists^A x.\ Godlike\ x\rfloor$ **nitpick**[*user-axioms*] **oops**

Nitpick does not only spare us the effort of searching for non-existent proofs but also provides us with very helpful information when it comes to fix an argument by giving us a text-based description of the countermodel found (as we can appreciate below for C10).

By employing the Isabelle proof assistant we prove non-valid a first formalization attempt of Lowe's modal ontological argument. This is, however, just the first of many series of iterations in our interpretive endeavor. Based on the information recollected so far, we can proceed to make the adjustments necessary to validate the argument. We will see how these adjustments have an impact on the inferential role of all concepts (necessariness, concreteness, dependence, explanation, etc.) and therefore on their meaning.

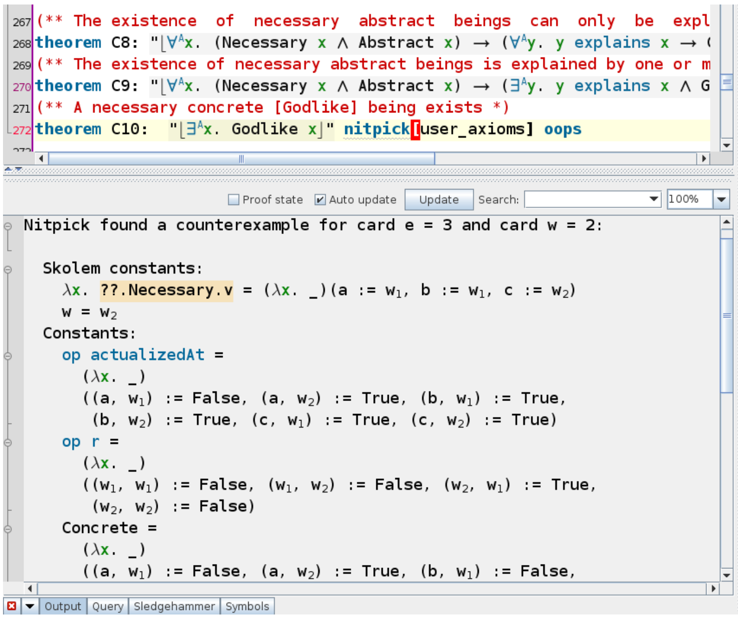

Figure 2: A screenshot of the Isabelle proof assistant showing a textual representation of the countermodel found by Nitpick for theorem C10.

Second Iteration Series: Validating the Argument I

By carefully examining the above countermodel for C10, it has been noticed that some necessary beings, which are abstract in the actual world, may indeed be concrete in other accessible worlds. Lowe has previously presented numbers as an example of such necessary abstract beings. It can be argued that numbers, while existing necessarily, can never be concrete in any possible world, so we add the restriction of abstractness being an essential property, i.e. a locally rigid predicate.

axiomatization where
  abstractness-essential: $\lfloor \forall x.\ Abstract\ x \rightarrow \Box Abstract\ x \rfloor$

theorem *C10*: $\lfloor \exists^A x.\ Godlike\ x \rfloor$
  nitpick[*user-axioms*] **oops** — Countermodel found

As Nitpick shows us, the former restriction is not enough to prove C10. We try postulating further restrictions on the accessibility relation R, which, taken together, would amount to it being an equivalence relation. This would make for a modal logic *S5* (cf. *Sahlqvist correspondence*), and thus the abstractness property becomes a (globally) rigid predicate.

axiomatization where
 T-axiom: *reflexive* R **and** — $\Box\varphi \rightarrow \varphi$
 B-axiom: *symmetric* R **and** — $\varphi \rightarrow \Box\Diamond\varphi$
 IV-axiom: *transitive* R — $\Box\varphi \rightarrow \Box\Box\varphi$

theorem *C10*: $\lfloor \exists^A x.\ Godlike\ x \rfloor$
nitpick[*user-axioms*] **oops** — Countermodel found

By examining the new countermodel found by Nitpick, we noticed that at some worlds there are non-existent concrete beings. We want to disallow this possibility, so we make concreteness an existence-entailing property.

axiomatization where *concrete-exist*: $\lfloor \forall x.\ Concrete\ x \to E!\ x \rfloor$

We carry out the usual 'sanity checks' to make sure the argument has not become trivialized.[26]

lemma *True*
nitpick[*satisfy*, *user-axioms*] **oops** — Model found: axioms are consistent
lemma $\lfloor \forall x.\ E!\ x \rfloor$
nitpick[*user-axioms*] **oops** — Countermodel found: necessitism is not valid
lemma $\lfloor \varphi \to \Box\varphi \rfloor$
nitpick[*user-axioms*] **oops** — Countermodel found: modal collapse is not valid

Since Nitpick could not find a countermodel for C10, we have enough confidence in its validity to ask Sledgehammer to search for a proof.

theorem *C10*: $\lfloor \exists^A x.\ Godlike\ x \rfloor$ **using** *Existence-def Necessary-def abstractness-essential concrete-exist P2 C1 B-axiom* **by** *meson*

Sledgehammer is able to find a proof relying on all premises but the two modal axioms T and IV. By the end of this series of iterations, we see that Lowe's modal ontological argument depends for its validity on three unstated (i.e. implicit) premises: the essentiality of abstractness, the existence-entailing nature of concreteness, and the modal axiom B ($\varphi \to \Box\Diamond\varphi$). Moreover, we shed some light on the meaning of the concepts of abstractness and concreteness, as we disclose further premises which shape their inferential role in the argument.

Third Iteration Series: Validating the Argument II

We present a slightly simplified version of the original argument (without the implicit premises stated in the previous version). In this variant premises P1 to P5 remain unchanged and none of the last three premises proposed by Lowe (P6 to P8) show up anymore. Those last premises have been introduced in order to interrelate the concepts of explanation and dependence in such a way that they play somewhat opposite roles. Now we want to go all the way and simply assume that explanation and dependence are inverse relations, for we want to understand how the interrelation of these two concepts affects the validity of the argument.

[26] These checks are constantly carried out after postulating axioms for every iteration, so we won't mention them anymore.

axiomatization where
dep-expl-inverse: $\lfloor \forall x\ y.\ y\ explains\ x \leftrightarrow x\ dependsOn\ y \rfloor$

Let us first prove the relevant partial conclusions.

theorem *C1*: $\lfloor \forall^A x.\ Abstract\ x \rightarrow (\exists\ y.\ Concrete\ y \wedge x\ dependsOn\ y) \rfloor$
using *P3 P4* **by** *blast*

theorem *C5*: $\lfloor \exists^A x.\ Concrete\ x \rfloor$
using *P2 P3 P4* **by** *blast*

theorem *C7*: $\lfloor \forall^A x.\ (Necessary\ x \wedge Abstract\ x) \rightarrow Explained\ x \rfloor$
using *Explained-def P3 P4 dep-expl-inverse* **by** *meson*

However, the conclusion C10 is still countersatisfiable, as shown by Nitpick:

theorem *C10*: $\lfloor \exists^A x.\ Godlike\ x \rfloor$
nitpick[*user-axioms*] **oops** — Countermodel found

Next, let us try assuming a stronger modal logic. We can do this by postulating further modal axioms using the *Sahlqvist correspondence* and asking Sledgehammer to find a proof. Sledgehammer is in fact able to find a proof for C10 which only relies on the modal axiom T ($\Box\varphi \rightarrow \varphi$).

axiomatization where
T-axiom: *reflexive R* **and** — $\Box\varphi \rightarrow \varphi$
B-axiom: *symmetric R* **and** — $\varphi \rightarrow \Box\Diamond\varphi$
IV-axiom: *transitive R* — $\Box\varphi \rightarrow \Box\Box\varphi$

theorem *C10*: $\lfloor \exists^A x.\ Godlike\ x \rfloor$ **using** *Contingent-def Existence-def P2 P3 P4 P5 dep-expl-inverse T-axiom* **by** *meson*

In this series of iterations we have verified a modified version of the original argument by Lowe. Our understanding of the concepts of *ontological dependence* and *metaphysical explanation* have changed after the introduction of an additional axiom constraining both: they are now inverse relations. This new understanding of the inferential role of the concepts of *ontological dependence* and *metaphysical explanation* has been reached on the condition that the ontological argument, as stated in natural language, must hold (in accordance to the *principle of charity*). Depending on our stance on this matter, we may either feel satisfied with this result or want to consider further alternatives. In the former case we would have reached a state of *reflective equilibrium*. In the latter we would rather carry on with our iterative process in order to further illuminate the meaning of the concepts involved in this argument.

Fourth Iteration Series: Simplifying the Argument

After some further iterations we arrive at a new variant of Lowe's argument: Premises P1 to P4 remain unchanged and a new premise D5 ("x depends for its existence on y := necessarily, x exists only if y exists") is added. D5 corresponds to the 'definition' of ontological dependence as put forth by Lowe in his article (though only for illustrative purposes). As mentioned before, this purported definition was never meant by him to become part of the argument. Nevertheless, we show here how, by assuming the left-to-right direction of this definition, we get in a position to prove the main conclusions without any further assumptions.

axiomatization where *D5*: $\lfloor\forall^A x\ y.\ x\ dependsOn\ y \rightarrow \Box(E!\ x \rightarrow E!\ y)\rfloor$

theorem *C1*: $\lfloor\forall^A x.\ Abstract\ x \rightarrow (\exists\ y.\ Concrete\ y \wedge x\ dependsOn\ y)\rfloor$
using *P3 P4* **by** *meson*

theorem *C5*: $\lfloor\exists^A x.\ Concrete\ x\rfloor$ **using** *P2 P3 P4* **by** *meson*

theorem *C10*: $\lfloor\exists^A x.\ Godlike\ x\rfloor$
using *Necessary-def P2 P3 P4 D5* **by** *meson*

In this variant we have been able to verify the conclusion of the argument without appealing to the concept of metaphysical explanation. We were able to get by with just the concept of ontological dependence by explicating it in terms of existence and necessity (as suggested by Lowe).

As a side note, we can also prove that the original premise P5 ("No contingent being can explain the existence of a necessary being") directly follows from D5 by redefining metaphysical explanation as the inverse relation of ontological dependence.

abbreviation *explanation*::($e \Rightarrow e \Rightarrow wo$) (**infix** *explains*)
where $y\ explains\ x \equiv x\ dependsOn\ y$

lemma *P5*: $\lfloor\neg(\exists^A x.\ \exists^A y.\ Contingent\ y \wedge Necessary\ x \wedge y\ explains\ x)\rfloor$
using *Necessary-def Contingent-def D5* **by** *meson*

In this series of iterations we have reworked Lowe's argument so as to get rid of the somewhat obscure concept of metaphysical explanation, thus simplifying the argument. We also got some insight into Lowe's concept of ontological dependence vis-à-vis its inferential role in the argument (by axiomatizing its relation with the concepts of existence and necessity in D5).

There are still some interesting issues to consider. Note that the definitions of existence (*Existence-def*) and being "dependent" (*Dependent-def*) are not needed in any of the highly optimized proofs found by our automated tools. This raises some suspicions concerning the role played by the existence predicate in the definitions of necessariness and contingency, as well as putting into question the need for a definition of being "dependent" linked to the ontological dependence relation. We will see in the following section that our suspicions are justified and that this argument can be dramatically simplified.

Fifth Iteration Series: Arriving at a Non-Modal Argument

A new simplified emendation of Lowe's argument is obtained after abandoning the concept of existence and redefining necessariness and contingency accordingly. As we will see, this variant is actually non-modal and can be easily formalized in first-order predicate logic.

A more literal reading of Lowe's article has suggested a simplified formalization, in which necessariness and contingency are taken as complementary predicates. According to this, our domain of discourse becomes divided in four main categories, as exemplified in the table below.[27]

	Abstract	Concrete
Necessary	Numbers	God
Contingent	Fiction	Stuff

consts *Necessary*::$e \Rightarrow wo$
abbreviation *Contingent*::$e \Rightarrow wo$ **where** *Contingent x* $\equiv \neg$(*Necessary x*)

consts *Concrete*::$e \Rightarrow wo$
abbreviation *Abstract*::$e \Rightarrow wo$ **where** *Abstract x* $\equiv$ $\neg$(*Concrete x*)

abbreviation *Godlike*::$e \Rightarrow w \Rightarrow bool$ **where** *Godlike x* $\equiv$ *Necessary x* $\wedge$ *Concrete x*

consts *dependence*::$e \Rightarrow e \Rightarrow wo$ (**infix** *dependsOn*)
abbreviation *explanation*::($e \Rightarrow e \Rightarrow wo$) (**infix** *explains*)
where *y explains x* $\equiv$ *x dependsOn y*

As shown below, we can even define the "dependent" predicate as *primitive* (i.e. bearing no relation to ontological dependence) and still be able to validate the argument. Being "independent" is defined as the negation of being "dependent", as before.

consts *Dependent*::$e \Rightarrow wo$
abbreviation *Independent*::$e \Rightarrow wo$ **where** *Independent x* $\equiv \neg$(*Dependent x*)

By taking, once again, metaphysical explanation as the inverse relation of ontological dependence and by assuming premises P2 to P5 we can prove conclusion C10.

[27] As Lowe explains in the article, "there is no logical restriction on combinations of the properties involved in the concrete/abstract and the necessary/contingent distinctions. In principle, then, we can have contingent concrete beings, contingent abstract beings, necessary concrete beings, and necessary abstract beings."

axiomatization where
P2: $\lfloor \exists\, x.\ Necessary\ x \wedge Abstract\ x \rfloor$ **and**
P3: $\lfloor \forall\, x.\ Abstract\ x \rightarrow Dependent\ x \rfloor$ **and**
P4: $\lfloor \forall\, x.\ Dependent\ x \rightarrow (\exists\, y.\ Independent\ y \wedge x\ dependsOn\ y) \rfloor$ **and**
P5: $\lfloor \neg(\exists\, x.\ \exists\, y.\ Contingent\ y \wedge Necessary\ x \wedge y\ explains\ x) \rfloor$

theorem *C10*: $\lfloor \exists\, x.\ Godlike\ x \rfloor$ **using** *P2 P3 P4 P5* **by** *blast*

Note that, in the axioms above, all actualist quantifiers have been changed into non-guarded quantifiers, following the elimination of the concept of existence from our argument: Our quantifiers range over *all* beings, because all beings exist. Also note that all modal operators have disappeared; thus, this new variant is directly formalizable in classical first-order logic.

Sixth Iteration Series: Modified Modal Argument I

In the following two series of iterations, we want to illustrate the use of the *computational metaphysics* approach in those cases where we must start our interpretive endeavor with no *explicit* pre-understanding of the concepts involved (i.e. when no definitions are available). In such cases, we start by taking all concepts as primitive without stating any definition explicitly. We will see how we gradually improve our understanding of these concepts in the iterative process of adding and removing axioms, thus framing their inferential role in the argument.

consts *Concrete*::$e \Rightarrow wo$
consts *Abstract*::$e \Rightarrow wo$
consts *Necessary*::$e \Rightarrow wo$
consts *Contingent*::$e \Rightarrow wo$
consts *dependence*::$e \Rightarrow e \Rightarrow wo$ (**infix** *dependsOn*)
consts *explanation*::$e \Rightarrow e \Rightarrow wo$ (**infix** *explains*)
consts *Dependent*::$e \Rightarrow wo$
abbreviation *Independent*::$e \Rightarrow wo$ **where** *Independent* $x \equiv \neg(Dependent\ x)$

In order to honor the original intention of the author, i.e. providing a *modal* variant of St. Anselm's ontological argument, we are required to make a change in Lowe's original formulation. In this variant we will restate the expressions "necessary abstract" and "necessary concrete" as "necessarily abstract" and "necessarily concrete" correspondingly. With this new adverbial reading of the former "necessary" predicate we are no longer talking about the concept of *necessariness*, but of *necessity* instead, so we use the modal box operator ($\Box$) for its formalization. Note that in this variant we are not concerned with the interpretation of the original natural-language argument anymore. We are interested, instead, in showing how the *computational hermeneutics* method can go beyond simple interpretation and foster a creative approach to assessing and improving philosophical arguments.

Premise P1 now reads: "God is, by definition, a necessari*ly* concrete being."

abbreviation *Godlike*::$e \Rightarrow wo$ **where** *Godlike* $x \equiv \Box$*Concrete* x

Premise P2 reads: "Some necessari*ly* abstract beings exist". The rest of the premises remains unchanged.

axiomatization where
P2: $\lfloor \exists\, x.\ \Box \mathit{Abstract}\ x \rfloor$ **and**
P3: $\lfloor \forall\, x.\ \mathit{Abstract}\ x \rightarrow \mathit{Dependent}\ x \rfloor$ **and**
P4: $\lfloor \forall\, x.\ \mathit{Dependent}\ x \rightarrow (\exists\, y.\ \mathit{Independent}\ y \wedge x\ \mathit{dependsOn}\ y) \rfloor$ **and**
P5: $\lfloor \neg(\exists\, x.\ \exists\, y.\ \mathit{Contingent}\ y \wedge \mathit{Necessary}\ x \wedge y\ \mathit{explains}\ x) \rfloor$

Without postulating any additional axioms, C10 ("A *necessarily* concrete being exists") can be falsified by Nitpick.

theorem *C10*: $\lfloor \exists\, x.\ \mathit{Godlike}\ x \rfloor$
nitpick oops — Countermodel found

An explication of the concepts of necessariness, contingency and explanation is provided below by axiomatizing their interrelation to other concepts. We will now regard necessariness as being *necessarily abstract* or *necessarily concrete*, and explanation as the inverse relation of dependence, as before.

axiomatization where
Necessary-expl: $\lfloor \forall\, x.\ \mathit{Necessary}\ x \leftrightarrow (\Box \mathit{Abstract}\ x \vee \Box \mathit{Concrete}\ x) \rfloor$ **and**
Contingent-expl: $\lfloor \forall\, x.\ \mathit{Contingent}\ x \leftrightarrow \neg \mathit{Necessary}\ x \rfloor$ **and**
Explanation-expl: $\lfloor \forall\, x\ y.\ y\ \mathit{explains}\ x \leftrightarrow x\ \mathit{dependsOn}\ y \rfloor$

Without any further constraints, C10 becomes again falsified by Nitpick.

theorem *C10*: $\lfloor \exists\, x.\ \mathit{Godlike}\ x \rfloor$
nitpick oops — Countermodel found

We postulate further modal axioms (using the *Sahlqvist correspondence*) and ask Isabelle's Sledgehammer tool for a proof. Sledgehammer is able to find a proof for C10 which only relies on the modal axiom T ($\Box\varphi \rightarrow \varphi$).

axiomatization where
T-axiom: *reflexive R* **and** — $\Box\varphi \rightarrow \varphi$
B-axiom: *symmetric R* **and** — $\varphi \rightarrow \Box\Diamond\varphi$
IV-axiom: *transitive R* — $\Box\varphi \rightarrow \Box\Box\varphi$

theorem *C10*: $\lfloor \exists\, x.\ \mathit{Godlike}\ x \rfloor$ **using** *Contingent-expl Explanation-expl Necessary-expl P2 P3 P4 P5 T-axiom* **by** *metis*

Seventh Iteration Series: Modified Modal Argument II

As in the previous variant, we will illustrate here how the meaning (as inferential role) of the concepts involved in the argument gradually becomes apparent in the process of axiomatizing further constraints. We follow on with the adverbial reading of the expression "necessary" but provide an improved explication of the concepts of necessariness and contingency. We think that this explication, in comparison to the previous one, better fits our intuitive understanding of necessariness. We will now regard necessariness as being *necessarily* abstract or concrete (and metaphysical explanation as the inverse of the ontological dependence relation, as before).

axiomatization where
Necessary-expl: $\lfloor \forall x.\ Necessary\ x \leftrightarrow \Box(Abstract\ x \vee\ Concrete\ x)\rfloor$ **and**
Contingent-expl: $\lfloor \forall x.\ Contingent\ x \leftrightarrow \neg Necessary\ x\rfloor$ **and**
Explanation-expl: $\lfloor \forall x\ y.\ y\ explains\ x \leftrightarrow\ x\ dependsOn\ y\rfloor$

These constraints are, however, not enough to ensure the argument's validity, as confirmed by Nitpick.

theorem *C10*: $\lfloor \exists x.\ Godlike\ x\rfloor$ **nitpick oops** — Countermodel found

After some iterations, we see that, by giving a more satisfactory explication of the concept of necesariness, we are also required to (i) assume the essentiality of abstractness (as we did in a former iteration), and (ii) restrict the accessibility relation by enforcing its symmetry (i.e. assuming the modal axiom *B*).

axiomatization where
abstractness-essential: $\lfloor \forall x.\ Abstract\ x \rightarrow \Box Abstract\ x\rfloor$ **and**
B-Axiom: *symmetric R* — $\varphi \rightarrow\ \Box\Diamond\varphi$

theorem *C10*: $\lfloor \exists x.\ Godlike\ x\rfloor$ **using** *Contingent-expl Explanation-expl Necessary-expl P2 P3 P4 P5 abstractness-essential B-Axiom* **by** *metis*

In each of the previous versions we have seen how our understanding of the concepts of necessity/contingency, explanation/dependence and abstractness/concreteness has gradually evolved thanks to the kind of hypothetico-deductive method which has been made possible by the real-time feedback provided by Isabelle's automated proving tools.

We think that, after this last series of iterations, the use of the *computational hermeneutics* method has been illustrated adequately. We have thus reached a state of *reflective equilibrium* and are free to terminate here our interpretive endeavor. It is important to note that this last version of Lowe's argument is by no means its 'best' or 'correct' formalization; it is just a consequence of the path we have followed by coming up with new ideas and testing them with the help of automated tools. In *computational hermeneutics* the journey is much more important than the destination.

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

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

[4] 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.

[5] C. Benzmüller and B. Woltzenlogel Paleo. The inconsistency in Gödel's ontological argument: A success story for AI in metaphysics. In *IJCAI 2016*, 2016.

[6] J. Blanchette, S. Böhme, and L. Paulson. Extending Sledgehammer with SMT solvers. *Journal of Automated Reasoning*, 51(1):109–128, 2013.

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

[8] N. Block. Semantics, conceptual role. In *Routledge Encyclopedia of Philosophy*. Taylor and Francis, 1998.

[9] R. B. Brandom. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1994.

[10] D. Davidson. Radical interpretation interpreted. *Philosophical Perspectives*, 8:121–128, January 1994.

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

[12] D. Davidson. Radical interpretation. In *Inquiries into Truth and Interpretation*. Oxford University Press, September 2001.

[13] D. Davidson. Reply to Kirk Ludwig. In U. Zeglen, editor, *Donald Davidson : truth, meaning and knowledge*. Routledge, reprint. edition, 2001.

[14] G. Eder and E. Ramharter. Formal reconstructions of St. Anselm's ontological argument. *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science*, 192(9), October 2015.

[15] D. Fuenmayor and C. Benzmüller. Computer-assisted reconstruction and assessment of E. J. Lowe's modal ontological argument. *Archive of Formal Proofs*, Sept. 2017. http://isa-afp.org/entries/Lowe_Ontological_Argument.html, Formal proof development.

[16] D. Fuenmayor and C. Benzmüller. Types, Tableaus and Gödel's God in Isabelle/HOL. *Archive of Formal Proofs*, May 2017. http://isa-afp.org/entries/Types_Tableaus_and_Goedels_God.html, Formal proof development.

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

[18] T. F. Godlove, Jr. Saving belief: on the new materialism in religious studies. In N. Frankenberry, editor, *Radical interpretation in religion*. Cambridge University Press, 2002.

[19] T. Hales, M. Adams, G. Bauer, T. D. Dang, J. Harrison, L. T. Hoang, C. Kaliszyk, V. Magron, S. Mclaughlin, T. Nguyen, and et al. A formal proof of the kepler conjecture. *Forum of Mathematics, Pi*, 5, 2017.

[20] G. Harman. (Nonsolipsistic) conceptual role semantics. In E. Lepore, editor, *Notre Dame Journal of Formal Logic*, pages 242–256. Academic Press, 1987.

[21] P. Horwich. *Meaning*. Oxford University Press, 1998.

[22] E. J. Lowe. Ontological dependence. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2010 edition, 2010.

[23] E. J. Lowe. A modal version of the ontological argument. In J. P. Moreland, K. A. Sweis, and C. V. Meister, editors, *Debating Christian Theism*, chapter 4, pages 61–71. Oxford University Press, 2013.

[24] A. Naumowicz and A. Korniłowicz. A brief overview of mizar. In S. Berghofer, T. Nipkow, C. Urban, and M. Wenzel, editors, *Theorem Proving in Higher Order Logics: 22nd International Conference, TPHOLs 2009, Munich, Germany, August 17-20, 2009. Proceedings*, pages 67–72. Springer, Berlin, Heidelberg, 2009.

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

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

[27] P. Pagin. Is compositionality compatible with holism? *Mind & Language*, 12(1):11–33, March 1997.

[28] P. Pagin. Meaning holism. In E. Lepore, editor, *The Oxford handbook of philosophy of language*. Oxford University Press, 1. publ. in paperback edition, 2008.

[29] F. J. Pelletier. Holism and compositionality. In W. Hinzen, E. Machery, and M. Werning, editors, *The Oxford Handbook of Compositionality*. Oxford University Press, 1 edition, February 2012.

[30] F. J. Pelletier, G. Sutcliffe, and C. Suttner. The development of CASC. *AI Commun.*, 15(2,3):79–90, Aug. 2002.

[31] J. Peregrin and V. Svoboda. *Reflective Equilibrium and the Principles of Logical Analysis: Understanding the Laws of Logic*. Routledge Studies in Contemporary Philosophy. Taylor and Francis, 2017.

[32] F. Portoraro. Automated reasoning. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2014 edition, 2014.

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

[34] G. Sutcliffe and C. Suttner. The TPTP problem library. *Journal of Automated Reasoning*, 21(2):177–203, Oct 1998.

[35] F. Wiedijk. *The Seventeen Provers of the World: Foreword by Dana S. Scott (Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[36] M. Williams. Meaning and deflationary truth. *Journal of philosophy*, XCVI(11):545–564, November 1999.

[37] T. Williamson. *Modal Logic as Metaphysics*. Oxford University Press, 2013.