Spinozian Model Theory

Justin Bledin and Yitzhak Melamed

Johns Hopkins University

Abstract

This paper is an excerpt from a larger project that aims to open a new pathway into Spinoza's *Ethics* by formally reconstructing an initial fragment of this text. The semantic backbone of the project is a custom-made Spinozian model theory that lays out some of the formal prerequisites for more fine-grained investigations into Spinoza's fundamental ontology and modal metaphysics. We implement Spinoza's theory of attributes using many-sorted models with a rich system of identity that allows us to clarify the puzzling status of such logical principles as the Substitution of Identicals and Transitivity of Identity in Spinoza's thought. The intensional structure of our Spinozian models also captures his proposal that states of affairs can be necessitated or excluded by the essences of particular things, an *essence-relative* modality that should be of interest to philosophers who have sought to rehabilitate the concept of essence in contemporary analytic metaphysics.

Keywords: Spinoza, formal history of philosophy, modal logic, many-sorted model theory, logic of essence

1 Introduction

Spinoza's magnum opus *Ethica More Geometrico Demonstrata* (*Ethics* for short) is not an easy or clear work, to put it mildly. [1] To make matters worse, many commentators have complained of shoddy construction, arguing that Spinoza's logical argumentation breaks down often and early in the text (see for instance [11,10,1]). In this paper, we present part of a broader effort to reevaluate this pessimistic story of "Spinoza, the Logician" by reconstructing an initial fragment of the *Ethics* through E1p15 within the modern framework of quantified modal logic [8].

In his opening definitions and axioms, Spinoza introduces the building blocks of his ontology—substance [*substantia*], attribute [*attributum*], mode

[1] All quotations from Spinoza's works and letters are from Curley's translation [17]. We rely on Gebhardt's critical edition [16] for the Latin text. Passages in the *Ethics* are referred to using the following standard abbreviations: 'a' for axiom, 'c' for corollary, 'e' for explanation, 'p' for proposition, 's' for scholium, and 'd' for a definition when it appears immediately to the right of the part of the book or a demonstration in all other cases (so, for example, E1d1 is the first definition of Part One of the *Ethics*, and E1p15d is the demonstration of the fifteenth proposition of Part One). We use the abbreviation 'KV' for *Short Treatise on God, Man, and His Well-Being* and 'Ep.' for *Letters*.

[*modus*], and God [*Deus*]—and presents several constraints on the ontological, conceptual, and causal relations that obtain between these protagonists. In the propositions themselves, he establishes core properties of substance, such as that it is self-caused (E1p7) and infinite in its own kind (E1p8). By the time he reaches E1p15, Spinoza has already established his *substance monism*: God, a substance with an infinity of attributes (E1d6), exists (E1p11), and is unique (E1p14), and all inheres in God (E1p15). To the extent that Spinoza's demonstrations fail in this crucial early stage of the *Ethics*, this threatens the metaphysical foundations of his entire project.

The semantic backbone of our reconstruction is a custom-made Spinozian model theory, which we develop in the present installment of our work. This theory lays out some of the formal prerequisites for more fine-grained formal investigations into Spinoza's fundamental ontology and modal metaphysics. We implement Spinoza's theory of attributes using many-sorted models with a rich system of identity—our models include no less than three distinct notions of numerical identity—that allows us to clarify the puzzling status of such logical principles as the Substitution of Identicals [2,3] and Transitivity of Identity [9] in Spinoza's thought. The intensional structure of our Spinozian models also captures his proposal that states of affairs can be necessitated or excluded by the essences of particular things, an *essence-relative* modality that should be of interest to philosophers who have sought to rehabilitate the concept of essence in contemporary analytic metaphysics [5,6,7].

Given Spinoza's metaphysical views, bringing in the modern apparatus of possible worlds—and, indeed, allowing for domains consisting of multiple entities—might seem like overkill, or even plain distortion. In addition to his substance monism, we take Spinoza to establish a *necessitarianism* later in Part One according to which every actual state of affairs is necessary—things could not have been otherwise. However, it is important to keep in mind that Spinoza has to *argue* for these doctrines, and some of his main conclusions are drawn only after a lot of careful preliminary work (see E1p33 and its scholia). So, we don't want to build too much of Spinoza's metaphysics directly into our models, which must be capable of representing not only the positions that Spinoza eventually arrives at in the *Ethics* but also alternative metaphysical possibilities that he rules out through his argumentation, such as universes with multiple co-existing substances and non-necessary facts.

That said, to capture some of Spinoza's own idiosyncratic views about the universe, our Spinozian models have a few non-standard twists. We motivate these in sections 2 and 3, where we implement Spinoza's theory of attributes and essence-relative notions of possibility and conceivability—and provide more overview of Spinoza's philosophy in the process. We then present our full Spinozian model theory in §4 and conclude in §5.

2 Modeling the Attributes

At the heart of Spinoza's ontology is the distinction between substance and mode. The hallmark of substance, according to Spinoza, is its *independence*.

In his definition of substance (E1d3), he tells us that substance is "in itself [*in se*]" (i.e., inheres in itself) and "conceived through itself [*per se concipitur*]". Later, in E1p7d, Spinoza proves that substance is independent in a third sense: substance is the cause of itself and is not caused by any other thing.

In contrast, modes are by their nature *dependent* beings. In his definition of mode (E1d5), Spinoza asserts that a mode is an "affection" (roughly, a quality) of substance, and then he spells out how modes are dependent in two senses in which substance is not: a mode is "in another [*in alio*] through which it is also conceived [*per alio concipitur*]"—that is, a mode inheres in and must be understood through something other than itself. In E1p16c1, Spinoza also establishes that modes are causally dependent in that they must be caused by another, namely God.

The other two protagonists of Spinoza's metaphysical system—attributes and God—raise some pressing interpretative puzzles. In E1d4, Spinoza defines an attribute as "what the intellect perceives of substance, as constituting its essence [*id, quod intellectus de substantia percipit, tanquam ejusdem essentiam constituens*]". Spinoza then defines God in E1d6 as "a being absolutely infinite [*ens absolutè infinitum*]", which is spelled out further as "a substance consisting of an infinity of attributes, of which each one expresses an eternal and infinite essence". Unlike Descartes [4], who rules out the possibility of one substance having more than one principal attribute, Spinoza allows substances—or rather the one divine substance, God—to have multiple attributes. Of God's many attributes, we humans have access to only two: Thought and Extension (see E2a5, E2p13, and Ep. 64). But God has infinitely more attributes beyond our epistemic and causal reach (E1d6 and Ep. 56 (IV/261/14)).

A common, though oversimplified, taxonomy divides interpretations of Spinoza's attributes into two camps: subjectivist and objectivist. While the subjective position goes back to Hegel, the locus classicus is Wolfson [19] where Spinoza is taken to claim that attributes are inventions of the finite perceiving mind. Because Wolfson's reading has been subjected to devastating (and to our mind justified) critique, we mention it only to set it aside. Our own working position is objectivist, in at least the sense that we do not take attributes to be inventions of the human mind. Following Garrett [9] (and echoing Melamed [12]), we regard Spinoza as a proponent of a "strong ontological pluralism" according to which one thing can have more than one "fundamental manner or kind of existence, reality, or being". The idea that there can be more than one kind of existence—this is the "ontological pluralism" part—might not strike at least some philosophers as especially peculiar given the distinction between concrete and abstract objects, particulars and universals, or the divine and mundane. On Garrett's interpretation, however, Spinoza takes this further in proposing that a *single* thing can have existence of different kinds—this is the "strong" part. Indeed, God is a substance having infinitely many fundamental kinds of existence, each of which might be regarded as one of God's attributes. Spinoza's "thinking substance" (E2p7s) is God existing as a thinking thing, "extended substance" (E1p15s, E2p7s) is God existing as extended, and like-

wise for the other unknown attributes shrouded in darkness—they too should be regarded as different kinds of existence (E1p20d: "each of [God's] attributes expresses existence"). The same goes for finite things: on Spinoza's ingenious solution to the Mind-Body Problem, your mind is *you* existing as a thinking thing while your body is *you* existing as an extended thing (E2p21s, E3p2s). [2]

To formally capture Spinoza's idea that one and the same thing can have existence of many different kinds, we adopt a *many-sorted* model theory [14,15,18]. Unlike single-sorted models for modal logic, which include a set of possible worlds $\mathcal{W}$ where each world $w \in \mathcal{W}$ is assigned a single domain $\mathcal{D}(w)$ consisting of the entities that exist in this world, many-sorted models assign a potentially infinite number of domains of quantification to each world. Where $S = \{s_1, s_2, ...\}$ is an index set of *sorts*, a many-sorted model $\mathcal{M}$ for S assigns each $w \in \mathcal{W}$ a domain $\mathcal{D}_{s_1}(w)$ of existents of sort s_1, a domain $\mathcal{D}_{s_2}(w)$ of existents of sort s_2, and so forth (we work with variable domain models). Given a sort $s \in S$, the variables x_s, y_s, ... range over things of this sort and $\forall x_s$, $\exists y_s$, ... quantify over $\mathcal{D}_s(w)$. While entities in different sortal domains of a world are generally regarded as distinct, we repurpose these models to allow for one and the same thing to exist in multiple sortal domains.

For purposes of modeling Spinoza's metaphysics, we work with an infinite set of sorts that includes the sort `Th` of thinking entities (or rather, entities existing as thinking), the sort `Ex` of extended entities (or entities existing as extended), and sorts corresponding to all the other attributes. We call these "secondary sorts" because our models also include a "primary sort" $\aleph$ whose domain $\mathcal{D}_\aleph(w)$ at a world $w \in \mathcal{W}$ consists of all the entities having any kind of existence in w conceived in their fullness as multifaceted beings: [3]

Spinozian sorts: $\mathcal{S}_{Spinoza} = \{\aleph, \texttt{Th}, \texttt{Ex}, ...\}$

The $\aleph$-sort affords a bird's-eye view of a pluralistic ontology. To theorize at this global layer is *not* to substract (or abstract away) the attributes from substance and its modes because on the interpretation we work with, to strip away all the attributes of a thing would be to deny it existence of *any* kind. On the contrary, entities are regarded from the "$\aleph$-perspective" as having *all* their attributes—for instance, Spinoza adopts this all-encompassing perspective in defining God as a being consisting of an infinity of attributes in E1d6.

To help keep track of the identity of entities across different sortal domains,

[2] While Garrett doesn't emphasize the role of the intellect in all this, we take E1d4 to impose a substantive condition on fundamental kinds of existence—these kinds correspond to how substance and its modes can be perceived by the *infinite* intellect (see E2p7s). Adopting one perspective, the infinite intellect perceives God as a thinking thing (E2p1) and its various modes as modes of thought. Adopting another perspective, the infinite intellect perceives God as an extended thing (E2p2) and its modes as modes of extension.

[3] For Spinoza, substance and its modes are *infinitely*-faceted, existing as thinking, extended, and so forth. However, our models can also represent alternative universes with entities existing in only finitely many sortal domains, and perhaps only one. While we often speak of the entities in $\mathcal{D}_\aleph(w)$ as "multifaceted", strictly speaking they needn't be.

we assume that Spinozian models come equipped with a family of (partial) *projection functions* $\pi_{\mathtt{Th}}$, $\pi_{\mathtt{Ex}}$, ... that at each world $w \in \mathcal{W}$ "project" the multifaceted beings dwelling in the primary domain $\mathcal{D}_{\aleph}(w)$ of this world to single-faceted entities in its secondary sortal domains (to project a multifaceted thing is to home in on its having existence of this or that kind):

> **Projection into secondary domains**
> For any world $w \in \mathcal{W}$ and secondary sort $s \in \{\mathtt{Th}, \mathtt{Ex}, ...\}$, the projection function $\pi_s(w)$ maps entities from $\mathcal{D}_{\aleph}(w)$ into $\mathcal{D}_s(w)$. When defined, $\pi_s(w)(a_{\aleph})$ is $a_{\aleph}$ in w as perceived by the (infinite) intellect as a being of sort s.[4]

For instance, if '$\mathsf{God}_{\aleph}$' denotes Spinoza's absolutely infinite substance at w, then $\pi_{\mathtt{Th}}(w)(\mathsf{God}_{\aleph})$ is the thinking substance (i.e., $\mathsf{God}_{\mathtt{Th}}$), $\pi_{\mathtt{Ex}}(w)(\mathsf{God}_{\aleph})$ is the extended substance (i.e., $\mathsf{God}_{\mathtt{Ex}}$), and so forth. Because any single entity is singular under any kind of existence, and the primary domain $\mathcal{D}_{\aleph}(w)$ of a world w includes all the entities with any kind of existence in w, we require that each $\pi_s(w)$ is a one-one injective function, each member of a secondary domain $\mathcal{D}_s(w)$ is the $\pi_s(w)$-projection of some member of $\mathcal{D}_{\aleph}(w)$, and (in the other direction) every multifaceted being $a_{\aleph} \in \mathcal{D}_{\aleph}(w)$ is projected into at least one of the secondary domains of w—i.e., $\pi_s(w)(a_{\aleph})$ is defined for some $s \in \{\mathtt{Th}, \mathtt{Ex}, ...\}$.

Given the different sortal layers in Spinozian models, we can identify a number of (genuine) identity relations. First, there is the "standard" identity relation $=$ that any element in any domain of a model stands in with respect to itself and to no other element (i.e., $=$ is the diagonal relation):

> **Standard identity**
> $a_s = b_{s'}$ at w iff $a_s, b_{s'}$ are the same element in a model.

Second, there is what we call the "projective" identity of any multifaceted being in the $\aleph$-domain with each of its projections. For $s \in \{\mathtt{Th}, \mathtt{Ex}, ...\}$,

> **Projective identity**
> $a_{\aleph} =_{\mathrm{P}} b_s$ at w iff $\pi_s(w)(a_{\aleph}) = b_s$.

Third, there is the "cross-attribute" or "trans-attribute" identity of these projections. For $s, s' \in \{\mathtt{Th}, \mathtt{Ex}, ...\}$,

> **Cross-attribute identity**
> $a_s =_{\mathrm{C}} b_{s'}$ at w iff there is some $c_{\aleph}$ s.t. $c_{\aleph} =_{\mathrm{P}} a_s$ and $c_{\aleph} =_{\mathrm{P}} b_{s'}$ at w.

All three relations are genuine identity relations in the sense that if $a_s = b_{s'}$, $a_s =_{\mathrm{C}} b_{s'}$, or $a_s =_{\mathrm{P}} b_{s'}$, then a_s and $b_{s'}$ are *one and the same thing*, though a_s and $b_{s'}$ might still differ in terms of their kind or kinds of existence. In

[4] Spinoza rephrases his definition of attribute in E2p7s, referring to an attribute as "whatever can be perceived by an *infinite intellect* as constituting the essence of substance" (italics added). So, apparently, Spinoza has God's infinite intellect in mind in E1d4.

monistic ontologies involving only a single kind of existence, there is room for only a single notion of identity. However, in a pluralistic setting where one is referring to and quantifying over entities with more than one fundamental kind of existence, it is useful to have multiple notions of identity in play to capture more fine-grained notions of sameness and difference.

Note that the three identity relations differ with respect to their logical properties. Standard identity is an equivalence relation, as is cross-attribute identity on the secondary sortal domains where it applies. However, projective identity is neither reflexive nor symmetric, and is transitive only in a vacuous sense as we can never have $a_s =_{\mathrm{P}} b_{s'}$ and $b_{s'} =_{\mathrm{P}} c_{s''}$ for any $a_s, b_{s'}, c_{s'''}$. That said, we can get failures of transitivity if we consider *combinations* of our identity relations. As discussed above, we can have $\mathsf{God}_{\aleph} =_{\mathrm{P}} \mathsf{God}_{\mathtt{Th}}$ (E2p1: "God is a thinking thing") and $\mathsf{God}_{\aleph} =_{\mathrm{P}} \mathsf{God}_{\mathtt{Ex}}$ (E2p2: "God is an extended thing"), but while $\mathsf{God}_{\mathtt{Th}} =_{\mathrm{C}} \mathsf{God}_{\mathtt{Ex}}$, $\mathsf{God}_{\mathtt{Th}} \neq \mathsf{God}_{\mathtt{Ex}}$. This transitivity failure reveals how standard identity is a stricter notion than cross-attribute identity over the secondary domains where these notions both apply. Cross-attribute identity is the appropriate notion of identity when we are counting substances and modes in the ontology but are not concerned with the distinction between different kinds of existence. From this coarse-grained perspective, $\mathsf{God}_{\mathtt{Th}}$ and $\mathsf{God}_{\mathtt{Ex}}$ are numerically identical because we are talking about one and the same substance. On the other hand, standard identity is the appropriate notion when we are counting things as distinct when they have different kinds of existence. From this more fine-grained perspective, $\mathsf{God}_{\mathtt{Th}}$ and $\mathsf{God}_{\mathtt{Ex}}$ are numerically distinct because the former is God existing as thinking while the latter is God existing as extended. [5]

Our identity relations also differ in terms of their substitutional properties. All the predicates introduced in this paper are referentially transparent contexts with respect to the standard identity relation. For instance, where '$\mathsf{Extended}(t)$' and '$\mathsf{Affection}(t, t')$' formalize t *is extended* and t *is an affection of* t', the following conditionals hold:

If $\mathsf{Extended}(a_s)$ and $a_s = b_{s'}$, then $\mathsf{Extended}(b_{s'})$.
If $\mathsf{Affection}(a_s, b_{s'})$, $a_s = c_{s''}$, and $b_{s'} = d_{s'''}$, then $\mathsf{Affection}(c_{s''}, d_{s'''})$.

In models that accurately capture Spinoza's philosophy, many predicates will also turn out to be referentially transparent with respect to projective identity in the restricted sense that if they hold with respect to some elements in the primary domain $\mathcal{D}_{\aleph}(w)$, and these elements are all projected into the same secondary sortal domain $\mathcal{D}_s(w)$, then the predicates hold with respect to these projections as well. For instance, where '$\mathsf{Substance}(t)$' and '$t \rightsquigarrow t'$' formalize t *is a substance* and t *causes* t', the following conditionals hold:

If $\mathsf{Substance}(a_{\aleph})$ and $a_{\aleph} =_{\mathrm{P}} b_s$, then $\mathsf{Substance}(b_s)$.
If $a_{\aleph} \rightsquigarrow b_{\aleph}$, $a_{\aleph} =_{\mathrm{P}} c_s$, and $b_{\aleph} =_{\mathrm{P}} d_s$, then $c_s \rightsquigarrow d_s$.

[5] See also [9] for discussion of the status of the transitivity of identity in Spinoza's philosophy.

On the other hand, many predicates will be referentially opaque contexts for projective identity in its full generality—for example, it can be the case that $\mathsf{Affection}(a_{\aleph}, b_{\aleph})$, $a_{\aleph} =_{\mathrm{P}} c_{\mathtt{Ex}}$, $b_{\aleph} =_{\mathrm{P}} d_{\mathtt{Th}}$, but $\neg\mathsf{Affection}(c_{\mathtt{Ex}}, d_{\mathtt{Th}})$. As Della Rocca observes in [2,3], "attribute contexts" like '$\mathsf{Extended}(t)$' and related attribute-sensitive predicates are also referentially opaque with respect to cross-attribute identity (though Della Rocca does not phrase his observation in these terms)—for example, we can have $\mathsf{Extended}(a_{\mathtt{Ex}})$, $a_{\mathtt{Ex}} =_{\mathrm{C}} b_{\mathtt{Th}}$, but $\neg\mathsf{Extended}(b_{\mathtt{Th}})$.

3 Modeling Possibility and Conceivability

While Spinoza's modal metaphysics remains the subject of considerable debate (see [13] for helpful discussion and references), we interpret him as a strict necessitarian. This commitment is *strongly* suggested in various places in the *Ethics*, such as in E1p29, where Spinoza proves that "in nature there is nothing contingent", and in E1p33, where he proves that "Things could have been produced by God in no other way, and in no other order than they have been produced." Even though Spinoza is a necessitarian, there are still good reasons to have multiple worlds available in our models.

First, as mentioned, Spinoza argues for his necessitarianism only in the second half of Part One—the argumentation only really gets going in E1p16, where this paper leaves off—so we don't want to presuppose this doctrine in our model theory, which should be capable of representing rival views. While any model that accurately incorporates Spinoza's modal commitments will be one in which the actual world is the sole metaphysical possibility, we allow for models that include more than one metaphysically possible world in order to represent alternative views that Spinoza rejects.

Furthermore, even in models encoding Spinoza's necessitarianism wherein actuality and metaphysical possibility coincide, there are benefits to having *metaphysically impossible* worlds around. With such worlds, we can capture Spinoza's rich modal metaphysics and more nuanced necessitarianism, which asserts not simply that things could not have been otherwise but that things could not have been otherwise *by virtue of God's essence*—the full natural order flows from the necessity of God's essence (E1p16). In the first scholium immediately following E1p33d, Spinoza goes on to distinguish between two different *sources* or *grounds* of the necessary existence/nonexistence of things: [6]

> A thing is called necessary either by reason of its essence or by reason of its cause. For a thing's existence follows necessarily either from its essence and definition or from a given efficient cause. And a thing is also called impossible from these same causes—namely, either because its essence, *or* definition, involves a contradiction, or because there is no external cause which has been determined to produce such a thing.

For Spinoza, everything that exists necessarily exists and everything that does

[6] Spinoza famously endorses a strong version of the Principle of Sufficient Reason according to which there is a cause or reason for the existence or nonexistence of each thing (E1p11d2).

not exist necessarily fails to exist, but there are different reasons that things are ruled into or out of existence. The existence and nonexistence of some things is necessitated by their own essence or nature. In E1p7, for instance, Spinoza proves that "it pertains to the nature of a substance to exist". As for nonexistence, there are square circles and other "Chimeras [*Chymaeram*]" that fail to exist by virtue of their essence (E1p11d2). In contrast, the existence or nonexistence of other things is due to their (external) efficient cause—thus, for example, the existence of a broken window (and the nonexistence of a non-broken window) might be due to the impact of a rock crashing through it. More generally, we can think of the essence or real definition of a thing (or things) as settling certain subject matters while leaving open how things stand with respect to other matters. Spinoza argues that God's essence necessitates the full *ordo naturae*, but the essence of any nonsubstance, taken by itself, must leave many subject matters unsettled, such as the matter of this nonsubstance's own existence (E1p24).

Though we might need as few as one possible world to represent the metaphysically possible, the abundance of worlds in our Spinozian models is helpful for modeling what is necessitated by the essences of things and what is possible relative to these essences, which for Spinoza can outstrip the metaphysically possible. At this point, it is helpful to assume that the domain assignments $\mathcal{D}_{\aleph}$, $\mathcal{D}_{\mathtt{Th}}$, $\mathcal{D}_{\mathtt{Ex}}$, ... map each world $w \in \mathcal{W}$ to sets of existents ("local domains") drawn from "global domains" $\mathfrak{D}_{\aleph}$, $\mathfrak{D}_{\mathtt{Th}}$, $\mathfrak{D}_{\mathtt{Ex}}$, ... of the respective sorts; that is, $\mathcal{D}_s(w) \subseteq \mathfrak{D}_s$ for each $s \in \mathcal{S}_{Spinoza}$. The global domains include all the things we wish to theorize about, whether existent or nonexistent, possible or impossible, conceivable or inconceivable—a global domain can even include Chimeras like square circles, mountains without valleys, and so forth. Let $\mathfrak{D}^*$ be the union of these global domains. We assume that along with the world-internal structure already introduced to implement Spinoza's theory of attributes, a Spinozian model includes an *essence function* $\mathcal{E}$ that assigns to each $a \in \mathfrak{D}^*$ the set of propositions necessitated or forced by its essence:

> **Essence-relative necessity**
> The essence function $\mathcal{E}$ maps each element $a \in \mathfrak{D}^*$ to a set of propositions $\mathcal{E}(a) \subseteq \mathcal{P}(\mathcal{W})$, where $P \in \mathcal{E}(a)$ iff P is true in virtue of a's essence. [7]

The worlds in the intersection $\bigcap \mathcal{E}(a)$, which we call the "essence set" of a, are compatible with every proposition necessitated by a's essence or nature whereas worlds outside this intersection are excluded by a's essence. Spinoza's claim in E1p7 that it pertains to the essence of a substance to exist entails that if a is a substance, then $\mathcal{E}(a)$ includes the proposition that a exists, and therefore every world in $\bigcap \mathcal{E}(a)$ is one in which a exists. In contrast, if a is a nonsubstance (i.e., a is a mode (E1p4d)), and thus, for Spinoza it exists by virtue of its efficient causes (E1p24), then $\mathcal{E}(a)$ cannot include the proposition

[7] To represent what follows from the essences of multiple things taken together, one could define essence functions on the set of nonempty subsets of $\mathfrak{D}^*$, rather than on $\mathfrak{D}^*$ itself.

that a exists, and $\bigcap \mathcal{E}(a)$ can include worlds in which a fails to exist.

As for metaphysical possibility itself, we assume that what is metaphysically possible is dependent on what is possible with respect to essences—specifically, to be metaphysically possible is to be possible relative to the essences of *all* things (in fact, for Spinoza only God's essence need be taken into account):

Metaphysical possibility
w is a metaphysical possibility iff $w \in \bigcap \mathcal{E}(a)$ for each $a \in \mathfrak{D}^*$.

Introducing the name '@' for the actual world in a model, we require that $@ \in \mathcal{W}$ be possible relative to the essence of any thing, which ensures that the actual world is metaphysically possible:

Actual is possible: $@ \in \bigcap \mathcal{E}(a)$ for each $a \in \mathfrak{D}^*$.

Spinoza's claim that God's essence fixes the order of Nature can be captured by the requirement that $\bigcap \mathcal{E}(\mathsf{God}_{\aleph}) = \{@\}$. This enforces that @ is the *only* metaphysically possible world in the model. But, again, Spinoza has to argue for this position, and so we also allow for models in which God's essence leaves open multiple metaphysical possibilities.

To summarize, there are three kinds of worlds in Spinozian models. First, there are metaphysically possible worlds, such as @, which are compatible with the essences of all things. Second, there are metaphysically impossible worlds which are compatible with the essence of no thing. These worlds will not play an important role in what follows and can be disregarded.[8] Third, there are metaphysically impossible worlds which, though ruled out by the essences of all things when taken together, are nevertheless compatible with the essence of some particular thing (or things) and can therefore be used to capture how this essence leaves various subject matters unsettled. A metaphysically impossible world $w \in \mathcal{W}$ lying in the essence set $\bigcap \mathcal{E}(a)$ of some $a \in \mathfrak{D}^*$ might still be regarded as an open possibility in the restricted sense that the essence of a alone doesn't rule out this world.

This brings us to the notion of *conceivability*, which is intimately related to essence-relative modality in Spinoza's philosophy and appears in several key texts at the beginning of the *Ethics* (see for example E1d1, E1a7, E1p10s, E1p11d1, and E1p14). At least in his early period, Spinoza seems to think that conceivability amounts to the possibility of positing certain ideas in an infinite intellect (see P4 in the KV). However, we want to remain fairly noncommittal about how Spinoza understands conceivability in his later philosophy. So, we hardwire conceivability into our Spinozian models by taking them to include a *conceivability function* $\mathcal{C}$ that assigns to each $a \in \mathfrak{D}^*$ the set of propositions conceivable about it in the "narrow sense"—when attending only to its essence (see [13])—which we call the "conceivability set" of a:

[8] These metaphysically impossible worlds might still be regarded as *epistemically possible* (see E1p33s1). The notion of epistemic possibility is crucial to Spinoza's theory of human psychology and the supervening disciplines of ethics and political philosophy.

Conceivability
The conceivability function $\mathcal{C}$ maps each member $a \in \mathfrak{D}^*$ to a set of propositions $\mathcal{C}(a) \subseteq \mathcal{P}(\mathcal{W})$, where $P \in \mathcal{C}(a)$ iff P is conceivable about a when considering only a's essence.

Spinoza's notion of conceivability is a rich topic that requires a great deal more attention than we can offer here (see [13] for further discussion). Particularly important is the connection between what pertains to the essence of a thing and what is conceivable about it given its essence, which can be made precise using the essence and conceivability functions in our models. For instance, E1a7 requires that if the proposition that a does not exist lies in the conceivability set $\mathcal{C}(a)$ ("If a thing can be conceived as not existing..."), then the proposition that a exists is not a member of $\mathcal{E}(a)$, and the essence set $\bigcap \mathcal{E}(a)$ can include worlds in which a fails to exist ("...its essence does not involve existence"). More generally, Spinoza seems to think that if the essence of a thing necessitates certain facts about this thing, then the thing cannot be conceived in ways that conflict with these essentialist facts, and this can be spelled out as the set-theoretic constraint that every proposition in its conceivability set is compatible with its essence set. [9]

4 The Spinozian Language and Model Theory

In this section, we present (most of) the formal Spinozian language used in our project and describe its model theory. First, the language: to represent the logical forms of sentences in the initial fragment of *Ethics* up through E1p15, we adopt a language whose logical symbols include the standard sentential connectives, the actualist quantifiers '$\forall$' and '$\exists$', the possibilist quantifiers 'Π' and 'Σ', and variables indexed to every Spinozian sort. For quantificational purposes, we also help ourselves to unindexed variables 'x', 'y', ... and overlined unindexed variables '$\bar{x}$', '$\bar{y}$', ... for denoting things of any sort in $\mathcal{S}_{Spinoza}$ and of any secondary sort in $\{\mathtt{Th}, \mathtt{Ex}, ...\}$ respectively. Whereas standard first-order languages have only a single symbol for identity, we have three:

Identity symbols: '$=$', '$=_{\mathrm{P}}$', and '$=_{\mathrm{C}}$' for standard, projective, and cross-attribute identity

The Spinozian language also includes the following modal operators:

Necessity-by-essence operators: '$\Box_t$' (read: *It is necessitated by the essence of t that...*) for each term t (constant or variable) in the language

Metaphysical necessity and possibility operators: '$\Box$' (read: *It is metaphysically necessary that...*) and '$\Diamond$' (read: *It is metaphysically possible that...*)

Conceivability operators: '$\blacklozenge_t$' (read: *It is conceivable about t when conceived solely in terms of its essence that...*) for each term t

[9] We do not build such correspondences between $\mathcal{E}$ and $\mathcal{C}$ directly into our Spinozian models so that axioms like E1a7 have some work to do.

The remaining logical symbols enable us to talk about sorts (notation: the sortal subscripts or lack thereof on argument positions of predicates indicate whether they can be instantiated by things of any sort (t), multifaceted things only ($t_{\aleph}$), single-faceted things only ($\bar{t}$), or things of some specific secondary sort ($t_{\mathtt{Th}}$, $t_{\mathtt{Ex}}$, ...)):

Sortal-projective predicates

$\mathsf{Same\text{-}sort}(t, t')$: *t and t' are the same sort of thing*
$\mathsf{All\text{-}sorts}(t_{\aleph})$: *$t_{\aleph}$ is projected into each of the infinitely many secondary sortal domains*

Turning to the nonlogical symbols of the language, we need a long laundry list of additional predicate for translating the text. Among these are the following predicates, which Spinoza defines in E1d1-E1d8:

$\mathsf{Causa\text{-}sui}(t)$: *t is a cause of itself*
$\mathsf{Finite\text{-}in\text{-}kind}(\bar{t})$: *$\bar{t}$ is finite in its own kind*
$\mathsf{Substance}(t)$: *t is substance*
$\mathsf{Attribute}(\bar{t})$: *$\bar{t}$ is an attribute*
$\mathsf{Mode}(t)$: *t is a mode*
$\mathsf{Affection}(t, t')$: *t is an affection of t'*
$\mathsf{God}(t_{\aleph})$: *$t_{\aleph}$ is God*
$\mathsf{Abs\text{-}infinite}(t_{\aleph})$: *$t_{\aleph}$ is absolutely infinite*
$\mathsf{Free}(t)$: *t is free*
$\mathsf{Eternal}(t)$: *t is eternal*

In addition to the predicate '$\mathsf{God}(t_{\aleph})$', the language also has the constants '$\mathsf{God}_{\aleph}$', '$\mathsf{God}_{\mathtt{Th}}$', '$\mathsf{God}_{\mathtt{Ex}}$', ... for referring directly to God, both as a multifaceted substance existing in the $\aleph$-domain and as a single-faceted substance (thinking substance, extended substance, and so on) existing in a secondary domain.

Well-formed formulae of the formal Spinozian language are generated from its lexicon through the usual grammar. We interpret these formulae relative to a pointed Spinozian model and to a variable assignment g that maps each variable of the language to some member of the corresponding global sortal domain(s)—where $\mathfrak{D}^*$ is the union of all the global domains and $\bar{\mathfrak{D}}^*$ is the union of only the global secondary domains, $g(x_s) \in \mathfrak{D}_s$, $g(x) \in \mathfrak{D}^*$, and $g(\bar{x}) \in \bar{\mathfrak{D}}^*$. Our official definition of a Spinozian model integrates the intraworld structure from section 2 with the interworld structure from section 3 and adds a function for interpreting the nonlogical symbols in the language:

Spinozian models
A many-sorted Spinozian model $\mathcal{M}$ is an ordered tuple consisting of a nonempty set $\mathcal{W}$ with designated point $@ \in \mathcal{W}$, a domain assignment $\mathcal{D}_s$ for each sort $s \in S_{Spinoza}$ mapping every world $w \in \mathcal{W}$ to a set of entities drawn from a global domain $\mathfrak{D}_s$, a projection function π_s for each secondary sort $s \in \{\mathtt{Ex}, \mathtt{Th}, ...\}$, an essence function $\mathcal{E}$, a conceivability function $\mathcal{C}$, and an interpretation function $\mathcal{I}$ mapping each constant in

the language to a member of the corresponding global domain and each n-adic nonlogical predicate symbol and world w to an n-ary relation over the global domains:

a. For each constant c_s, $\mathcal{I}(c_s) \in \mathfrak{D}_s$.
b. For each nonlogical predicate P, $\mathcal{I}(P(t_1, ..., t_n), w) \subseteq \mathfrak{D}^{*n}$.

Note that we interpret constant and predicate symbols over the global domains of the model, and not just over the local domains of worlds. This keeps the model theory flexible: constants can refer to both existing and non-existing things at a world, and predicates can be instantiated by both existents and nonexistents.

The semantics for the non-modal fragment of the language is relatively standard. We first compute the extensions of terms in the usual way:

Term denotations
The denotation $[\![t]\!]_{\mathcal{M},g,w}$ of term t at w with respect to $\mathcal{M}$ and g is defined as follows:

a. $[\![c_s]\!]_{\mathcal{M},g,w} = \mathcal{I}(c_s)$.
b. $[\![x_s]\!]_{\mathcal{M},g,w} = g(x_s)$, $[\![x]\!]_{\mathcal{M},g,w} = g(x)$, and $[\![\bar{x}]\!]_{\mathcal{M},g,w} = g(\bar{x})$.

We then compositionally assign satisfaction conditions to well-formed formulae using these denotations. Starting with atomic formulae, there are three cases to consider: predications, equations, and sortal-projective claims. To evaluate an n-adic nonlogical predicate symbol applied to n terms, we check to see whether the denotations of these terms stand in the n-ary relation expressed by the predicate:

Interpretation of predications
$\mathcal{M}, g, w \models P(t_1, ..., t_n)$ iff $\langle [\![t_1]\!]_{\mathcal{M},g,w}, ..., [\![t_n]\!]_{\mathcal{M},g,w} \rangle \in \mathcal{I}(P(t_1, ..., t_n), w)$

Standard, projective, and cross-attribute identity claims are evaluated by checking whether the denotations of terms on either side of the relevant identity symbol are identical in the senses discussed in Section 2:

Interpretation of identity claims
$\mathcal{M}, g, w \models t = t'$ iff $[\![t]\!]_{\mathcal{M},g,w} = [\![t']\!]_{\mathcal{M},g,w}$
$\mathcal{M}, g, w \models t_{\aleph} =_{\mathrm{P}} \bar{t}$ iff $[\![t_{\aleph}]\!]_{\mathcal{M},g,w} =_{\mathrm{P}} [\![\bar{t}]\!]_{\mathcal{M},g,w}$
$\mathcal{M}, g, w \models \bar{t} =_{\mathrm{C}} \bar{t}'$ iff $[\![\bar{t}]\!]_{\mathcal{M},g,w} =_{\mathrm{C}} [\![\bar{t}']\!]_{\mathcal{M},g,w}$

As for the sortal predicates '$\mathsf{Same\text{-}sort}(t, t')$' and '$\mathsf{All\text{-}sorts}(t_{\aleph})$', the former checks whether its arguments are of the same sort while the latter checks whether the multifaceted entity denoted by its argument is projected into each of the infinitely many secondary sortal domains:

Interpretation of sortal-projective claims

$\mathcal{M}, g, w \models \mathsf{Same\text{-}sort}(t, t')$	iff	$[\![t]\!]_{\mathcal{M},g,w} \in \mathfrak{D}_s$ iff $[\![t']\!]_{\mathcal{M},g,w} \in \mathfrak{D}_s$ for each sort $s \in \mathcal{S}_{Spinoza}$
$\mathcal{M}, g, w \models \mathsf{All\text{-}sorts}(t_{\aleph})$	iff	$\pi_s(w)([\![t_{\aleph}]\!]_{\mathcal{M},g,w})$ is defined for each secondary sort $s \in \{\mathtt{Th}, \mathtt{Ex}, ...\}$

Moving on to the sentential connectives, we assume that they have the classical semantics:

Interpretation of sentential connectives

$\mathcal{M}, g, w \models \neg\varphi$	iff	$\mathcal{M}, g, w \not\models \varphi$
$\mathcal{M}, g, w \models \varphi \wedge \psi$	iff	$\mathcal{M}, g, w \models \varphi$ and $\mathcal{M}, g, w \models \psi$
$\mathcal{M}, g, w \models \varphi \vee \psi$	iff	$\mathcal{M}, g, w \models \varphi$ or $\mathcal{M}, g, w \models \psi$
$\mathcal{M}, g, w \models \varphi \rightarrow \psi$	iff	$\mathcal{M}, g, w \not\models \varphi$ or $\mathcal{M}, g, w \models \psi$
$\mathcal{M}, g, w \models \varphi \equiv \psi$	iff	$\mathcal{M}, g, w \models \varphi \rightarrow \psi$ and $\mathcal{M}, g, w \models \psi \rightarrow \varphi$

We also give a standard treatment of universal/existential quantification, though we have a range of quantificational options corresponding to the different quantificational domains in our models (reflected in the availability of both actualist and possibilist quantifiers and the range of variable types in the language). Starting with actualist quantification over specific local secondary domains and letting $g_{[x_s \mapsto a_s]}$ be the variant assignment that is exactly like the variable assignment g except it sends the variable x_s to a_s, we evaluate quantified statements of the form '$\forall x_s\varphi$' and '$\exists x_s\varphi$' as follows:

Interpretation of actualist quantifiers

$\mathcal{M}, g, w \models \forall x_s\varphi$	iff	for all $a_s \in \mathcal{D}_s(w)$, $\mathcal{M}, g_{[x_s \mapsto a_s]}, w \models \varphi$
$\mathcal{M}, g, w \models \exists x_s\varphi$	iff	for some $a_s \in \mathcal{D}_s(w)$, $\mathcal{M}, g_{[x_s \mapsto a_s]}, w \models \varphi$

Actualist quantification with unindexed variables is analogous—where $\mathcal{D}^*(w)$ is the union of all the local domains of w (i.e., the set of all existents of any sort in w) and $\bar{\mathfrak{D}}^*(w)$ is the union of only the local secondary domains (i.e., the set of all existents of any secondary sort in w), we give the following additional semantic entries:

Interpretation of actualist quantifiers (continued)

$\mathcal{M}, g, w \models \forall x\varphi$	iff	for all $a \in \mathcal{D}^*(w)$, $\mathcal{M}, g_{[x \mapsto a]}, w \models \varphi$
$\mathcal{M}, g, w \models \exists x\varphi$	iff	for some $a \in \mathcal{D}^*(w)$, $\mathcal{M}, g_{[x \mapsto a]}, w \models \varphi$
$\mathcal{M}, g, w \models \forall \bar{x}\varphi$	iff	for all $\bar{a} \in \bar{\mathfrak{D}}^*(w)$, $\mathcal{M}, g_{[\bar{x} \mapsto \bar{a}]}, w \models \varphi$
$\mathcal{M}, g, w \models \exists \bar{x}\varphi$	iff	for some $\bar{a} \in \bar{\mathfrak{D}}^*(w)$, $\mathcal{M}, g_{[\bar{x} \mapsto \bar{a}]}, w \models \varphi$

We also allow for possibilist quantification over the global domains of a model. Whereas '$\forall$' and '$\exists$' quantify over only the existing things in a world, the possibilist universal and existential quantifiers 'Π' and 'Σ' quantify over all things, whether they exist or not, and whether they are conceivable at the world of evaluation or not. We provide the following clauses for general quantified statements of the form '$\Pi x\varphi$' and '$\Sigma x\varphi$' (the remaining cases are similar):

Interpretation of possibilist quantifiers

$\mathcal{M}, g, w \models \Pi x \varphi$ iff for all $a \in \mathfrak{D}^*$, $\mathcal{M}, g_{[x \mapsto a]}, w \models \varphi$

$\mathcal{M}, g, w \models \Sigma x \varphi$ iff for some $a \in \mathfrak{D}^*$, $\mathcal{M}, g_{[x \mapsto a]}, w \models \varphi$

In some parts of the *Ethics*, Spinoza clearly has actualist quantification in mind. In others, he needs possibilist quantification. In still others it is unclear what he intends to quantify over. In our broader project, we adopt a conservative methodology and try to get by with only actualist quantification as much as possible. We introduce possibilist quantification only when it is absolutely required (starting with our treatment of E1p11d2).

The remaining entries are for the modal operators whose semantics involves the essence and conceivability functions. The essence of a thing necessitates that φ iff the proposition expressed by φ with respect to $\mathcal{M}$ and g (i.e., the set $\{w : \mathcal{M}, g, w \models \varphi\}$) lies in the essence function $\mathcal{E}$ applied to this thing:

Interpretation of necessary-by-essence operators

$\mathcal{M}, g, w \models \Box_t \varphi$ iff $\{w : \mathcal{M}, g, w \models \varphi\} \in \mathcal{E}(\llbracket t \rrbracket_{\mathcal{M},g,w})$

Metaphysical necessity/possibility is necessity/possibility relative to the essences of all members of the global domains of the model:

Interpretation of metaphysical modality operators

$\mathcal{M}, g, w \models \Box \varphi$ iff for all $v \in \mathcal{W}$ s.t. $v \in \bigcap \mathcal{E}(a)$ for each $a \in \mathfrak{D}^*$, $\mathcal{M}, g, v \models \varphi$

$\mathcal{M}, g, w \models \Diamond \varphi$ iff for some $v \in \mathcal{W}$ s.t. $v \in \bigcap \mathcal{E}(a)$ for each $a \in \mathfrak{D}^*$, $\mathcal{M}, g, v \models \varphi$

Finally, it is conceivable about a thing that φ when attending to only its essence iff the proposition expressed by φ with respect to $\mathcal{M}$ and g is a member of the conceivability set of this thing:

Interpretation of conceivability operators

$\mathcal{M}, g, w \models \blacklozenge_t \varphi$ iff $\{w : \mathcal{M}, g, w \models \varphi\} \in \mathcal{C}(\llbracket t \rrbracket_{\mathcal{M},g,w})$

Having recursively assigned satisfaction conditions to every formulae of the Spinozian language relative to a pointed model and variable assignment, we can next define truth for its sentences in the usual way, and we identify valid arguments as those that preserve truth in all pointed models.

5 Conclusion

With this model theory in place, one can next get down to work formally reconstructing Spinoza's demonstrations in the *Ethics* (as we have done in the full version of this project). The strict requirement of formal proof provides a powerful diagnostic tool for identifying tacit premises, redundancies, and potential errors in Spinoza's "Geometric manner". Spoiler: pace Leibniz [11], Bennett [1], and others, we find Spinoza to be a skilled logician whose deductive argumentation, for the most part, holds together remarkably well. While many

of Spinoza's proofs are enthymemes that require implicit unstated premises to go through, prolonged exegetical gymnastics isn't required to fill in most of the holes, and many of the unstated premises are trivial. Our positive assessment extends to Spinoza's modal reasoning: far from being an incompetent modal logician, Spinoza operates nimbly with complex modal concepts in many of his demonstrations, which is all the more impressive given that he had nothing like the modern technology of modal logic at his disposal.

Making this case on Spinoza's behalf must be left for another occasion. But even from developing the core model theory in this paper, we hope to have already helped to undermine a common perception among philosophers and scholars of the history of philosophy that precise philosophical formalization is inconsistent with historical precision. Precise philosophy and precise history of philosophy needn't come at the expense of one another, and in the current study we strived to achieve both kinds of precision by developing a rigorous formal architecture for theorizing about Spinoza's ontology and modal metaphysics.

References

[1] Bennett, J., "A Study of Spinoza's Ethics," Hackett, Indianapolis, 1984.

[2] Della Rocca, M., *Spinoza's argument for the identity theory*, Philosophical Review **102** (1993), pp. 183–213.

[3] Della Rocca, M., "Representation and the Mind-Body Problem in Spinoza," Oxford University Press, New York, 1996.

[4] Descartes, R., "The Philosophical Writings of Descartes *(3 volumes). Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch*," Cambridge University Press, Cambridge, 1985.

[5] Fine, K., *Essence and modality*, Philosophical Perspectives **8** (1994), pp. 1–16.

[6] Fine, K., *The logic of essence*, Journal of Philosophical Logic **24** (1995), pp. 241–273.

[7] Fine, K., *Semantics for the logic of essence*, Journal of Philosophical Logic **29** (2000), pp. 543–584.

[8] Fitting, M. and R. L. Mendelsohn, "First-Order Modal Logic," Kluwer Academic Publishers, Dordrecht, 1998.

[9] Garrett, D., *The indiscernibility of identicals and the transitivity of identity in Spinoza's logic of the attributes*, in: Y. Y. Melamed, editor, *Cambridge Critical Guide to Spinoza's Ethics*, Cambridge University Press, Cambridge, 2017 pp. 12–42.

[10] Jarrett, C., *The logical structure of Spinoza's* Ethics*, part 1*, Synthese **37** (1978), pp. 15–65.

[11] Leibniz, G. W., "Philosophical Paper and Letters*. Edited by Leroy E. Loemker*," D. Reidel, Dordrecht, 1969.

[12] Melamed, Y. Y., *Spinoza's deification of existence*, Oxford Studies in Early Modern Philosophy **6** (2012), pp. 75–104.

[13] Newlands, S., "Reconceiving Spinoza," Oxford University Press, Oxford, 2018.

[14] Schmidt, A., *Über deduktive theorien mit mehreren sorten von grunddingen*, Mathematische Annalen **115** (1938), pp. 485–506.

[15] Schmidt, A., *Die zulässigkeit der behandlung mehrsortiger theorien mittels der üblichen einsortigen prädikatenlogik*, Mathematische Annalen **123** (1951), pp. 187–200.

[16] Spinoza, "Opera *(4 volumes). Edited by Carl Gebhardt*," Carl Winter, Heidelberg, 1925.

[17] Spinoza, "The Collected Works of Spinoza *(2 volumes). Edited and translated by Edwin Curley*," Princeton University Press, Princeton, 1985-2016.

[18] Wang, H., *Logic of many-sorted theories*, Journal of Symbolic Logic **17** (1952), pp. 105–116.

[19] Wolfson, H. A., "The Philosophy of Spinoza *(2 volumes)*," Harvard University Press, Cambridge, 1934.