Computational Metaphysics: Kurt Gödel's Ontological Argument
Extended Abstract

Edith Förster
Otto-Friedrich-Universität Bamberg

Abstract

Recent research identified a possible cross-fertilization of philosophy and computer science. Computational means might help us to better understand philosophical arguments and discover insights that weren't accessible by traditional means. At the same time, philosophical questions might challenge existing technical solutions and inspire refinement. Kurt Gödel's variant of the Ontological Argument constitutes a precedent and will be featured in this paper.

1 Introduction

Dwelling on the existence of god and whether one could find a definitive proof for it has been a venerable and fascinating endeavor for philosophers over the course of centuries, starting back in the middle ages and even earlier. At a time when the topic fell out of vogue in the 20th century, logician Kurt Gödel elaborated his own version of an Ontological Argument proving the existence of god. His version involves methods of formalizing arguments and proofs which developed at the same time as machines that from their very fundaments in hardware on are based on nothing else but logic. It thus becomes possible to translate philosophical arguments in a language that allows them to be processed by machines. This might allow for new insights concerning the philosophical questions, even maybe – as an ultimate goal – to the point where it becomes possible to settle all ongoing debates and quarrels on the subject. While for the latter, there is still a big question mark, at least it is possible to show how exactly the involvement of computers can help in the discovery of philosophical insights – while at the same time challenging computer scientists to develop new technical methods to do justice to such a task. Gödel's Ontological Argument, featured in this paper relying on the original article that presented "Computational Metaphysics",[Kirchner *et al.*, 2019] can be seen as a starting point, illustrating the whole new methodology and its potential for many other topics as a case study.

2 Definitions and background

2.1 Computational Metaphysic

The article invokes the following definition in its abstract: "Computational philosophy is the use of mechanized computational techniques to unearth philosophical insights that are either difficult or impossible to find using traditional philosophical methods. That part of computational philosophy that focuses on metaphysics is computational metaphysics."[Kirchner *et al.*, 2019] As a working definition of metaphysics for the current context, Stanford Encyclopedia might be a basis. It differentiates between "old" and "new" metaphysics. In the classical approach, this discipline addressed topics like Being as such, first causes, unchanging things and other questions, beyond mere physical "things". Modern Metaphysics, on the other side, is concerned with Modality, Space and Time, Persistence and Constitution, Causation, Freedom and Determinism, the Mental and the Physical.[van Inwagen and Sullivan, 2021]

2.2 Ontological Arguments

Ontological Arguments have a long tradition in philosophy, with Anselm of Canterbury being the first one to present such an argument. Again referring back to Stanford Encyclopedia, it can be said that these arguments share the same basic idea: Their first step is to prove that the existence of a being that is god is *possible*. As a second step, they try to show that from this possibility it follows logically that god indeed does exist.[Oppy, 2021] Keeping this blueprint in mind will become helpful when diving into the details of Kurt Gödel's form of the argument.

2.3 Methodologies

This section summarizes the technical details provided in the original paper on Computational Metaphysics.[Kirchner *et al.*, 2019] For applying computational methods to philosophical problems, a language of much greater expressivity than only first order logic is required. This confinement initially limited the usefulness of automated theorem provers for philosophy to a great extent. In a series of papers, researchers tried to develop solutions to this problem, ending up in particular with two methodologies:
1) Shallow Semantic Embeddings (SSEs). This denotes using the syntactic capabilities of a higher-order theorem prover

such as Isabelle/HOL to represent the semantics of a target logic and define the original syntax of this logic within the prover. This is sufficient for modal, higher-order, and non-classical logics and for investigating the validity of philosophical arguments. If it is possible to prove an argument is true in the SSE, soundness for the proof in the target logic follows immediately. This SSE technique was the starting point for a natural encoding of Gödel's modern variant of the ontological argument in second-order S5 modal logic, with various formalizations and variants, as we'll see.
2) Developing additional abstraction layers: Here, the goal is to represent the deductive system of philosophical theories with a reasoning system that goes beyond the deductive systems of classical modal logics. This extends the concept of SSEs to make interactive and automated reasoning in Isabelle/HOL possible and allows for a deeper analysis of the semantic properties of the target logic, such as completeness.

3 Gödel's proof

3.1 Part 1: God's Existence is possible

Following the blueprint just explained, Gödel in the first part of his argument tries to establish the possibility of a godlike being's existence. He starts with a definition and three axioms, deriving three theorems from them.[1]

Definition 1 *Being Godlike is equivalent to having all positive properties.*
Axiom 1 *Exactly one of a property or its negation is positive.*
Axiom 2 *Any property entailed by a positive property is positive.*
Axiom 3 *The combination of any collection of positive properties is itself positive.*

From the axioms A2 and A2 follows theorem T1:

Theorem 1 *Every positive property is possibly instantiated (if a property P is positive, then it is possible that some being has property P).*

From definition D1 and axiom A3 follows:

Theorem 2 *Being Godlike is a positive property.*

From T1 and T2 follows:

Theorem 3 *Being Godlike is possibly instantiated.*

3.2 Part 2: God's Existence is necessary, if possible

In the second part of his argument, Gödel introduces two additional definitions and two new axioms, leading to three theorems. At the very end, he concludes that the godlike being must necessarily exist.

[1]The presentation of Gödel's argument given here follows a pre-printed paper.[Benzmüller and Fuenmayor, 2018]

Definition 2 *A property E is the essence of an individual x iff <u>x has E and</u> all of x's properties are entailed by E.*[2]
Axiom 4 *Positive (negative) properties are necessarily positive (negative).*

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

Theorem 4 *Being Godlike is an essential property of any Godlike individual.*
Definition 3 *Necessary existence of an individual is the necessary instantiation of all its essences.*
Axiom 5 *Necessary existence is a positive property.*

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

Theorem 5 *The property of) being Godlike, if instantiated, is necessarily instantiated.*

Finally, from T3, T5 (together with some implicit modal axioms[3] that will be discussed later on, e.g. S5) the existence of (at least a) God follows:

Theorem 6 *Being Godlike is actually instantiated.*

4 Implementation and different variants

4.1 The inconsistency problem and Scott's variant

The definition of essence introduced by Gödel comes with certain problems. It states: *A property Y is the essence of an individual x iff all of x's properties are entailed by Y, i.e., iff* $\forall Z(Zx \rightarrow Y \Rightarrow Z)$, where $Y \Rightarrow Z$ means that $\Box\forall x(Y\,x \rightarrow Zx)$. This definition doesn't require an individual *x* to actually exemplify its essence – although intuitively, one might assume this. But in the form given, it would be possible to derive as a lemma that an empty property (e.g., being non-self-identical) is an essence of any individual *x*. As a result, the proof is no longer valid. Scott added a conjunct to the definition of essence: *A property Y is the essence of an individual x iff x has property y and all of x's properties are entailed by Y.* This preserves the argument's consistency. Philosophers had missed the inconsistency, but it got detected in an experiment with the automated theorem prover Leo-II.[Kirchner *et al.*, 2019]

4.2 Results discovered with ATPs

Additionally, ATPs were able to discover which axioms were otiose, and which properties of the modal operator were required for the argument. It turned out that even as emended by Scott, the argument implies modal collapse: $\varphi \equiv \Box\varphi$, in other words: There can only be models of the premises to the argument in which exactly one possible world exists. This

[2]The underlined part in definition D2 has been added by Scott. Gödel originally omitted this part; see the section on Scott's variant.

[3]The modal logic S5, for which the proof was intended, features among others the following principle: *if a statement A is possibly the case, then A is necessarily possibly the case.* This has been debated among researchers.

implies that everything is determined, which creates a great conflict to the strong human intuition of having free will and that things could have been otherwise. As a result, efforts were made to preserve the argument's validity while getting rid of the undesired side-effect of modal collapse, leading to additional variants of the argument.[Kirchner *et al.*, 2019]

4.3 Anderson's variant

One idea was to introduce a change to the premises that govern the primitive notions of a positive property. Originally, the axiom reads: *Y is positive if and only if the negation of Y is non-positive.* Anderson's suggestion is to discard one part and preserve only one direction of the biconditional, namely: *If a property is positive, then its negation is not positive.* As a result, now "indifferent" properties can exist, in other words properties that are neither positive nor the negation of a positive property.[Benzmüller and Fuenmayor, 2018] This change, of course, also has an effect on the argument's validity. In order to make the argument valid again, Anderson proposes modifications to premises that govern definition of essence and Godlikeness:

essence*: *A property E is an essence* of an individual x if and only if all of x's necessary/essential properties are entailed by E and (conversely) all properties entailed by E are necessary/essential properties of x.*

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

This makes the argument valid again, as could be proven by computational means. At the same time, it has introduced a certain vagueness into the conception of Godlikeness, as the new definition allows now distinct Godlike entities to exist, which differ only by properties that are neither positive nor non-positive.[Kirchner *et al.*, 2019; Benzmüller and Fuenmayor, 2018]

4.4 Fitting's variant

Another researcher suggests that there is a subtle ambiguity in Gödel's argument, stemming from not making explicit whether the notion of a positive property applies to extensions or intensions of properties. An intensional concept may be exemplified by different objects under different circumstances: The concept of "first human ever to reach the south pole" might in some possible world have been instantiated by Robert Scott and his team instead of Roald Amundsen' expedition. However, an object referred to by extensional terms is always the same in all imaginable circumstances; thus the term is said to "designate rigidly". For proper names and enumerations of individuals this is arguably the case.[Benzmüller and Fuenmayor, 2018]

To better understand the difference between both interpretations of "being Godlike", Fitting formalizes Scott's emendation in an intensional type theory, such that he can now encode and compare both alternatives. The property of being Godlike would in Fitting's interpretation be represented by the λ-expression $[\lambda x\ \forall Y\ (PY \rightarrow Yx)]$, where P is the second-order property of being a positive property. In this understanding, the variable Y in the λ-expression ranges over properties whose extensions are fixed from world to world, while P is a second-order property whose extension among the first-order properties can vary. Gödel's original version of the argument presents positiveness and essence as second-order properties. Fitting however suggests that the expressions denoting the first-order properties to which positiveness and essence apply are not rigid designators and might have different extensions in different circumstances. In his variant, positiveness and essence apply only to the extensions of first-order properties, where the expressions denoting these extensions are rigid designators. For a positive property in some possible world, this would imply that its extension is the same in all circumstances. This leads to a rigid property, in other words a property that is exemplified by exactly the same individual(s) in all possible worlds. In Fitting's understanding, only rigid properties can be positive.[Kirchner *et al.*, 2019]

The overall result is the following one: This variant can prevent modal collapse and also leave Gödel's argument (including Scott's amendment) largely unchanged, up to an additional (implicit) weaker assumption: that any individual having some positive property has it necessarily, this positive property becoming constitutive for its identity. What it took to get there was further formalization in a more complex and expressive logic.[Benzmüller and Fuenmayor, 2018]

5 Generalizability and Future Research

The precedent of the Gödelean Ontological Argument revealed that interdisciplinary studies of metaphysics with computational means yield further insights and a refinement in the understanding of the implications that come with different assumptions. Implementing logics and metaphysical theories led to the development of new technologies for automated reasoning environments. At the same time, computational means allowed to detect inconsistencies in an argument that were not found by philosophers.

A next step could be to tackle questions such as the following: Are there any alternative reconstructions in which some of the axioms of the variants described could be derived as theorems? How prove that there is a "unique" object that exemplifies the property of being God, without running into modal collapse? How might various notions of equality in various logical settings help with restoring uniqueness? Would different notions of equality yield monotheism or not? From a computer scientist's perspective, one might ask whether the property of "being Godlike" could be formalized in a way resembling the mathematical concept of an ultrafilter. With a glance back into the history of philosophy, it could also be asked whether the merging of necessity and possibility (e.g. modal collapse) is really the denial of free will, as there have been many philosophers who claimed otherwise.

With both disciplines, computer science and philosophy, heavily relying on reasoning and argumentation, this seems to indicate great potential for further cross-fertilization – not only for the topic of Ontological Arguments, but also for other subjects such as normative reasoning with deontic logic.

References

[Benzmüller and Fuenmayor, 2018] Christoph Benzmüller and David Fuenmayor. Can computers help to sharpen our understanding of ontological arguments? (invited paper). 08 2018.

[Kirchner *et al.*, 2019] Daniel Kirchner, Christoph Benzmüller, and Edward N. Zalta. Computer science and metaphysics: A cross-fertilization. *Open Philosophy*, 2(1):230–251, 2019.

[Oppy, 2021] Graham Oppy. Ontological Arguments. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.

[van Inwagen and Sullivan, 2021] Peter van Inwagen and Meghan Sullivan. Metaphysics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.