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

David Fuenmayor
Freie Universität Berlin, Germany
david.fuenmayor@fu-berlin.de

Christoph Benzmüller[*]
Freie Universität Berlin, Germany
University of Luxembourg, Luxembourg
c.benzmueller@fu-berlin.de

Abstract

Computers may help us to better understand (not just verify) arguments. In this article we defend this claim by showcasing the application of a new, computer-assisted interpretive method to an exemplary natural-language argument with strong ties to metaphysics and religion: E. J. Lowe's modern variant of St. Anselm's ontological argument for the existence of God. Our new method, which we call *computational hermeneutics*, has been particularly conceived for use in interactive-automated proof assistants. It aims at shedding light on the meanings of words and sentences by framing their inferential role in a given argument. By employing automated theorem reasoning technology within interactive proof assistants, we are able to drastically reduce (by several orders of magnitude) the time needed to test the logical validity of an argument's formalization. As a result, a new approach to logical analysis, inspired by Donald Davidson's account of radical interpretation, has been enabled. In computational hermeneutics, the utilization of automated reasoning tools effectively boosts our capacity to expose the assumptions we indirectly commit ourselves to every time we engage in rational argumentation and it fosters the explicitation and revision of our concepts and commitments.

[*] Funded by Volkswagen Foundation.

Part I: Introductory Matter

The traditional conception of logic as an *ars iudicandi* sees as its central role the classification of arguments into valid and invalid ones by identifying criteria that enable us to judge the correctness of (mostly deductive) inferences. However, logic can also be conceived as an *ars explicandi*, aiming at rendering the inferential rules implicit in our socio-linguistic argumentative praxis in a more orderly, more transparent, and less ambiguous way, thus setting the stage for an eventual critical assessment of our conceptual apparatus and inferential practices.

The novel approach we showcase in this article, called *computational hermeneutics*, is inspired by Donald Davidson's account of *radical interpretation* [18, 15]. It draws on the well-known *principle of charity* and on a holistic account of meaning, according to which the meaning of a term can only be given through the explicitation of the inferential role it plays in some theory (or argument) of our interest. We adopt the view that the process of logical analysis (aka. formalization) of a natural-language argument is itself a kind of interpretation, since it serves the purpose of *making explicit* the inferential relations between concepts and statements.[1] Moreover, the output of this process: *a* logical form, does not need to be unique, since it is dependent on a given background logical theory, or, as Davidson has put it:

> "... much of the interest in logical form comes from an interest in logical geography: to give the logical form of a sentence is to give its logical location in the totality of sentences, to describe it in a way that explicitly determines what sentences it entails and what sentences it is entailed by. The location must be given relative to a specific deductive theory; so logical form itself is relative to a theory." ([16] p. 140)

Following the *principle of charity* while engaging in a process of logical analysis, requires us to search for plausible implicit premises, which would render the given argument as being logically valid and also foster important theoretical virtues such as consistency, non-circularity (avoiding 'question-begging'), simplicity, fruitfulness, etc. This task can be seen as a kind of conceptual *explication*.[2] In computational

[1] In recent times, this idea has become known as *logical expressivism* and has been championed, most notably, by the adherents of semantic inferentialism in the philosophy of language. Two paradigmatic book-length expositions of this philosophical position can be found in the works of Brandom [13] and Peregrin [37].

[2] Explication, in Carnap's sense, is a method of conceptual clarification, aimed at replacing an unclear 'fuzzy' pre-theoretical concept: an *explicandum*, by a new more exact concept with clearly defined rules of use: an *explicatum*, for use in a target theory. While Carnapian in spirit, our idea of explication focuses mostly on the activity of conceptual *explicitation* by the means of formal logic.

hermeneutics we carry it out by providing definitions (i.e. by directly relating a *definiendum* with a *definiens*) or by introducing formulas (e.g. as axioms) which relate a concept we are currently interested in (*explicandum*) with some other concepts which are themselves explicated in the same way (in the context of the same or some other background theory). The circularity involved in this process is an unavoidable characteristic of any interpretive endeavor and has been historically known as the *hermeneutic circle*. Thus, computational hermeneutics contemplates an iterative process of 'trial and error' where the adequacy of some newly introduced formula or definition becomes tested by computing, among others, the logical validity of the *whole* formalized argument. In order to explore the generally very wide space of possible formalizations (and also of interpretations) for even the simplest argument, we have to test its validity at least several hundreds of times (also to account for logical pluralism). It is here where the recent improvements and ongoing consolidation of modern automated theorem proving technology, in particular for higher-order logic (HOL), become handy. A concrete example of the application of this approach using the *Isabelle/HOL* [31] proof assistant to the logical analysis and interpretation of an ontological argument will be provided in the last section.

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

Philosophical and Religious Arguments

Is it possible to find meaning in religious argumentation? Or is religion a conversation-stopper? Do religious beliefs provide a conceptual framework through which a believer's world-view is structured to such an extent, that the interpretation of religious arguments becomes a hopeless case? (given the apparent incommensurability between the conceptual schemes of speakers and interpreters of different creeds). 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. We argue for the role of logic as a common ground for understanding in general and, in particular, for theological argumentation. We reject therefore the view that deep religious convictions constitute an insurmountable obstacle for successful interreligious communication (e.g.

between believers and lay interpreters). Such views have been much discussed in religious studies. Terry Godlove, for instance, has convincingly argued in [24] against what he calls the “framework theory" in religious studies, according to which, for believers, religious beliefs shape the interpretation of most of the objects and situations in their lives. Here Godlove relies on Donald Davidson’s rejection of “the very idea of a conceptual scheme" [17].

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 [8, 9, 23, 7], we can bear witness to the previous claims. While looking for the most appropriate formalization of an argument, we have been led to consider further unstated assumptions (implicit premises) needed to reconstruct the argument as logically valid, and thus to ponder how much we may have departed from the original argument and to what extent we are still doing justice to the intentions of its author. We 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).[3] Reflecting on this experience, we have become motivated to work out a computer-assisted interpretive approach drawing on semantic holism, which is especially suited for finding meaning in theological and metaphysical discourse.

We want to focus our inquiry on the issue of understanding a particular type of arguments and the role computers can play in it. We are thus urged to distinguish the kind of arguments we want to address from others that, on the one hand, rely

[3] More specifically, Eder and Ramharter [19] propose 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.

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 that many of the concepts introduced into these and many other kinds of philosophical discussions remain quite fuzzy and unclear ("explicanda" in Carnap's 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 conceptual explication can be carried out *systematically* by giving definitions or axiomatizing conceptual interrelations, and then using automated reasoning tools to explore the space of possible logical inferences. This approach, which we name computational hermeneutics, will be illustrated in the case study presented in the last section.

Top-down versus Bottom-up Approaches to Meaning

Above we have discussed the challenge of finding meaning in religious arguments. Determining meanings in philosophical contexts, however, has generally been considered a problematic task, especially when one wants to avoid the kind of ontological commitments resulting from postulating the existence, for every linguistic expression, of some obscure abstract being in need of definite identity criteria (cf. Quine's slogan "no entity without identity"). We want to talk here of the *meaning* of a linguistic expression (particularly of an argument) as *that* which the interpreter needs to grasp in order to *understand* it, and we will relate this to such blurry things as the inferential role of expressions.

In a similar vein, 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).[4] Accordingly, we take the meaning of each sentence as resulting from the meaning of its constituent words

[4] Ideally, an argument would be analyzed as an island isolated from any external linguistic or pre-linguistic goings-on, to the extent that its validity would depend solely on what is explicitly stated (premises, inference rules, etc.); 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

(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 (theoretical or colloquial) of its main concepts and then working our way up to an understanding of its sentences and their inferential interrelations.[5]

The bottom-up approach is the one usually employed in the formal verification of arguments (logic as *ars iudicandi*). However, it leaves open the question of how to arrive at the meaning of words beyond our initial pre-understanding of them. This question is central to our project, since we are interested in understanding (logic as *ars explicandi*) more than mere verifying. Thus, we want to complement the atomistic bottom-up approach with a holistic top-down one, by proposing a computer-supported method aimed at determining the meaning of expressions from their inferential role vis-à-vis argument's validity (which is determined for the argument *as a whole*), much in the spirit of Donald Davidson's program of *radical interpretation*.[6]

Radical Interpretation and the Principle of Charity

What is the use of radical interpretation in religious and metaphysical discourse? The answer is trivially stated by Davidson himself, who convincingly argues that "all understanding of the speech of another involves radical interpretation" ([15], p. 125). Furthermore, the impoverished evidential position we are faced with when interpreting metaphysical and theological 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 utterances without relying on any prior understanding of their language.[7]

meaning compositionally (in particular, we see no place for idioms in philosophical arguments). Unsurprisingly, these demands are never met in their entirety in real-world arguments.

[5] There is a well-known tension between the holistic nature of inferential roles and a compositional account of meaning. In computational hermeneutics, we aim at showing both approaches in action (top-down and bottom-up), thus demonstrating their compatibility in practice. For a theoretical treatment of the relationship between compositionality and meaning holism, we refer the reader to [35, 33, 34].

[6] The connections between Davidson's truth-centered theory of meaning and theories focusing on the inferential role of expressions (e.g. [13, 27, 12]) have been much discussed in the literature. While some authors (Davidson included) see both holistic approaches as essentially different, others (e.g. [45], [28], p. 72) have come to see Davidson's theory as an instance of inferential-role semantics. We side with the latter.

[7] For an interesting discussion of the relevance of Davidson's philosophy of language in religious studies, we refer the reader to [25].

Davidson's program builds on the idea of taking the notion of truth as basic and extracting from it an account of translation or interpretation satisfying two general requirements: (i) it must reveal the compositional structure of language, and (ii) it can be assessed using evidence available to the interpreter [15, 18].

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 some object language L, a sentence of the form: «"s" is true in L iff p» (aka. T-schema) can be derived, where p acts as a translation of s into a sufficiently expressive metalanguage used for interpretation (note that in the T-schema the sentence p is being *used*, while s is only being *mentioned*). Thus, by virtue of the recursive nature of Tarski's definition of truth [43], the *compositional* structure of the object-language sentences becomes revealed.

From the point of view of computational hermeneutics, the sentence s is interpreted in the context of a given argument. The object language L thereby corresponds to the idiolect of the speaker (natural language plus some technical terms and background information), and the metalanguage is constituted by formulas of our chosen logic of formalization (some expressive logic XY) plus the turnstyle symbol $\vdash_{XY}$ signifying that an inference (argument) is valid in logic XY. As an illustration, consider the following instance of the T-schema used for some theological argument about monotheism: «"There is only one God" is true [in English, in the context of argument A] iff $A_1, A_2, ..., A_n \vdash_{HOL}$ "$\exists x.\ God\ x \wedge \forall y.\ God\ y \rightarrow y{=}x$"», where $A_1, A_2, ..., A_n$ correspond to the formalization of the premises of argument A and the turnstyle $\vdash_{HOL}$ corresponds to the standard logical consequence relation in higher-order logic (HOL). By comparing this with the T-schema («"s" is true in L iff p») we can notice that the *used* metalanguage sentence p can be paraphrased in the form: «"q" follows from the argument's premises [in HOL]» where the *mentioned* sentence q corresponds to the formalization (in some chosen logic) of the object sentence s. In this example we have considered a sentence playing the role of a conclusion which is being supported by some premises. It is however also possible to consider this same sentence in the role of a premise: «"There is only one God" is true [in the context of argument A] iff $A_1, A_2, ...,$ "$\exists x.\ God\ x \wedge \forall y.\ God\ y \rightarrow y{=}x$"$, ..., A_n \ \vdash_{HOL}\ C$»; now the truth of the sentence is postulated so that it can be used to validate C.[8] Most importantly, this example aims at illustrating how the interpretation of a sentence relates to its logical formalization and the inferential role it plays in a background

[8] We may actually want to weaken the double implication in this case, or work with an alternative notion of logical consequence. Moreover, other roles can be conceived for such a sentence in the context of an argument, for instance, it can also play the role of an unwanted conclusion: a sentence which we want to make sure it remains false no matter how we analyze the argument.

argument.

The second general requirement (ii) states that the interpreter has 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' that cause sentences to be true (or, in our case, arguments to be valid). In our approach, formal logic serves as a common ground for understanding. Computing the logical validity of a formalized argument constitutes the kind of objective (or, more appropriately, intersubjective) evidence needed to secure the adequacy of our interpretations, under the *charitable* assumption that the speaker follows (or at least accepts) similar logical rules as we do. In computational hermeneutics, the computer acts as an (arguably unbiased) arbiter deciding on the truth of a sentence in the context of an argument. In order to account for logical pluralism, computational hermeneutics targets the utilization of different kinds of classical and non-classical logics through the technique of *semantical embeddings* (see e.g. [6, 4]), which allows us to take advantage of the expressive power of classical higher-order logic (as a metalanguage) in order to embed the syntax and semantics of another logic (as an object language). Using the technique of semantical embeddings we can, for instance, embed a modal logic by defining the modal operators as meta-logical predicates. A framework for automated reasoning in different logics by applying the technique of semantical embeddings has been successfully implemented using automated theorem proving technology [21, 5].

Underlying his account of radical interpretation, there is a central notion in Davidson's theory: the *principle of charity*, which he holds as a condition for the possibility of engaging in any kind of interpretive endeavor. In a nutshell, the principle says that "we make maximum sense of the words and thoughts of others when we interpret in a way that optimizes agreement" [17]. The principle of charity builds on the possibility of intersubjective agreement about external facts among speaker and interpreter and can thus be invoked to make sense of a speaker's ambiguous utterances and, in our case, to presume (and foster) the validity of the argument we aim at interpreting. Consequently, in computational hermeneutics we assume from the outset that the argument's conclusions indeed follow from its premises and disregard formalizations that do not do justice to this postulate.

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,[9] automated reasoning has moved to close proximity of logic and philosophy, thanks to substantial theoretical developments in the last decades. Nevertheless, its 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 (inductive, abductive, analogical, etc.). 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. [9]). We will, on this occasion, restrain ourselves to the computer-supported interpretation of existing arguments, that is, to a situation where the given nodes/statements in the argument constitute a coarse grained "island proof structure" that needs to be rigorously assessed.

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 collection of such 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[10] 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 trying to prove 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 the-*

[9] 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.

[10] For instance, some of the proof methods commonly employed by the *Isabelle/HOL* proof assistant are: term rewriting, classical reasoning, tableaus, model elimination, ordered resolution and paramodulation.

orem proving has grown out of this endeavor and its software programs are known as *proof assistants.*[11]

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[12] and TPTP (Thousands of Problems for Theorem Provers) [42] 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 [36], whose problems are selected from the TPTP library.

Automated theorem provers (particularly focusing on higher order logics) 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) [40]. The monumental proof of Kepler's conjecture by Thomas Hales and his research team has been recently formalized and verified using the HOL-Light and Isabelle proof assistants as part of the *Flyspeck project* [26].

Isabelle [31] 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 semantical embeddings [6].

Part II: The Computational Hermeneutics Method

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, a main task of this paper is to illustrate a central *qualitative* advantage of computer-assisted argumentation: It enables a different, *holistic* approach to philosophical argumentation.

[11] A survey and system comparison of the most famous interactive proof assistants has been carried out in [44]. The results of this survey remain largely accurate to date.

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

Holistic Approach: Why Feasible Now?

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. Our philosopher is working with pen and paper and she follows some chosen proof procedure (e.g. natural deduction or sequent calculus). Depending on her calculation skills, this may take some minutes, if not much longer, to be carried out. It seems clear that she cannot allow herself many of such experiments on such conditions.

Now compare the above scenario to another one in which our working philosopher can carry out such an experiment in just a few seconds and with no effort, by employing an automated theorem prover. 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.[13]

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 take things to a whole new level (note that we are talking here of many hundreds of such trial-and-error 'experiments' that would take weeks or months if using pen and paper). Most importantly, this qualitative leap opens the door for the possibility of automating the process of logical analysis for natural-language arguments with regard to their subsequent computer-assisted critical evaluation.

The Approach

Computational hermeneutics is a holistic iterative enterprise, where we evaluate the adequacy of some candidate formalization of a sentence by computing the logical validity of the whole argument. We start with formalizations of some simple statements (taking them as tentative) and use them as stepping stones on the way to the formalization of other argument's sentences, repeating the procedure until arriving at a state of *reflective equilibrium*: A state where our beliefs and commitments have

[13] The situation is obviously idealized, since, as is well known, most of theorem-proving problems are computationally complex 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* [32, 1, 41, 8, 9] suggests, the lucky situation depicted above is not rare.

the highest degree of coherence and acceptability.[14] In computational hermeneutics, 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 premises and conclusions by making gradual adjustments while getting automatic feedback about the suitability of our speculations. In this fashion, by engaging in a dialectic questions-and-answers ('trial-and-error') interaction with the computer, we work our way towards a proper understanding of an argument by circular movements between its parts and the whole (hermeneutic circle).

A rough outline of the iterative structure of the *computational hermeneutics* approach is as follows:

1. **Argument reconstruction** (initially in natural language):

    a. **Add or remove sentences and choose their truth-values.** Premises and desired conclusions would need to become true, while some other 'unwanted' conclusions would have to become false. Deciding on these issues expectedly involves a fair amount of human judgment.

    b. **Establish inferential relations,** i.e., determine the extension of the logical consequence relation: which sentences should follow (logically) from which others. This task can be done manually or automatically by letting our automated tools find this out for themselves, provided the logic for formalization has been selected and argument has already been roughly formalized (hence the mechanization of this step becomes feasible only after at least one outermost iteration). Automating this task frequently leads to the simplification of the argument, since current theorem provers are quite good at detecting idle axioms (see e.g. Isabelle's *Sledgehammer* tool [10]).

2. **Selection of a logic for formalization,** guided by determining the logical structure of the natural-language sentences occurring in the argument. This task can be partially automated (using the *semantical embeddings* technique) by

[14] We have been inspired by John Rawls' notion of *reflective equilibrium* as a state of balance or coherence between a set of general principles and particular judgments (where the latter follow from the former). We arrive at such a state through a deliberative give-and-take process of mutual adjustment between principles and judgments. More recent methodical accounts of reflective equilibrium have been proposed as a justification condition for scientific theories [20] and objectual understanding [2], and also as an approach to logical analysis [39].

searching a catalog of different embedded logics (in HOL) and selecting a candidate logic (modal, free, deontic, etc.) satisfying some particular syntactic or semantic criteria.

3. **Argument formalization (in the chosen logic),** while getting continuous feedback from our automated reasoning tools about the argument's correctness (validity, consistency, non-circularity, etc.). This stage is itself iterative, since, for every sentence, we charitably (in the spirit of the *principle of charity*) try several different formalizations until getting a correct argument. Here is where we take most advantage of the real-time feedback offered by our automated tools. Some main tasks to be considered are:

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

   b. **Vary the logical form of already formalized sentences.** This can be done systematically and automatically by relying upon a catalog of (consistent) logical variations of formulas (see *semantical embeddings*) and the output of automated tools (ATPs, model finders, etc.).

   c. **Bring related terms together,** either by introducing definitions or by axiomatizing new interrelations among them. These newly introduced formulas can be translated back into natural language to be integrated into the argument in step (1.a), thus being disclosed as former *implicit* premises. The process of searching for additional premises with the aim of rendering an argument formally correct can be seen as a kind of abductive reasoning ('inference to the best explanation') and thus needs human support (at least at the current state of the art).

4. **Are termination criteria satisfied?** That is, have we arrived at a state of *reflective equilibrium*? If not, we would come back to some early stage. Termination criteria can be derived from the adequacy criteria of formalization found in the literature on logical analysis (see e.g. [3, 14, 38, 39]). An equilibrium may be found after several iterations without any significant improvements.[15]

[15] In particular, inferential adequacy criteria lend themselves to the application of automated deduction tools. Consider, for instance, Peregrin and Svoboda's [39] proposed criteria:
(i) The *principle of reliability*: "ϕ counts as an adequate formalization of the sentence S in the logical system L only if the following holds: If an argument form in which ϕ occurs as a premise or as the conclusion is valid in L, then all its perspicuous natural language instances in which S appears as a natural language instance of ϕ are intuitively correct arguments."

> Furthermore, the introduction of automated reasoning and linguistic analysis tools makes it feasible to apply these criteria to compute, in seconds, the degree of ’fitness’ of some candidate formalization for a sentence (in the context of an argument).

Part III: Lowe’s Modal Ontological Argument

In this section, the main contribution of this article, we illustrate the computer-supported interpretation of a variant of St. Anselm’s ontological argument for the existence of God, using *Isabelle/HOL*.[16] This argument, which was introduced by the philosopher E. J. Lowe in an article named “A Modal Version of the Ontological Argument" [30], serves here as an exemplary case for an interesting and sufficiently complex, systematic argument with strong ties to metaphysics and religion. The interpretation of Lowe’s 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, i.e., 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 (i.e. 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 in order to use the modal *S5* axioms to deduce its necessary ex-

(ii) The *principle of ambitiousness*: “ϕ is the more adequate formalization of the sentence S in the logical system L the more natural language arguments in which S occurs as a premise or as the conclusion, which fall into the intended scope of L and which are intuitively perspicuous and correct, are instances of valid argument forms of S in which ϕ appears as the formalization of S." ([39] pp. 70-71).

[16] We refer the reader to [22] for further details. That computer-verified article has been completely written in the Isabelle proof assistant and thus requires some familiarity with this system.

istence as a conclusion.[17] 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 methodical 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:

(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:

(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.)

[17] As shown in [8], modal logic *KB* actually suffices to prove Scott's variant of Gödel's argument; this was probably not known to Lowe though.

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

Lowe also introduces some informal definitions which should help the reader to understand some 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 use computational hermeneutics to interpret iteratively the argument shown above (by reconstructing it formally in different variations and in different logics). 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".[18]

In order to shed light on 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:[19]

[18] 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. 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.

[19] 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.

(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.

The choice of a modal logic for formalization has brought to the foreground an interesting technical constraint: The Isabelle proof assistant (as well as others) does not natively support modal logics. We have used, therefore, a technique known as *semantical embedding*, which allows us to take advantage of the expressive power of higher-order logic in order to embed the syntax and semantics of an object language. Here we draw on previous work on the embedding of multimodal logics in HOL [6], which has successfully been applied to the analysis and verification of ontological arguments (e.g. [9, 8, 7, 23]). Using this technique, we can embed a modal logic K by defining the $\Box$ and $\Diamond$ operators using restricted quantification over the set of *reachable* worlds (using a *reachability relation* R as a guard). Note that, in the following definitions, the type *wo* is declared as an abbreviation for *w⇒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**) — Reachability relation
abbreviation *mbox* :: *wo⇒wo* (□-)
  where $\Box\varphi \equiv \lambda w.\forall v.\ (w\ \mathbf{R}\ v) \longrightarrow (\varphi\ v)$
abbreviation *mdia* :: *wo⇒wo* (◇-)
  where $\Diamond\varphi \equiv \lambda w.\exists v.\ (w\ \mathbf{R}\ v) \wedge (\varphi\ v)$

The ‘lifting’ of the standard logical connectives to type *wo* is straightforward. Validity is consequently defined as truth in *all* worlds and represented by wrapping the formula in special brackets ($\lfloor - \rfloor$).

abbreviation *valid::wo⇒bool* (⌊-⌋) **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*: $\lfloor(\Box(\varphi \rightarrow \psi)) \rightarrow (\Box\varphi \rightarrow \Box\psi)\rfloor$ **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).[20]

We hence restrict our quantifiers so that they range only over those entities that 'exist' (i.e. are actualized) at a given world. This approach is known as *actualist quantification* and is implemented, using the semantical 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.

consts *isActualized*::*e*⇒*wo* (**infix** *actualizedAt*)

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

The corresponding binder syntax is defined below.

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

We use a model finder (Isabelle's Nitpick tool [11]) to verify that actualist quantification validates neither the Barcan formula nor its converse. For the conjectured lemma, Nitpick finds a countermodel, i.e. a model (satisfying all axioms) which falsifies the given formula. The formula is consequently non-valid (as indicated by the Isabelle's "oops" keyword).

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

[20] 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 on the metaphysics of modality (see [46]).

Unrestricted (aka. possibilist) 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) the concept of existence becomes formalized (explicated) in the usual form by using a restricted particular quantifier ($\approx$ stands for the unrestricted identity relation on all objects), (ii) necessariness becomes formalized as existing necessarily, and (iii) contingency becomes formalized as existing possibly but not necessarily.

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

definition *Necessary*::*e*⇒*wo* **where** *Necessary* $x \equiv\ \Box E!\ x$
definition *Contingent*::*e*⇒*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-existing objects). The semantics is *positive* because we have placed no restriction regarding predication on non-existing 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, S5, etc.*) by restricting the reachability 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.[21] Thus, we start by formalizing concreteness as a *primitive* world-dependent predicate and then derive abstractness from it, namely as its negation.

[21] 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."

consts *Concrete*::*e*⇒*wo*
abbreviation *Abstract*::*e*⇒*wo* **where** *Abstract x* ≡ ¬(*Concrete x*)

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

abbreviation *Godlike*::*e*⇒*wo* **where** *Godlike x* ≡ *Necessary x* ∧ *Concrete x*

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

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

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 terms 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*⇒*e*⇒*wo* (**infix** *dependsOn*)
definition *Dependent*::*e*⇒*wo* **where** *Dependent x* ≡ $\exists^A y.\ x\ \mathit{dependsOn}\ y$
abbreviation *Independent*::*e*⇒*wo* **where** *Independent x* ≡ ¬(*Dependent x*)

We have formalized ontological dependence as a *primitive* world-dependent relation and refrained from any explication (as suggested by Lowe).[22] Moreover, an entity is *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: $\lfloor \forall^A x.\ \mathit{Abstract}\ x \rightarrow \mathit{Dependent}\ x \rfloor$ **and**

[22] 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* [29] for further discussion.

P4: $\lfloor\forall^A x.\ \mathit{Dependent}\ x \longrightarrow (\exists^A y.\ \mathit{Independent}\ y \wedge x\ \mathit{dependsOn}\ y)\rfloor$

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.[23] These expressions rely on a single binary relation, which will initially be taken as *primitive.* This relation has been called *metaphysical explanation* by Lowe.[24]

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

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

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 *metaphysical explanation* to *ontological dependence.*[25]

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

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.[26] 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 (a

[23] Note that we have omitted the expressions “can" and “needs to" in our formalization, since they seem to play here only a rhetorical role. As in the case of attributive adjectives discussed before, 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.*

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

[25] 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".

[26] 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."

screenshot showing Nitpick's text-based representation of such a model is provided below).

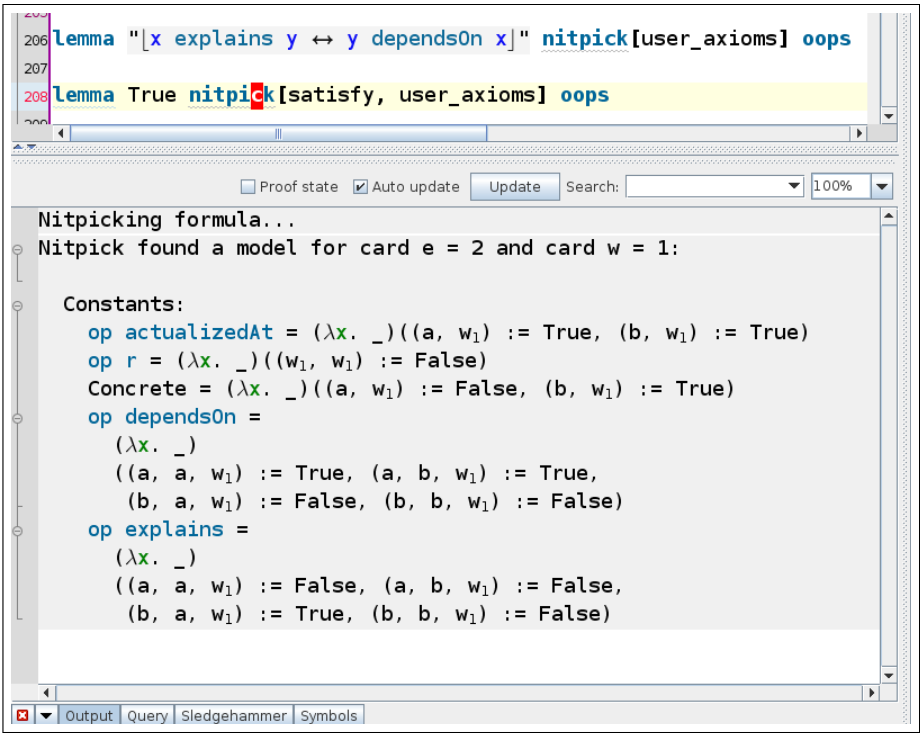
lemma $\lfloor\forall\, x\ y.\ x\ \mathit{explains}\ y \leftrightarrow y\ \mathit{dependsOn}\ x\rfloor$ **nitpick**[*user-axioms*] **oops**

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 $\lfloor\forall\, x.\ \mathit{Explained}\ x \leftrightarrow \mathit{Dependent}\ x\rfloor$
  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 search for a model satisfying some tautological sentence (here we use a trivial 'True' proposition), thus demonstrating the satisfiability of the argument's axioms. Nitpick's output is a text-based representation of the found model (or a message indicating that no model, up to a predefined cardinality, could be found). This information is very useful to inform our future decisions. The screenshot below (taken from the Isabelle proof assistant) shows the model found by Nitpick, which satisfies the argument's formalized premises:

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



```
lemma "⌊x explains y ↔ y dependsOn x⌋" nitpick[user_axioms] oops

lemma True nitpick[satisfy, user_axioms] oops
```

```
Nitpicking formula...
Nitpick found a model for card e = 2 and card w = 1:

  Constants:
    op actualizedAt = (λx. _)((a, w1) := True, (b, w1) := True)
    op r = (λx. _)((w1, w1) := False)
    Concrete = (λx. _)((a, w1) := False, (b, w1) := True)
    op dependsOn =
      (λx. _)
      ((a, a, w1) := True, (a, b, w1) := True,
       (b, a, w1) := False, (b, b, w1) := False)
    op explains =
      (λx. _)
      ((a, a, w1) := False, (a, b, w1) := False,
       (b, a, w1) := True, (b, b, w1) := False)
```

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 consists of two individual objects a and b and and a single world w_1, which is not connected via the reachability relation R to itself. We furthermore have in world w_1: b is concrete, a is not; a depends on b and itself, while b depends on no other object; b is the only object that explains a and a explains no object.

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. that all beings necessarily exist) or *modal collapse* (i.e. that all truths are necessary). Since both would trivially validate the argument.

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 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 use automated theorem provers. Isabelle comes with various different provers tailored for specific kinds of problems and thus employing different approaches, strategies and heuristics. We typically make extensive use of Isabelle's *Sledgehammer* tool [10], which integrates several state-of-the-art external theorem provers and feeds them with different combinations of axioms and the conjecture in question. If successful, *Sledgehammer* returns valuable dependency information (the exactly required axioms and definitions to prove a given conjecture) back to Isabelle, which then exploits this information to (re-)construct a trusted proof with own, internal proof automation means. The entire process often only takes a few seconds.

By using Sledgehammer we can here verify the validity of our partial conclusions (C1, C5 and C7) and even find the premises they rely upon.[27]

(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 reachability relation R is an equivalence relation. This exploits the *Sahlqvist correspondence* which relates modal axioms to constraints on a model's reachability relation: reflexivity, symmetry, seriality, transitivity and euclideanness imply axioms T, B, D, IV, V respectively (and also the other way round).

[27] We prove theorems in Isabelle here by using the keyword "by" followed by the name of an Isabelle-internal and thus trusted proof method (generally, 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). As explained, these methods were automatically suggested and applied by the Sledgehammer tool. The interactive user in fact does not need to know, or learn, much about these methods in the beginning (he will benefit a lot though, if he does).

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**

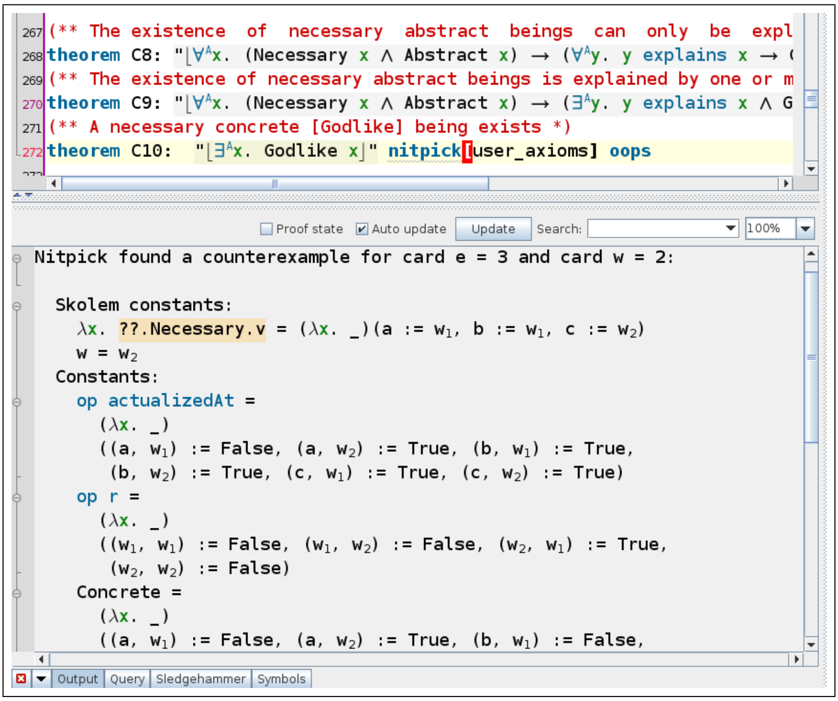
(C9) *The existence of necessary abstract beings is explained by one or more necessary concrete (Godlike) beings.*

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**

Note that 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 (counter-)model found. We present below another screenshot showing Nitpick's counterexample for C10:



```
(** The existence  of  necessary  abstract  beings  can  only  be  expl
theorem C8: "⌊∀^Ax. (Necessary x ∧ Abstract x) ⟶ (∀^Ay. y explains x ⟶ (
(** The existence of necessary abstract beings is explained by one or m
theorem C9: "⌊∀^Ax. (Necessary x ∧ Abstract x) ⟶ (∃^Ay. y explains x ∧ G
(** A necessary concrete [Godlike] being exists *)
theorem C10:  "⌊∃^Ax. Godlike x⌋" nitpick[user_axioms] oops
```

```
Nitpick found a counterexample for card e = 3 and card w = 2:

  Skolem constants:
    λx. ??.Necessary.v = (λx. _)(a := w1, b := w1, c := w2)
    w = w2
  Constants:
    op actualizedAt =
      (λx. _)
      ((a, w1) := False, (a, w2) := True, (b, w1) := True,
       (b, w2) := True, (c, w1) := True, (c, w2) := True)
    op r =
      (λx. _)
      ((w1, w1) := False, (w1, w2) := False, (w2, w1) := True,
       (w2, w2) := False)
    Concrete =
      (λx. _)
      ((a, w1) := False, (a, w2) := True, (b, w1) := False,
```

By employing the Isabelle proof assistant we have proven 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.

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 reachable 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: $\lfloor\forall x.\ \mathit{Abstract}\ x \rightarrow \Box \mathit{Abstract}\ x\rfloor$

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

Again, we have used model finder Nitpick to get a counterexample for C10, so the former restriction is not enough to prove this conclusion. We try postulating further restrictions on the reachability relation *R*, which, taken together, would amount to it being an equivalence relation. This would make for a modal logic *S5* (see *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*: $\lfloor\exists^{A} x.\ \mathit{Godlike}\ x\rfloor$
  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*: $\lfloor\forall x.\ \mathit{Concrete}\ x \rightarrow E!\ x\rfloor$

We carry out the usual ‘sanity checks’ to make sure the argument has not become trivialized.[28]

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 $\lfloor \varphi \rightarrow \Box\varphi \rfloor$
  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 another automated reasoning tool: Isabelle’s *Sledgehammer* [10] to search for a proof.

theorem *C10*: $\lfloor \exists^{A} x.\ \mathit{Godlike}\ x \rfloor$ **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.* Thus, by the end of this series of iterations, we have seen 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* ($\varphi \rightarrow \Box\Diamond\varphi$). 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

In this iteration series we want to explore the critical potential of computational hermeneutics. In this slightly simplified variant (without the implicit premises stated in the previous version), premises P1 to P5 remain unchanged, while none of the last three premises (P6 to P8) show up anymore. Those last premises have been introduced by Lowe in order to interrelate the concepts of explanation and dependence in such a way that they play somewhat opposite roles, without one being the inverse of the other. Nonetheless, we will go all the way and assume that explanation and dependence are indeed inverse relations, for we want to understand how the interrelation of these two concepts affects the validity of the argument.

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

Let us first prove the relevant partial conclusions.

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

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

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

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

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

theorem *C10*: $\lfloor \exists^{A} x.\ \mathit{Godlike}\ x\rfloor$
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*: $\lfloor \exists^{A} x.\ \mathit{Godlike}\ x\rfloor$ **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* (in the context of Lowe's argument) has changed after the introduction of an additional axiom constraining both: they are now inverse relations. This new understanding of the inferential role of the above concepts of dependence and 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 expressions 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*: $\lfloor\forall^{A} x\ y.\ x\ \mathit{dependsOn}\ y \rightarrow \Box(E!\ x \rightarrow E!\ y)\rfloor$

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

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

theorem *C10*: $\lfloor\exists^{A} x.\ \mathit{Godlike}\ x\rfloor$
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 $y\ \mathit{explains}\ x \equiv x\ \mathit{dependsOn}\ y$

lemma *P5*: $\lfloor\neg(\exists^{A} x.\ \exists^{A} y.\ \mathit{Contingent}\ y \wedge \mathit{Necessary}\ x \wedge y\ \mathit{explains}\ x)\rfloor$
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 and being-dependent (axioms "Existence-def" and "Dependent-def" respectively) 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

In the next iterations, we want to explore once again the critical potential of computational hermeneutics by challenging another of the author's claims: that this argument is a *modal* one. A new simplified version of Lowe's argument is obtained after abandoning the concept of existence altogether 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.[29]

	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 being-dependent as a *primitive* predicate (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.

[29] 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."

```
consts Dependent::e⇒wo
abbreviation Independent::e⇒wo where Independent x ≡ ¬(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.

```
axiomatization where
P2: ⌊∃ x. Necessary x ∧ Abstract x⌋ and
P3: ⌊∀ x. Abstract x → Dependent x⌋ and
P4: ⌊∀ x. Dependent x → (∃ y. Independent y ∧ x dependsOn y)⌋ and
P5: ⌊¬(∃ x. ∃ y. Contingent y ∧ Necessary x ∧ y explains x)⌋

theorem C10: ⌊∃ x. Godlike x⌋ using  P2 P3 P4 P5 by blast
```

Note that, in the axioms above, all restricted (actualist) quantifiers have been changed into unrestricted (possibilist) quantifiers, following the elimination of the concept of existence from our argument: Our quantifiers now range over all beings, because all beings exist. Also note that 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 hermeneutics* approach in those cases where we must start our interpretive endeavor with no *explicit* understanding of the concepts involved. 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⇒wo
consts Abstract::e⇒wo
consts Necessary::e⇒wo
consts Contingent::e⇒wo
consts dependence::e⇒e⇒wo (infix dependsOn)
consts explanation::e⇒e⇒wo (infix explains)
consts Dependent::e⇒wo
abbreviation Independent::e⇒wo where Independent x ≡ ¬(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 “necessari*ly* abstract" and “necessari*ly* concrete" respectively. With this new adverbial reading 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. It can be argued that in this variant we are not concerned with the interpretation of the *original* natural-language argument anymore. We are rather interested 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*⇒*wo* **where** $\mathit{Godlike}\ x \equiv \Box \mathit{Concrete}\ x$

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

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

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

theorem *C10*: $\lfloor\exists\, x.\ \mathit{Godlike}\ x\rfloor$ **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 reachability relation by enforcing its symmetry (i.e. assuming the modal axiom *B*).

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

theorem *C10*: $\lfloor\exists\, x.\ \mathit{Godlike}\ x\rfloor$ **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 being-necessary (necessariness), being-contingent (contingency), explanation, dependence, abstractness, concreteness, etc. has gradually evolved thanks to the iterative holistic method 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 do not claim that this formalization of Lowe's argument is its best or most adequate one; 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 our view, while the third variant may be the closest one to Lowe's original formulation, it is this latter (seventh) variant the one which strikes the best balance between interpretation and critical assessment of this argument. We encourage the reader to continue with this process until arriving to his/her own *reflective equilibrium* (possibly by building upon our computer-verified work [22] available at the *Archive of Formal Proofs*).[30]

Conclusion

We have argued for the role of formal logic as an *ars explicandi* and the possibility of applying it to foster our understanding of rational arguments (in particular metaphysical and theological ones). We understand the give-and-take process aiming at an adequate formal reconstruction of a natural-language argument in itself as a kind of interpretive endeavor. Moreover, we have argued that, by using automated reasoning technology to systematically explore the many different inferential possibilities latent in a formalized argument, we can make explicit the inferential role played by its constituent expressions and thus better understand their meaning in the given interpretation context.

As a computer-assisted method, computational hermeneutics aims at complementing our human ingenuity with the data-processing power of modern computers and at using this synergy to make interpretation more effective. In a similar vein, we currently work on how to apply this approach in the computer science field of *natural language understanding*. Specifically, we want to tackle the problem of formalization: how to search *methodically* for the most appropriate logical form(s) of a given natural-language argument, by casting its individual statements into expressions of some sufficiently expressive logical language. Being able to automatically extract a formal representation for some piece of natural-language discourse, by taking into

[30] The Archive of Formal Proofs (`www.isa-afp.org`) is a collection of proof libraries, examples, and larger scientific developments, mechanically checked using the Isabelle proof assistant. It is organized in the way of a scientific journal and submissions are refereed.

account its holistically-determined logical location in a web of possible inferences, is an important step towards the deep semantic analysis and critical assessment of non-trivial natural-language discourse. Further applications in areas like knowledge/ontology extraction, semantic web and legal informatics are currently being contemplated.

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. Baumberger and G. Brun. Dimensions of objectual understanding. *Explaining understanding. New perspectives from epistemology and philosophy of science*, pages 165–189, 2016.

[3] M. Baumgartner and T. Lampert. Adequate formalization. *Synthese*, 164(1):93–115, 2008.

[4] C. Benzmüller. A top-down approach to combining logics. In J. Filipe and A. Fred, editors, *Proc. of the 5th International Conference on Agents and Artificial Intelligence (ICAART)*, volume 1, pages 346–351, Barcelona, Spain, 2013. SCITEPRESS – Science and Technology Publications, Lda.

[5] C. Benzmüller. Recent successes with a meta-logical approach to universal logical reasoning (extended abstract). In S. A. da Costa Cavalheiro and J. L. Fiadeiro, editors, *Formal Methods: Foundations and Applications - 20th Brazilian Symposium, SBMF 2017, Recife, Brazil, November 29 - December 1, 2017, Proceedings*, volume 10623 of *Lecture Notes in Computer Science*, pages 7–11. Springer, 2017.

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

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

[8] C. Benzmüller and B. Woltzenlogel Paleo. Automating Gödel's ontological proof of God's existence with higher-order automated theorem provers. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *ECAI 2014*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 93 – 98. IOS Press, 2014.

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

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

[11] 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.

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

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

[14] G. Brun. Reconstructing arguments. formalization and reflective equilibrium. *Logical analysis and history of philosophy*, 17:94–129, 2014.

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

[16] D. Davidson. *Essays on actions and events: Philosophical essays*, volume 1. Oxford University Press on Demand, 2001.

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

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

[19] 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.

[20] C. Elgin. *Considered judgment*. Princeton University Press, 1999.

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

[22] 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.

[23] 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.

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

[25] 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.

[26] 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.

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

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

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

[30] 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.

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

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

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

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

[35] 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.

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

[37] J. Peregrin. *Inferentialism: Why rules matter*. Springer, 2014.

[38] J. Peregrin and V. Svoboda. Criteria for logical formalization. *Synthese*, 190(14):2897–2924, 2013.

[39] 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.

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

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

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

[43] A. Tarski. The concept of truth in formalized languages. *Logic, semantics, metamathematics*, 2:152–278, 1956.

[44] 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.

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

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

Received \jreceived