Artificial Intelligence and Religion: Recent Advances and Future Directions

with Andrea Vestrucci, "Introduction: Five Steps Towards a Religion-AI Dialogue"; Lluís Oviedo, "AI and Theology: Looking for a Positive—But Not Uncritical—Reception"; Christoph Benzmüller, "Symbolic AI and Gödel's Ontological Argument"; Sara Lumbreras, "Lessons from the Quest for Artificial Consciousness: The Emergence Criterion, Insight-Oriented AI, and Imago Dei"; Marius Dorobantu, "Artificial Intelligence as a Testing Ground for Key Theological Questions"; and Andrea Vestrucci, "Artificial Intelligence and in God's Existence: Connecting Philosophy of Religion and Computation."

ARTIFICIAL INTELLIGENCE AND IN GOD'S EXISTENCE: CONNECTING PHILOSOPHY OF RELIGION AND COMPUTATION

by Andrea Vestrucci

Abstract. The exploration of metaphysical arguments in the symbolic AI environment provides clarification and raises unexpected questions about notions in philosophy of religion and theology. Recent attempts to apply automatic theorem prover technology to Anselm's ontological argument have led to a simplification of the argument. This computationally discovered simplification has given rise to logical observations. The article assesses one of these observations: the application of the diagonal method (in Cantor's version) to Anselm's argument. The evaluation of the applications of theorem provers to metaphysical and theological arguments contributes to the following topics in philosophy of religion: the limits of natural theology, the relationship between religion and STEM, and theology's scientificity.

Keywords: artificial intelligence and religion; artificial intelligence and theology; computational metaphysics; diagonal argument; formal theology; ontological argument; theology as science; religion and technology

Andrea Vestrucci, Faculty of Information Systems and Applied Computer Sciences, University of Bamberg, Germany. Starr King School, Oakland, US; e-mail: avestrucci@sksm.edu.

Three Approaches to AI–Religion Research

In recent years, the interdisciplinary investigation on the relationships and interactions between artificial intelligence (AI) and religion developed into at least three approaches.

The first approach has religion at its center: it applies outcomes from AI and computer science to formulate new philosophical, sociological, or theological insights on spiritual and religious topics and to analyze changes in practices (Sutinen and Cooper 2021). This approach might also assess the impact of technological outcomes, developments, and habits—such as digitalization and the diffusion of the internet—on spiritual notions and religious beliefs and rituals (Phillips 2019; Campbell and Tsuria 2021).

The second approach has computer science at its center: it analyzes and presents the impact that discoveries and advances in computer science and AI have on our culture, society, or world vision by using metaphors based on theological and philosophical concepts (Schmidhuber 2012; Marcus 2013). In this approach, religious ideas and theological concepts are the means to investigate how the relationship with technology and AI reframes or recalibrates our orientations in thinking and in the world (Obadia 2022).

The third approach seeks a full interdisciplinary connection between research in religion and in symbolic AI: it aims to have both poles as its center, by building a theoretical and experimental interconnection between metaphysics and computer science. This interconnection consists in exploring, in an automated reasoning environment, informal metaphysical arguments featuring concepts in philosophy of religion and theology. Different programs are applied to metaphysical arguments, after a process of formalization and translation of such arguments into the software syntax. This is the approach provided by the program in computational metaphysics, inaugurated by Fitelson and Zalta (2007) as application of the research on axiomatic theory of abstract objects. Recent contributions to this research program include works by Christoph Benzmüller and his teams in Berlin, Luxembourg, and Bamberg (Benzmüller and Woltzenlogel Paleo 2016; Kirschner, Benzmüller, and Zalta 2019; Benzmüller 2020). This third direction of research produces effective cross-contributions between the two disciplines: on one hand, AI programs are applied to higher abstract languages formalizing complex metaphysical/theological informal concepts and arguments; on the other hand, the feedback from the theorem prover often provides unexpected and useful insights on the metaphysical and theological concepts and arguments.

My article aims to provide a first introduction to this third approach, in order to invite philosophers and theologians to further familiarize with outputs from computational metaphysics and the new perspectives it can grant on their work material. At the same time, my article focuses on

assessing one outcome of the program in computational metaphysics: Oppenheimer and Zalta's computationally discovered simplification of Anselm's ontological argument in *Proslogion* II (Oppenheimer and Zalta 2011, 2021).

In what follows, I outline Oppenheimer and Zalta's formalization of Anselm's argument, their translation of such formalization in the language of theorem prover Prover9, and their analysis of the simplification that resulted by running the program (sections "Oppenheimer and Zalta's Formalization of Anselm's Argument" and "Oppenheimer and Zalta's Analysis of the Computational Simplification"). Section "Diagonalization of the Ontological Argument" presents an observation concerning the application of Cantor's diagonal method to the ontological argument. In section "Philosophical and Theological Relevance of Symbolic AI Applications," I explore why familiarization with the research in computational metaphysics is relevant in current debates in philosophy of religion.

Oppenheimer and Zalta's Formalization of Anselm's Argument

In order to understand what Oppenheimer and Zalta inferred from their computationally discovered simplification of Anselm's ontological argument, I need to first outline their formalization of the argument (Oppenheimer and Zalta 1991).

The logical background is Russell's theory of well-defined definite descriptions, that is, definite descriptions that have a unique denotation, such that there is one and only one object satisfying the description. In Oppenheimer and Zalta's language, a definite description is formalized as "$\imath x \varphi$," meaning "the unique x such that φ" (1991, 2–3). They use a free logic to formalize definite descriptions: this means that "definite descriptions can't be substituted into universal claims without first knowing that they are well-defined" (2011, 4), that is, without first knowing that the description has indeed a denotation. This is important in order to avoid forcing a description to have a denotation: that a description has a denotation must first be *proved*. Of course, as Oppenheimer and Zalta acknowledge (1991, 5), being well-defined or having a unique denotation does not mean that the object denoted by the description exists *in reality*. This specification is fundamental in the context of Anselm's ontological argument, since such "existence in reality" is precisely what must be proved.

The logical principles governing definite descriptions in Oppenheimer and Zalta's language are the following (1991, 6–7; 2011, 3):

Description Axiom: the thing that satisfies φ satisfies also ψ if and only if there is a y that satisfies φ, everything that satisfies φ is identical with y,

and y satisfies ψ. In other words, the x that satisfies φ satisfies also ψ if and only if there is a unique x that satisfies φ, and x satisfies ψ.

$$\psi\left[\imath x\varphi/z\right] \equiv \exists y\left(\varphi\left[y/x\right] \wedge \forall u\left(\varphi\left[u/x\right] \to u = y\right) \wedge \psi\left[y/z\right]\right).$$

$\psi[\imath x\varphi/z]$ results from substituting $\imath x\varphi$ for z everywhere in ψ, and ψ is an atomic or identity formula in which z occurs free. By abbreviating $\exists x(\varphi \wedge \forall y(\varphi[y/x] \to y = x))$ into $\exists!x\varphi$, from Description Axiom we have the following two logical principles:

<u>Description Theorem 1</u>: if there is a unique object satisfying φ, then φ is well-defined.

$$\exists!x\varphi \to \exists y\left(y = \imath x\varphi\right).$$

<u>Lemma 1</u>: if there is an object that is identical with the (unique) denotation of a well-defined description, then this object is the denotation of that description.

$$\tau = \imath x\varphi \to \varphi[\tau/x], \quad \text{for any } \tau.$$

From Lemma 1, it is deduced <u>Description Theorem 2</u>: if a description has a unique denotation, then any free variables in that description can be substituted with that denotation.

$$\exists y\left(y = \imath x\varphi\right) \to \varphi[\imath x\varphi/x].$$

These are the logical principles. The nonlogical principles are five (Oppenheimer and Zalta 1991, 8–14; 2011, 5–6):

<u>Connectedness of the binary Relation G "greater than"</u>: for any two objects x and y, either the former is greater than the latter, or vice versa, or the two are the same object.

$$\forall x\forall y\left(xGy \vee yGx \vee x = y\right).$$

For Oppenheimer and Zalta, Relation G is not required to be an *ordering* relation: it is just a *connected* relation (1991, 13). The difference between ordering relation and connected relation is that an ordering relation must be transitive ($xGy \wedge yGz \to xGz$) and antisymmetric (for each distinct x and y, $xGy \to \neg(yGx)$). Relation G only requires that all pairs of conceivable objects are in the relation, without ordering these objects from a least element to a greatest element. Thus, Relation G does not require transitivity: what matters is that Relation G connects any two conceivable objects and sees which one is greater than the other. This implies that, as per Relation G, *there are not two distinct conceivable objects that have the same "greatness" (whatever it means): if two conceivable objects have the same*

"greatness," then the two objects coincide, they are the same. Therefore, it is not clear why Relation G is not antisymmetric, since antisymmetry means $xGy \land yGx \to x = y$, and this is indeed part of the formula of Relation G. I will return to this in a moment.

Premise 1: there is an x which is conceivable and such that no other object y is greater than x and also conceivable.

$$\exists x \left(Cx \land \neg\exists y \left(yGx \land Cy \right) \right).$$

If we replace "$(Cx \land \neg\exists y(yGx \land Cy))$" with "$\varphi_1$" then Premise 1 reads: $\exists x\varphi_1$. "None greater" differs from "greatest": there is a fundamental difference between the property "element (of a set) than which there is no greater" and the property "greatest element (of a set)." Assuming a nonempty set T and a binary relation R, the element x of T satisfying the first property is called "Maximal element x" $=_{df} \forall s \in T : x \leq s \to x \geq s$. The element x of T that satisfies the second property is called "Greatest element x" $=_{df} \forall s \in T : xGs$. Let's ask: assuming a set C of all conceivable things, is $x\varphi_1$ the maximal or the greatest element of C? The definition of maximal element would fit the "negative" wording "none greater." However, a maximal element is not required to be comparable to all other elements of a set (i.e., the relation establishing the maximal element is not required to apply to all elements of the set): there can be elements with which the maximal element is *not* in relation, hence there can be more than one maximal element. This contradicts Relation G, because it is defined to apply to *all pairs of elements to the set*. This means that $x\varphi_1$ is the greatest element of set C.[1] But this would imply that Relation G *is an ordering one*—and, as mentioned, for Oppenheimer and Zalta Relation G is not required to be an ordering one. We have an impasse: either $x\varphi_1$ is the greatest element of set C and Relation G is an ordering one, or Relation G is not an ordering one and nothing assures us that there is an element than which *all* other elements of set C are less great.[2]

This impasse is apparently solved by the third nonlogical principle, Lemma 2: if there is an x, which satisfies φ_1 (i.e., being conceivable and such that nothing greater can be conceived), then this x is unique.

$$\exists x\varphi_1 \to \exists! x\varphi_1.$$

This follows from Relation G: as mentioned, if two elements have the same "greatness," then they are identical. So, if we have more than one maximal element of set C, these elements are the same. There is only one maximal element in set C. However, this implies that Relation G applies to *all* elements of C—which, again, is not required by the definition of maximal element.

There is a fundamental characteristic constituting Anselm's ontological argument: the distinction between existing *in intellectu* and existing *in re*. This distinction is formalized as distinction between existential quantifier $\exists x$ and existence predicate $E!x$. In Oppenheimer and Zalta's formalization, $\exists x$ is not existentially loaded: it simply means "there is an object x such that," without claiming that this x exists also in reality. This implies that there are objects that do not exist in reality: $\exists x(\neg E!x)$. The distinction between existential quantifier and existence predicate follows from Zalta's theory of abstract objects (Zalta 1983; see also http://mally.stanford.edu/theory.html), which can be encapsulated in the following principle: $A!x \leftrightarrow \neg \Diamond E!x$: an object x is abstract if it is not possible for x to exist in reality.

The distinction between existential quantifier and existence predicate is as the basis of the fourth nonlogical principle, <u>Premise 2</u>: if the object x with the property φ_1 (being conceivable and such that nothing greater can be conceived) fails to exist *in reality*, then there is another object, which is greater than object x and conceivable.[3]

$$E!\imath x\varphi_1 \rightarrow \exists y \left(yG\imath x\varphi_1 \wedge Cy \right).$$

The final nonlogical principle is the <u>Definition of God</u>: $g =_{df} \imath x\varphi_1$, where g is a constant with the same denotation of the description. From these logical and nonlogical premises, Oppenheimer and Zalta formalized their version of Anselm's ontological argument (1991: 14–15).

Oppenheimer and Zalta's Analysis of the Computational Simplification

As part of the program in computational metaphysics, Oppenheimer and Zalta aimed to test their formalization of Anselm's argument in an automated reasoning environment. In order to do so, they represented all logical and nonlogical premises into the syntax of the first-order theorem prover Prover9.

Figure 1 shows the representation of all logical and nonlogical premises—for details, see Oppenheimer and Zalta (2011, 7–12). The "Assumptions" panel features, from top to bottom: Description Theorem 1 (line 1); Lemma 1 (line 2); Description Theorem 2 (line 3); a Sorting Principle for definite description (line 4); Connectedness of Relation G (line 5); Definition of the property "none greater" (line 6); Premise 1 (line 7); Lemma 2 (line 8); Premise 2 (line 9); Definition of God (line 10). The "Goals" panel features the conclusion of the argument, that is, what is to be proved: that g exists in reality, or $E!g$.

Some notes on this representation. "Relation1" stands for a 1-place relation. "Ex1(Relation,x)" (resp. "Ex2(Relation,x,y)") means that an object x (resp. two objects x and y) exemplifies a 1-
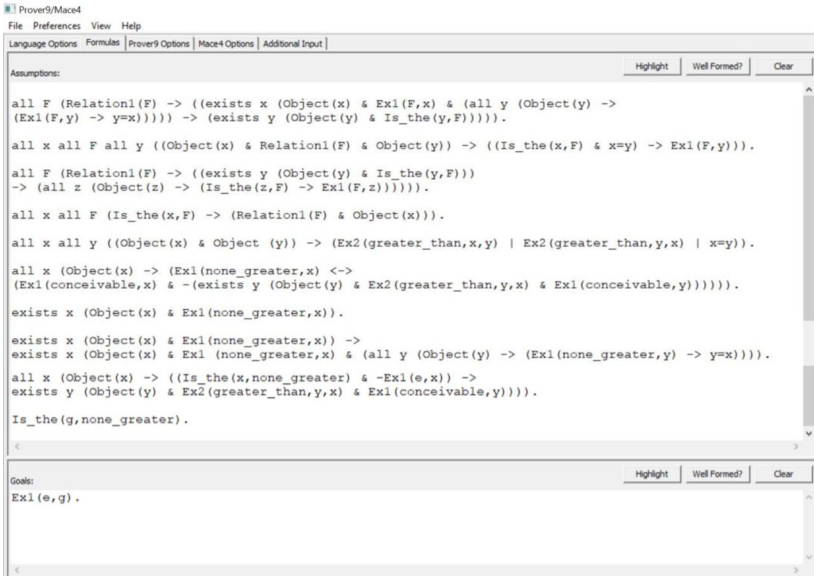
Figure 1. Ontological Proof on Prover9.

place (resp. 2-place) `Relation`; those are used to represent, respectively, the "none greater" property and the "greater than" relation. The formula "`Is_the`" represents definite descriptions. The Sorting Principle (line 4) formalizes that "`Is_the`" is satisfied by an `Object` for a `Relation`. The formulation of connectedness of Relation *G* (line 5) confirms that the Relation spans over *all* pairs of elements of set *C*. The Definition of "none greater" (line 6) serves to explain which is the "`Relation1`" to which the "`Ex1`" in Premise 1 refers (line 7). The property "`e`" in Premise 2 (line 9) stands for "existing in reality."

Prover9 works by clausification: premises are reduced to a disjunction of atomic formulas (or negation of atomic formulas), universal quantifiers are dropped, existential quantifiers undergo Skolemization (the existentially quantified variable is replaced by a function whose arguments are the universally quantified variables - if any - that precede the existential quantifier), the negation of the conclusion is also clausified, and this clausified negation of the conclusion is assumed as a premise. Thus, Prover9 works in a *Reductio* environment: the proof consists in deriving a contradiction from the resulting set of assumptions. (For a more detailed explanation, see https://mally.stanford.edu/cm/ontological-argument/old/clausifyingDescThm1.html.)

By running Prover9 and analyzing the output, Oppenheimer and Zalta discovered that Prover9 needed only Description Theorem 2, Sorting Principle, Definition of "none greater," Premise 2, and definition of God (and, clearly, the conclusion to be proved, i.e., that *g* exists *in re*) to build a proof.

Thus, Prover9 did *not* need Description Theorem 1, Premise 1, Lemma 2, and the connectedness of Relation G.

For Oppenheimer and Zalta this simplification rises serious logical questions. Some of their observations include (2011, 16–19): (1) Such simplification "reveal[s] that minimal logical and non-logical machinery is necessary for formulating an ontological argument for the existence of God," so that it might appear that "the computationally-simplified version of the argument reveals that it has a subtle logical beauty" (2011, 17); (2) Since Premise 1 and Lemma 2 are not necessary, then there is not (logical) need to justify the use of the definite description $\iota x \varphi_1$ because (a) the *Reductio*, by negating $E!\iota x \varphi_1$, does not assume that the description has denotation, and (b) the use of definite description is justified within the proof itself, that is, within the *Reductio* environment; (3) The simplification elicits a generalization of Premise 2 (see below); (4) The simplification seems to clarify that the ontological argument makes use of the diagonal argument (see next section); (5) Since the connectedness of Relation G is not necessary, then Relation G does not need to be connected either: "it simply has to satisfy Premise 2, i.e., be such that $E!\iota x \varphi_1 \rightarrow \exists y(yG\iota x \varphi_1 \wedge Cy)$" (2011, 18). Thus, the "burden" of the ontological proof would fall on the "shoulders" of Premise 2 alone.

The analysis of the computationally discovered simplification shows a flaw in the original formalization of the ontological argument. This flaw concerns the soundness of Premise 2: as mentioned, the simplified ontological argument does not need to justify the use of definite descriptions (since Premise 1 and Lemma 2 are expunged), and this justification can neither come from Premise 2 alone, because Premise 2 does not presuppose that there is a well-defined definite description for φ_1. Therefore, Premise 2 can be true (and thus the argument can work) only by using the conclusion of the ontological argument, that is, that the definite description denotes (is well-defined) and that such denotation exists *in re* (2011, 20). This is a vicious circle. The alternative is to find an independent support for the use definite description. But this means that, "though the simplified ontological argument is valid, Premise 2 is questionable and to the extent that it lacks independent justification, the simplified argument fails to demonstrate that God exists" (2011, 21).

As mentioned, in light of the computational simplification Oppenheimer and Zalta argued that Premise 2 could be generalized into the following Premise 2': $\neg E!x \rightarrow \exists y(yGx \wedge Cy)$ (2011, 17; see also 2021, 13): Premise 2' states that if something does not exist *in re*, then something else is greater than that and conceivable. In 2021, to undermine some criticisms in Garbacz (2012), Oppenheimer and Zalta further generalized Premise 2 into the following universal claim, called Premise 2'' (2021, 13):

$$\forall x \left((\varphi_1 \wedge \neg E!x \rightarrow \exists y \left(yGx \wedge Cy \right) \right).$$

Premise 2" states that for every object x, if x is such that none greater can be conceived and it does not exist *in re*, then, there is another conceivable object y which is greater than x. By substituting Premise 2" to Premise 2, the argument needs both Premise 1 and the connectedness of Relation G to reach the conclusion (2021, 13–14). This confirms that the computational simplification showed that there was a problem with Premise 2: the application of Prover9 was necessary to detect and address an issue concerning the relationship between Premise 2 and the claim the ontological argument aims to demonstrate, God's existence *in re*.

In what follows, I further discuss one of Oppenheimer and Zalta's observations on the computational simplification, and I present a theological interpretation of it.

Diagonalization of the Ontological Argument

I focus on Oppenheim and Zalta's observation that the computational simplification shows that Anselm's ontological argument makes use of the diagonal argument:

> [T]he new analysis of the argument brings out much more clearly that it deploys diagonal reasoning for a positive conclusion. [...] Anselm diagonalizes when he applies the description to itself in line (6), i.e., when he invokes Description Theorem 2 after concluding within the *Reductio* that there is something which is the conceivable thing such that nothing greater is conceivable. <u>Description Theorem 2 allows him to infer that the object denoted by the description satisfies the matrix of the description, i.e., that it is itself conceivable and such that nothing greater is conceivable. Since the description itself is instantiated within its own matrix, we have a clear case of diagonalization.</u> But here the diagonal argument leads to an existence claim, rather than to a nonexistence claim as in Russell's Paradox. More generally, diagonal arguments have been used to reach negative claims, such as in Cantor's proofs that the power set of a set A can't be mapped 1-to-1 to a subset of A and that there is no 1-to-1 mapping from the set of real numbers to the set of natural numbers. Diagonal arguments have also been employed to generate *aporiai*, or puzzles such as the Liar Paradox. (Oppenheimer and Zalta 2011, 18, my underlining; see also Oppenheimer and Zalta 2021, 11)

The underlined part seems to me to be the core of the quotation. Let's try to unpack it. First, I recall some terminology: "(definite) description" = "$\imath x \varphi_1$"; "matrix of the description" = "φ_1" = "$Cx \wedge \neg \exists y(yGx \wedge Cy)$"; Description Theorem 2 states that $\exists y(y = \imath x \varphi) \rightarrow \varphi[\imath x \varphi / x]$, that is, any free variable in a description matrix can always be substituted by the object denoted by the description.

Back the quotation. The "line (6)" mentioned in the quotation corresponds to the following line in the tabular form (2011, 13; see also line 4 at page 7): $C\imath x \varphi_1 \wedge \neg \exists y(yG\imath x \varphi_1 \wedge Cy)$. This line reads: the unique object satisfying the property being-conceivable-and-such-that-no-other-

conceivable-object-is-greater-than-it is itself conceivable and such that no other conceivable object is greater than it. To see how this derives from Description Theorem 2, let's proceed gradually. Line 5 of the tabular form asserts that there is a unique object that satisfies the matrix of the description: $\exists y(y = \imath x\varphi_1)$. Notice that this is the antecedent of Description Theorem 2. Therefore, by applying Description Theorem 2, we have $\varphi_1[\imath x\varphi_1/x]$, that is, we can substitute any free variable x occurring in the matrix of the description ($\varphi_1 = Cx \wedge \neg\exists y(yGx \wedge Cy)$) with the unique object $\imath x\varphi_1$ that satisfies the matrix of the description. I recall that "$\imath x\varphi_1$" abbreviates "$x(Cx \wedge \neg\exists y(yGx \wedge Cy))$"; thus, if we substitute this to any x in the description matrix we have:

$$C\imath x \left(Cx \wedge \neg\exists y \left(yGx \wedge Cy\right)\right) \wedge \neg\exists y \left(yG\imath x \left(Cx \wedge \neg\exists y \left(yGx \wedge Cy\right)\right) \wedge Cy\right).$$

This line is the expansion of the aforementioned line (6). It reads: the unique object that satisfies the description matrix (i.e., the property of being conceivable and such that no greater object is conceivable) also uniquely exemplifies (i.e., it is *the* instance of) exactly this description matrix (or property), that is, "it is itself conceivable and such that nothing greater is conceivable." Is this "a clear case of diagonalization"?

"Diagonalization" is the name given to a method that Georg Cantor used in 1891 to redemonstrate the two theorems mentioned in the quotation: that the cardinality of set $\mathbb{R}$ of real numbers is bigger than the cardinality of set $\mathbb{N}$ of natural numbers, and that the cardinality of the powerset of any set A is larger than the cardinality of set A. Cantor already proved the theorems in 1874, but he praised the method introduced in 1891 for its simplicity (Cantor 1891, 76).

I illustrate diagonalization starting from the first theorem. Cantor considers a set M of elements e, which are infinite sequences of two digits, m and w, so that each $M \ni e = (x_1, x_2 \ldots, x_v \ldots)$, where $x_1, x_2 \ldots, x_v \ldots$ are either m or w (Cantor 1891, 75–76). Let's take these elements e randomly and building a 1-to-1 correspondence with the elements of $\mathbb{N}$, that is, 1 2 3 4 5 6… Building this correspondence equates to *enumerate* the elements of M. Cantor organizes this enumeration in an array (a square matrix), where on the left we have the number of the elements e, and on the top we have the number of the digits composing each e_n: $a_{n,1}$, $a_{n,2}$, $a_{n,3}$, where n is the number of the element e to which the digits belong:

$$
\begin{aligned}
e_1 &= \left(a_{1,1}, a_{1,2}, \ldots, a_{1,\mu}, \ldots, a_{1,v}, \ldots\right) \\
e_2 &= \left(a_{2,1}, a_{2,2}, \ldots, a_{2,\mu}, \ldots, a_{2,v}, \ldots\right) \\
e_3 &= \left(a_{3,1}, a_{3,2}, \ldots, a_{3,\mu}, \ldots, a_{3,v}, \ldots\right) \\
&\ldots \\
e_\mu &= \left(a_{\mu,1}, a_{\mu,2}, \ldots, a_{\mu,\mu}, \ldots, a_{\mu,v}, \ldots\right) \\
&\ldots
\end{aligned}
$$

Cantor builds a new element of M, e_D, by changing all digits located on the diagonal $a_{v,v}$ of the array, that is, all digits that occupy the position having the same number (n) of the element e_n the digit belong to: we change the first digit of e_1, the second digit of e_2, the third digit of e_3, ..., the μth digit of e_μ... We change these diagonal digits by substituting m with w in case $a_{v,v} = m$, or w with m in case $a_{v,v} = w$. Question: what is the position of this element e_D in the list enumerating the elements of M? Perhaps e_D is in position number p? No, because the pth digit of e_D is built on the modification of the digit $a_{p,p}$ of the element e in position p (i.e, e_p). Since e_D is built on the modification of a digit of *each* element of M, then e_D presupposes the whole array of 1-to-1 correspondences between the elements of M and N. Therefore, e_D is *nowhere* in the array. There is no number $n \in \mathbb{N}$ which corresponds to e_D. If we consider the digit $a_{i,j}$ of each e (including e_D) as the decimal digits of irrational numbers in the interval (0,1), then each e (including e_D) is an element of $\mathbb{R}$, but (at least) e_D is *not* in a one-to-one correspondence with an element of $\mathbb{N}$. Thus, $\mathbb{R}$ has a cardinality which is "bigger" than the cardinality of $\mathbb{N}$: its infinity is nondenumerable. QED

The second theorem is a generalization of the first. Let's take a random set A of infinite cardinality, and its powerset $P(A)$—the powerset of a set A is the set of all subsets of A. Let's consider a function $f(x) : A \rightarrow P(A)$ from elements of A to elements of $P(A)$, and let's assume this function is surjective, that is, such that for all elements p of $P(A)$ there is an element x of A so that $p = f(x)$. In other words, we assume that the function covers all elements of the codomain $P(A)$. Let's consider now D, a subset of A, constituted by all those elements of A that are not in the range of $f(x)$: the function cannot map these elements to themselves in $P(A)$. Subset D is defined as such:

$$D = \{x \in A : x \notin f(x)\}.$$

D is a subset of A thus it must feature in $P(A)$. Since per hypothesis the function mapping A to $P(A)$ is surjective, then also D, being a subset of A, must be mapped by the function, that is, there must be elements $d \in A : f(d) = D$. But D is constituted by elements that do not belong to their image of the function. Therefore, we have that $d \in f(d) = D \Leftrightarrow d \notin f(d)$. This is a contradiction. Thus, the function that maps elements of A to elements to its powerset is not surjective. The cardinality of $P(A)$ is greater than the cardinality of A. QED

Subset D is the *diagonal* set for the function $f(x)$, because it is built analogously to the element e_D in the previous proof. In fact, in case the cardinality of A is infinite, the hypothetically surjective function $f(x) : A \rightarrow P(A)$ defines an infinite array with, on the left, all members x of A, and on the top the same members as argument of the function $f(x)$.

The array is constituted by values 1 or 0 depending on whether or not an x is member of $P(A)$, that is, whether or not $x \in f(x)$. Let's consider all diagonal values, and let's change them from 1 to 0 and vice versa. We create a new row "D" (a new subset of A) constituted by all members of A to which $f(x)$ does *not* apply: if $(a, f(a)) = 0$, then a is in D and $D \neq f(a)$; if $(a, f(a)) = 1$ then a is not in D and, again, $D \neq f(a)$. Thus, $D = \{x \in A : x \notin f(x)\}$. Row D is nowhere in the array because it presupposes all values of the array. However, it should feature in it, since it is a subset of A. Thus, we have a contradiction.

Both "diagonal objects" (it would be more appropriate to call them "antidiagonal"), e_D and D, share the same structure: (1) They are built on the modification of the diagonal values on an array which lists, on the left, and with no specific order, the elements x of a set, and on the top the elements x (in the same order), so that the values of the array result from a binary relation between left and top elements (for the first theorem, it is the correspondence between enumeration of the elements of M and enumeration of the elements' digits; for the second theorem the binary relation is a function)[4]; (2) Point 1 (the modification of the diagonal values) is justified by the fact that the definition of the diagonal objects includes the *negation* of the binary relation that structures the array (in the first theorem the digits of e_D are per construction in *no* coordinates (x,x); in the second theorem the members of D are *not* members of the function of which they are the arguments)[5]; (3) Thus, the diagonal objects simultaneously belong and do not belong to the set of elements x, because the diagonal objects presuppose all values of the array.

In light of this, it seems that the instantiation of the description $\iota x \varphi_1$ within its own matrix is not a case of diagonalization for at least two reasons. First, because for Oppenheimer and Zalta the belonging of $\iota x \varphi_1$ to set C of conceivable things is not problematic (I will shortly return to this). Second, because $\iota x \varphi_1$ is not an element that is built on the basis of considering all elements of set C. On the contrary, $\iota x \varphi_1$ is an element of set C that enters on stage since the beginning of the ontological proof. It is perhaps the *only* element of C about which we know something: that it satisfies φ_1. Thus, I would say that the instantiation of the description within its own matrix is a case of self-reference, but not all cases of self-references are cases of diagonalization (Buldt 2016).

However, it is indeed possible to apply the diagonal method (in Cantor's version) to Anselm's ontological argument (in Oppenheimer and Zalta's version). We consider the following array: on the left, the elements c of set C in random order; on the top, the same elements in the same random order; and the values of each coordinate (c_i, c_j) corresponds to whether the binary relation G "greater than" is satisfied or not, with 1 if yes and 0 if not (0 also applies to the case $c_n G c_n$, i.e., $c_n = c_n$).

Let's focus on $\iota x \varphi_1$. First, per construction of φ_1, the row corresponding to $\iota x \varphi_1$ must have all 1 apart from the value of coordinates $(\iota x \varphi_1, \iota x \varphi_1)$ which is, of course, 0. Where is this row in the array? This row must be unique: no other rows can have all 1. One way to make sure the row is unique is to order the rows from the least to the greatest, but this way is blocked by the "greater than" relation not being an ordering one. In other words, since the "greater than" relation is not transitive, we cannot deduce from the array a row with all 1 (from aGb and bGc nothing assures that aGc); for the same reason, as mentioned previously, we cannot deduce from the array that this row is unique. Hence, the only way to make sure that the row with all 1 is there and is unique is to build such row on the array itself, i.e., on values of all other rows. One possibility is to select the values on the diagonal (they are all zeros being all cases of $c_n G c_n$) and to build the values of the row $\iota x \varphi_1$ by changing all 0 on the diagonal into 1. The choice of the diagonal is the most linear one to build a new row on the other rows' values. Moreover, the only entries we know for sure the values are the one located on the diagonal. Finally, even if there is an element m whose row has all 1 (e.g., an hypothetical "second-to-$\iota x \varphi_1$" element that also exists *in re*), the row of this element m certainly has 0 on position m (corresponding to the diagonal entry (m,m)), thus, $\iota x \varphi_1$ is greater than element m because its row has 1 on position m; it follows that the array cannot display the entry whose coordinates are $(\iota x \varphi_1, \iota x \varphi_1)$. Hence, Anselm's ontological proof can be indeed thought as diagonal argument that preserves both the non-ordering nature of G and the uniqueness of $\iota x \varphi_1$. I leave to a future article the application of the fixed-point lemma (Gödel's "variant" of the diagonal argument; see Gaifman 2006) to Anselm's ontological argument.

There is another (less rigorous) way to build $\iota x \varphi_1$ as a diagonal object: this has to do with the vagueness of the "none greater" property φ_1—in specific, to the extent of this property. Set C is the set of all conceivable objects. Thus, set C also includes the aforementioned array that lists all conceivable objects. In this case the array is a self-referential object, since the array itself must feature somehow within the array. Now, considering the property φ_1, the self-referential array seems *not* to be greater than the object that satisfies the property φ_1. In fact, it seems plausible that the "none greater" property also applies to any metalinguistic treatment of the object satisfying the property (in other words, "$\iota x \varphi_1$" must be somehow lesser that $\iota x \varphi_1$). Thus, $\iota x \varphi_1$ cannot be an object of the array, and yet the array shall list all conceivable objects, and here we have another case of diagonalization.

The philosophical and theological consequences are significant. The fact that the object uniquely satisfying the property φ_1 is a diagonal object means that the conceivability of God is built on all elements of set C—that is, the conceivability of God presupposes all conceivable objects.

Therefore, "God" (the conceivable object called "God") simultaneously belongs and does not belong to the set of conceivable objects. In other words, God can be placed within the set of conceivable objects only if it is *not* placed within the set of conceivable objects. God is *a conceivable "something" that lies beyond conception*.

Interestingly enough, this result echoes very closely what Anselm himself writes in chapter 15 of *Proslogion*: "Ergo domine, non solum es quo maius cogitari nequit sed es *quiddam maius quam cogitari possit*." In my translation: "Therefore, o Lord, you are not only that than which none greater can be thought, but you are *something greater than what can be thought*." God is not just that than which nothing greater can be conceived: God is that which is greater than everything that can be conceived. This seems to shake the very hypothesis of God being a concept, an object of thought, a word—the noun "God." As mentioned, this is implicitly present in the vagueness of the extent of the "none greater" property: does the extent of such property also include the metalinguistic level? If yes, then how can a discourse on the object satisfying the "none greater" property be possibly "not greater" than this object? And how is a conceivable object $c = (\imath x\varphi_1 + y)$ less "great" than $\imath x\varphi_1$? These questions would invite to deepen the idea that the notion of God impacts the very constitution of language, and our (metalinguistic) notions about language. This would mean that the notion of God might impact the distinction between object language and metalanguage, as I explore in Vestrucci (2019b).

To sum up: either $\imath x\varphi_1$ is not a diagonal object and is indeed a member of set C (without considering some issues arising from the peculiarity of set C and the vagueness of property φ_1), or $\imath x\varphi_1$ is a diagonal object and thus its belonging to set C is at least problematic. But we should not consider this as a negative result, as much as we should not consider the results of Cantor's applications of the diagonal method as negative. On the contrary, they are all *positive* results: they clarify the nature and properties of sets and their members. In Cantor's case, the application of the diagonal method leads to the construction of specific mathematical objects (Gaifman 2006, 711–12) and to the formulation of theorems (Simmons 1990, 283–84). In the case of Anselm's proof, the application of the diagonal method deepens the nature of set C and opens to further questions on the property "none greater" and the object satisfying it.

Philosophical and Theological Relevance of Symbolic AI Applications

The application of symbolic AI programs to metaphysical and theological arguments and concepts turns out to be relevant for at least three areas of current investigation in philosophy of religion. These areas are: the discussion of criticisms to natural theology, the relationship between theology and mathematics, and the interdisciplinary research in religion and AI.

Following the division between external and internal criticisms to natural theology (Kojonen 2017), the use of automated reasoning assistants contributes to the debates on the incoherency of the concept of God, the epistemology of natural theology (as two of the "external critiques"), and the notion of evidentialism (as one of the "internal critiques"). Concerning the incoherency of the concepts of God, the assessment of the outputs from the theorem provers helps to deepen and recalibrate the limits, extents, and provability of our conceptualizations of God. More specifically, it clarifies whether, and how, our concepts of God are compatible with our premises and presuppositions about a divine being, or clash against our expectations—or implicit presuppositions—about the adherence of our concepts with the ("divine") object they aim to signify.

As far as the epistemology of natural theology is concerned, the exploration of metaphysical and theological arguments in an automated reasoning environment provides useful and unexpected insights on the logical structure and justifiability of the arguments. For instance, the discussion about the applicability of the diagonal method to the ontological argument might present new insights on the discussion on the equilibrium between knowability and unknowability of God. If, per diagonal method, the concept of God simultaneously belongs and does not belong to the set of conceivable things, then God is simultaneously conceivable and beyond conceivability. This fosters alternative approaches to the question about the extent, limit, and perfectibility of our knowledge about God. This might even create fruitful bridges with the theology of the *deus absconditus*, the specific approach in revelatory theology according to which divine revelation is simultaneously revelation *of* God's mystery and revelation that God *is* a mystery (see Vestrucci 2019a, 100–10).

Concerning the critique of evidentialism, the feedback that theorem provers provide about the logical structure of an informal argument might be used as external evidence or disproof of such argument. One of the greatest advantages in the use of theorem provers is the impossibility to cheat: the reasoning assistant shows immediately, and irrevocably, if an argument does not work logically. If in the automated reasoning environment the conclusion cannot be deduced by the premises, then it simply cannot be deduced: there is no possibility to trick the argument to make it work, for example by using vaguely defined concepts, or by formulating argumentations affected by fallacies (e.g., *petitiones principii*).

It is important to underline that, in spite of the name "theorem provers," the aim of using such software in a philosophical or theological setting is not simply to *prove* once and for all the informal arguments. Rather, the aim is to *improve* them, to better understand their complexity, to discover new aspects previously unnoticed, and to detect redundancies and over-complications. This improvement is part of the ongoing investigation on the logical depth, fruitfulness, and potentialities of metaphysical and theological reflections.

I pass to the second area of relevance: the philosophical debate on the relationship between theology and mathematics. The assessment of theorem provers applications to metaphysical and theological arguments contributes to the research on the consonance between the objects of theology and the objects of mathematics, both in the research practice and in the methodology of such research.

Concerning the research practice, working in an automated reasoning environment forces metaphysics and theology to speak "mathematically" by translating an informal argument into algorithmic syntax. This operation presupposes at least the following practices: to consider each aspect of the argument distinctively, to outline the axioms and principles of the argument, and to detect and clarify the semantic vagueness sometimes affecting theological and metaphysical concepts. For instance, the issues with the simplified version of the ontological argument might invite to further clarify the "none greater" property. Moreover, the distinction between quantifier and existence predicate helps in clarifying the definition of abstract objects, which can amend a possible aporia in the matter (Harrison 2017, 483).

Concerning the methodological aspect, the application of theorem provers to metaphysical and theological arguments forces to work in an axiomatic environment and to evaluate how efficiently—or not—metaphysics and theology work in such environment. This contributes significantly to the current revival of the classical issue of the scientificity of theology—that is, whether and how theology can be considered a science (Pannenberg 1973; Peterson 2008; Melville 2010; Munchin 2011; Göcke 2018). In the famous debate with theologian Karl Barth, mathematical logician (and former theologian) Heinrich Scholz argued that theology is not a science because its language cannot be axiomatized as in mathematics, and because its propositions cannot be tested by an objective third party—that is, by empirical experiments or in an axiomatic environment (Scholz 1971). Theology, according to Scholz, could claim the status of science only if its propositions are not only affirmed, but demonstrated. The axiomatic environment and the theorem prover outputs invite to consider Scholz's requirement not as a limit, but as a regulative idea, and as a challenge for the future of theology.

The third aspect of relevance is the interdisciplinary research on the relationship between religion and AI. The literature on this topic is growing fast. A large number of publications focuses on the possibility to formulate modeling for concepts and arguments belonging to human studies. One example concerns the attempts to approach the study of belief and replicate believing processes via computational models and AI programs (for a review on recent outputs, see Vestrucci, Lumbreras, and Oviedo 2021). Another example is transhumanism: the evaluation of how technology (including AI programs) interacts and interconnects with human body and mind contributes to rethinking what makes the human and what the

future of humankind would be, very often via a religious vocabulary (Mercer and Trothen, 2014; Smith 2018). A third example is the aforementioned investigation on the impact that technology has not only on our religious tenets and their plausibility and communicability, but also on our religious practices and rituals.

The applications of symbolic AI programs to metaphysical and theological arguments, and the assessment of such applications, contribute to debates in philosophy of religion, and to the interdisciplinary research on religion and AI, in ways that are unique, and different from the majority of other directions. This approach does not merely present analogies between concepts from computer science and from humanities, nor does it outline reflections based on the passive reception of technological outputs. Rather, it *applies* computational tools to develop the research in philosophy and theology; it carefully *translates* philosophical and theological concepts, inferences, and arguments into machine syntax; and it *assesses* the results of these applications *from within*, that is, from the actual effort of using and operating with AI programs. As such, this direction of research is a concrete model for building applied bridges and deep, fruitful dialogues not only between disciplines, but also between *competencies* in religion and AI. There are exciting, open ways to explore.

Conclusion

In this article, I have presented an example of close interconnection between philosophy of religion and AI research. This example focused on Oppenheimer and Zalta's analysis of their computationally-discovered simplification of Anselm's ontological argument via the application of an automated reasoning program. The article formulated an assessment of one of the observations on this computationally discovered simplification, concerning the use of the diagonal method in the ontological argument. The assessment showed that the implementation of such a method could make the concept of God simultaneously belonging and not belonging to the set of conceivable objects. This invites to reconsider and further investigate the thinkability of the concept "God." As such, the assessment of the AI exploration of metaphysical and theological arguments effectively contributes to the current philosophical debates on theology's language and epistemology, its relationship with mathematics, and the relationship between religion and AI. This close interplay between computation and philosophy of religion fosters an applied and concretely interdisciplinary research.

Acknowledgments

Notes

1. Moreover, the definition of greatest element seems to better fit the idea (proper to Anselm's proof) of "absoluteness" of the element satisfying the property. In fact, we could imagine the cardinality of set C to increase in time, with new conceivable things becoming its elements proportionally to the process of *learning*. I'd say that the ontological argument itself contributes to this increase of the cardinality of C by introducing the notions of "conceivable object" and of element satisfying the "none greater" property. From this "temporal" perspective, set C in time t_1 might have a new element, which was unconceivable at time t_0 and which is greater than the maximal element of C at t_0. For this reason, the definition of greatest element seems to better fit the idea of $x\varphi_1 : x\varphi_1$ is assured to be greater than any (past, present, and future) element of C. I plan to deepen this temporal approach to the ontological argument in a future article.

2. Oppenheimer and Zalta use the metaphor of arrows pointing away from x to y whenever x is greater than y. They write: "Now, in any model in which '∃$x\varphi_1$' is true, i.e., in which there is a conceivable object such that no conceivable object is greater, there has to be at least one object *having no arrows pointing towards it*! Such an object is called a maximal object'" (1991, 13). But this is not supported by the definition of maximal element: the "incomparable" objects, that is, the objects to which Relation G does not apply, also have also no arrow pointing towards them.

3. In other words, if the object with property φ_1 is only abstract, then another object can be conceived to be greater than this object (and the first object does not satisfy the property φ_1).

4. In the Grelling-Nelson Paradox (1908), another case of diagonalization, the left side of the array lists all possible properties, the top lists the English string representing these property (e.g., the word "written" for the property of being written), the values of the array refer to whether the binary relation "is denoted by" is satisfied or not (e.g., we have 1 for the coordinates [written, "written"], but 0 for the coordinates [happy, "happy"]).

5. In the Grelling-Nelson Paradox the diagonal object corresponds to the property of being heterological, that is, being a word *not* denoting its own meaning (negation of the binary relation): the row corresponding to this property is built on the modification of all values of the diagonal of the array.

References

Benzmüller, Christoph. 2020. "A (Simplified) Supreme Being Necessarily Exists, Says the Computer: Computationally Explored Variants of Gödel's Ontological Argument." In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning — Applications and Systems*, International Joint Conferences on Artificial Intelligence Organization (IJCAI), 779–89.

Benzmüller, Christoph, and Bruno Woltzenlogel Paleo. 2016. "The Inconsistency in Gödel's Ontological Argument: A Success Story for AI in Metaphysics." In Subbarao Kambhampati *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, Palo Alto (CA): AAAI Press, 936–42.

Buldt, Bernd. 2016. "On Fixed Points, Diagonalization, and Self-Reference." In *Von Rang und Namen. Philosophical Essays in Honour of Wolfgang Spohn*, edited by Wolfgang Freitag, Hans Rott, Holger Sturm, and Alexandra Zinke, 47–64. Leiden: Brill.

Campbell, Heidi A, and Ruth Tsuria, eds. 2021. *Digital Religion: Understanding Religious Practice in Digital Media*. London: Routledge.

Cantor, Georg. 1891. "Über eine elementare Frage der Mannigfaltigkeitslehre." *Jahresbericht der Deutschen Mathematiker-Vereinigung* 1: 75–78.

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

Gaifman, Haim. 2006. "Naming and Diagonalization, from Cantor to Godel to Kleene." *Logic Journal of the IGPL* 14 (5): 709–28.

Garbacz, Paweł. 2012. "PROVER9's Simplification Explained Away." *Australasian Journal of Philosophy* 90 (3): 585–92.

Göcke, Benedikt Paul, ed. 2018. *Die Wissenschaftlichkeit der Theologie Band 1 Historisch und systematische Perspektiven*. Münster: Aschendorff.

Grelling, Kurt, and Leonard Nelson. 1908. "Bemerkungen zu den Paradoxien von Russell und Burali-Forti." *Abhandlungen der Friesischen Schule II*, 301–34. Göttingen: Vandenhoeck & Ruprecht.

Harrison, Victoria S. 2017. "Mathematical Objects and the Object of Theology." *Religious Studies* 53 (4): 479–96.

Kirschner, Daniel, Christoph Benzmüller, and Edward N. Zalta. 2019. "Computer Science and Metaphysics: A Cross-Fertilization." *Open Philosophy* 2:230–51.

Kojonen, Erkki Roppe. 2017. "Natural Theology in Evolution: A Review of Critiques and Changes." *European Journal of Philosophy of Religion* 9 (2): 83–117.

Marcus, Gary. 2013. "Can Science Lead to Faith." *The New Yorker* https://www.newyorker.com/tech/annals-of-technology/can-science-lead-to-faith. Last accessed on 9th august 2022. Last access 9th of august 2022

Melville, Stewart, ed. 2010. Science and Religion in Dialogue*, 2 volumes*. Oxford: Wiley-Blackwell.

Mercer, Calvin, and Tracy J. Trothen, eds. 2014. *Religion and Transhumanism: The Unknown Future of Human Enhancement*. Santa Barbara, CA: Praeger.

Munchin, David. 2011. *Is Theology a Science? The Nature of the Scientific Enterprise in the Scientific Theology of Thomas Forsyth Torrance and the Anarchic Epistemology of Paul Feyerabend*. Leiden: Brill.

Obadia, Lionel. 2022. "(Online) Spelling the (Digital) Spell: Talking About Magic in the Digital Revolution." *Sophia* 61 (1): 23–40.

Oppenheimer, Paul E., and Edward N Zalta. 2021. "On Anselm's Ontological Argument in ProslogionII." *History of Philosophy & Logical Analysis* 1:1–25.

Oppenheimer, Paul E., and Edward N. Zalta 1991. "On the Logic of the Ontological Argument." https://mally.stanford.edu/ontological.pdf. Originally published in James Tomberlin, ed. *Philosophical Perspectives 5: The Philosophy of Religion*, 509–29. (Atascadero: Ridgeview). Last access 9th of August 2022.

———. 2011. "A Computationally Discovered Simplification of the Ontological Argument." https://mally.stanford.edu/abstracts/ontological-computational.html. Last accessed on 9th august 2022. Originally published in *Australasian Journal of Philosophy* 89 (2): 333–49. Last access 9th of August 2022.

Pannenberg, Wolfgang. 1973. *Wissenschaftstheorie und Theologie*. Frankfurt am Main: Suhrkamp.

Peterson, Gregory R. 2008. "In Praise of Folly? Theology and the University." *Zygon: Journal of Religion and Science* 43 (3): 563–77.

Phillips, Peter M. 2019. *The Bible, Social Media and Digital Culture*. London: Routledge.

Schmidhuber 2012. This is actually a TedX talk: https://www.youtube.com/watch?v=T1Ogwa76yQo. A transcript of the talk is available here: https://www.kurzweilai.net/in-the-beginning-was-the-code. Last accessed on 9th august 2022.

Scholz, Heinrich. 1971. "Wie ist eine evangelische Theologie als Wissenschaft möglich?" In *Theologie als Wissenschaft: Aufsätze und Thesen*, edited by G. Sauter, 221–64. München: Keiser. (The article originally appeared in *Zwischen den Zeiten* 9 (1931)).

Simmons, Keith. 1990. "The Diagonal Argument and the Liar." *Journal of Philosophical Logic* 19 (3): 277–303.

Smith, Wesley J. 2018. "Transhumanism: A Religion for Postmodern Times." *Religion and Liberty* 28 (4). https://www.acton.org/religion-liberty/volume-28-number-4/transhumanism-religion-postmodern-times

Sutinen, Erkki, and Anthony-Paul Cooper. 2021. *Digital Theology: A Computer Science Perspective*. Bingley (UK) London: Emerald Group Publishing.

Vestrucci, Andrea, Sara Lumbreras, and Lluis Oviedo. 2021. "Can Artificial Intelligence Help Us to Understand Belief: Sources, Advances, Limits, and Future Directions." *International Journal of Interactive Multimedia and Artificial Intelligence* 7 (1): 24–33.

Vestrucci, Andrea. 2019a. *Theology as Freedom: On Martin Luther's De servo arbitrio*. Tubingen: Mohr Siebeck.

———. 2019b. "Metalanguage and Revelation: Rethinking Theology's Language and Relevance." *Logica Universalis* 13:551–75.

Zalta, Edward N. 1983. *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Dordrecht: Riedel.