Equivalence and Theory Expansion

Diego Arana Segura*
diego.arana@rutgers.edu

Abstract

This paper presents a novel framework for understanding theoretical equivalence that reconciles two familiar approaches to the problem: formal and content-based. Formal approaches are based on the logical and syntactical features of theories, while content-based approaches focus on their content as construed in various metaphysical approaches to semantics. I argue that these approaches are complementary and that a deeper view of equivalence emerges when we consider a theory's expansion potential—its capacity to be embedded in broader theoretical contexts. This notion links content to syntax, as syntactical structure constrains how a theory's expressions can be used, and this use in turn determines the theory's possibilities for representing the world across different theoretical contexts. The framework is applied to various cases where formal and content-based approaches face difficulties, showing that it can explain the equivalence and inequivalence of theories where previous approaches failed, and that it can better justify the strategies employed by proponents of these approaches in each case.

1 Introduction

Two theories are said to be metaphysically equivalent when they say the same about the world, and merely differ in how they say it. Unlike epistemic or cognitive equivalence, which concern how agents come to know or believe theories, metaphysical equivalence focuses on the *aspects of the world* that theories capture.

*I would like to thank Ted Sider for feedback on numerous drafts, without which I could not have written this paper. I am also grateful to Fernanda Birolli, Sam Elgin, Michael Glanzberg, Adrian Liu, Brian McLaughlin, Travis McKenna, Jill North, Eileen Nutting, Andrew Rubner, Ezra Rubenstein, Jonathan Schaffer, Alexander Skyles, Bram Vaassen, Minghui Yang, the participants at the 2022 Rutgers WIP talk, 2023 Rutgers Third-year Seminar, 2023 Rutgers Dissertation Seminar, and the Central APA 2023 Division Meeting for their helpful comments and suggestions.

This notion of equivalence often arises in debates in philosophy of physics (Rynasiewicz, 1996; Wallace and Timpson, 2009; Weatherall, 2014), philosophy of mathematics (Linnebo and Rayo, 2012), and metametaphysics (Hirsch, 2009). Claims of equivalence are often used by a third party as a tool to defuse or question the legitimacy of some debates about how the world is, but they are also used to defend realist views by bypassing epistemic problems, like certain cases of underdetermination of theory by evidence (Glymour, 1977). These claims are typically justified by constructing translation schemes between opposing languages or formalisms, showing evidence that such schemes are constructible, or showing that the (model-theoretic) semantics of the theories are similar in some sense—such as having isomorphic or dual categories of models.

But what *is* it for two theories to be equivalent? What does equivalence *amount to*? The literature roughly divides into two approaches to this question. The first approach considers equivalence as a relation based on formal or syntactical criteria, analyzing the syntactical or model-theoretic structure of the theories (Glymour, 2013; Barrett and Halvorson, 2016a; Barrett and Halvorson, 2022; Dewar, 2023). The second approach focuses on the content of the theories as construed in various metaphysical approaches to semantics (Hirsch, 2009; Coffey, 2014; Teitel, 2021). I will refer to these as the *formal* and *content-based* approaches, respectively.

Each approach addresses problems that the other misses, but they are often perceived as being in tension, and proponents of one tend to avoid (Hirsch, 2009; Barrett and Halvorson, 2016b; Halvorson, 2019) or even call into question (Coffey, 2014; Teitel, 2021; Dewar, 2023) notions characteristic of the other. In this paper, I aim to bridge this gap by proposing a framework that integrates elements of both. A key observation is that proponents of both approaches often employ similar strategies when their criteria deliver unsatisfactory verdicts. Specifically, they frequently embed the theories in larger languages and focus on the properties of these embeddings rather than on the original theories themselves.

I argue that this is not accidental, and that these strategies can be understood as attempts to determine the *expansion potential* of the theories in question—their capacity to be embedded in broader theoretical contexts, which depends on how their expressions are *used*. This notion links content to syntactic and formal structure, as this structure constrains how a theory's expressions can be used, and this use in turn determines the theory's possibilities for representing the world across different theoretical contexts. The framework I propose is based on this idea, and is applied to a variety of cases where formal and content-based approaches face difficulties.

The paper is structured as follows. In Sections 2 and 3, I examine formal and content-based approaches to theoretical equivalence, respectively, and discuss the problems they face. In both of these sections, I show that proponents of these approaches often rely on the notion of expansion, and on features of it that are not explained in terms of the purely formal or purely content-based considerations. In Section 4, I introduce the concept of expansion potential and demonstrate how it can integrate the insights of both approaches. In Section 5, I apply the framework to various cases, showing how it can explain the equivalence and inequivalence of theories where previous approaches faced difficulties, and that it can better justify the strategies employed by proponents of these approaches in each case. Finally, in Section 6, I conclude by discussing further implications of the current approach.

2 Formal Approaches

Formal approaches to theoretical equivalence are based on the formal and syntactical features of theories. They involve criteria such as logical equivalence, definitional equivalence, Morita equivalence, mutual interpretability, and the like (Barrett and Halvorson, 2016b). These approaches analyze theories as sets of sentences within formal languages or as model-theoretic entities, establishing equivalence based on formal properties like derivability and model structure.

Most criteria within this approach establish equivalence between theories considered as sets of sentences in "uninterpreted" formal languages. For instance, logical equivalence holds between two theories T_1 and T_2 in the same language L when they have the same closure under L's consequence relation $\vdash$, which can be defined purely syntactically or by specifying a model-theoretic semantics. This allows us to define the notion of logical equivalence in a manner largely independent of the particular features of the world theories describe. Definitional equivalence and Morita equivalence extend this idea to theories stated in different languages. Specifically, both hold between two theory-language pairs $\langle T_1, L_1\rangle$ and $\langle T_2, L_2\rangle$, with L_1 and L_2 being possibly distinct languages, when there exists a common language L^+ containing (iterated) definitional or Morita extensions T_1^+ of T_1 and T_2^+ of T_2, such that T_1^+ and T_2^+ are logically equivalent in L^+. But the construction of these extensions is purely syntactic, which makes the criteria largely independent of the contents of the theories involved.[1]

A main criticism of formal approaches stems precisely from their independence from content, which can lead to overgenerating equivalence claims in cases where formally equivalent theories differ significantly in meaning. For example, the statements 'All lions have stripes' and 'All tigers have stripes'—hereby, the *Sklar example*—might be logically or model-theoretically equivalent under some criteria, but they make different claims about the world (Sklar, 1982; Sider, 2020; Teitel, 2021). Proponents of formal approaches are generally aware of this issue and often maintain that formal equivalence is only a *necessary* but not *sufficient* condition for theoretical equivalence. This has led to the widespread view that formal structure must be supplemented with *empirical content* to yield an adequate criterion (Quine, 1975).

[1] That is, whether $\langle T^+, L^+\rangle$ is a definitional or Morita extension of $\langle T, L\rangle$ is determined by the syntactic properties of the languages involved.

However, this idea faces two difficulties. The first is the now familiar objection that the theory-observation distinction is not as clear-cut as it might seem (Teitel, 2021, p. 4129). While in specific contexts we may identify empirical content reliably—e.g. agreement of stress-energy in GR or Born-Rule Expectations in QM—, an even somewhat general account of what constitutes a theory's empirical content across cases remains elusive. The second difficulty concerns the role of structure: once empirical content is brought in to do the substantive work in equivalence judgments, it is unclear what contribution formal structure continues to make. What exactly does it add to a *worldly* equivalence judgment, beyond helping us track syntactic or inferential patterns?

A typical response appeals to the idea that structure fixes the meaning of theoretical terms by specifying their *functional role* within the theory. Thus, theoretical terms get their content not in isolation, but in virtue of the role they play in a theory, and the theory gets its content from the empirical claims it makes about the world. First, I want to note that this response commits one to some kind of holism, which I don't object to, and is even congenial to the current approach. Second, that it simply pushes the bump: even if we had an account of how theoretical terms get their content via their connection to the entire theory, we would still need an account of what the content of the entire theory is. Formal approaches to equivalence are unable to provide such an account, unless one accepts that the content of two *formally* equivalent "entire theories"—whatever that means—are the same, which is either obscure or implausible.

Given this, there is pressure to explain equivalence without either collapsing formally equivalent theories, or explicitly invoking notions of content. For reasons of space, I will not be able to review all such attempts, but an interesting attempt is found in Dewar (2023). Here, Dewar, proposes a solution that relies on embedding the theories to be compared into a common language and justifying the equivalence or inequivalence of the theories through the *formal* articulation of this common language. To illustrate his approach, he uses two versions of Maxwell's equations, M_ρ and M_μ, which differ only in the symbol used for charge density ('ρ' vs. 'μ'). In cases like this, one should obviously conclude that the difference is simply notational, and that the theories are equivalent. However, nothing in the formal properties of the theories themselves distinguishes this case from cases of genuinely inequivalent theories which just happen to be syntactically isomorphic, like the Sklar example. Dewar argues that the difference in this case is that the theories are inter-derivable within a joint language, and that the rules of this language permit the derivations that justify the equivalence claim. More precisely, if $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$ are the languages of M_ρ and M_μ, respectively, and $\mathcal{L}^+$ is the joint language, then the judgment of equivalence is justified by the fact that any sentence ϕ_ρ in $\mathcal{L}_\rho$ and its counterpart ϕ_μ in $\mathcal{L}_\mu$ are inter-derivable within $\mathcal{L}^+$. Additionally, the rules of $\mathcal{L}^+$ permit the derivations $M_\rho \vdash \phi_\mu$ and $M_\mu \vdash \phi_\rho$ if $\phi_\mu \in M_\mu$ and $\phi_\rho \in M_\rho$, and one can also derive '$\rho = \mu$' within $\mathcal{L}^+$ (p. 16).

However, notice that this is a case where we *start* with two theories M_ρ and M_μ in languages $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$, respectively, and *produce* a joint language $\mathcal{L}^+$ where the two theories are inter-derivable.[2] In this case, the antecedently accepted rules of $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$ are not sufficient to derive '$\rho = \mu$' within $\mathcal{L}^+$, or any of the other sentences that justify the equivalence claim. Granted, we can *choose* to add these as rules or axioms to $\mathcal{L}^+$, but this is not forced on us by the formal properties of the languages involved. Without further justification, it is unclear why we should accept the specific rules of $\mathcal{L}^+$ that facilitate these derivations. That this move is unjustified can be seen by further reflecting on the Sklar example. Suppose we found ourserves in a similar situation, where we need to *produce* a joint language with 'lion' and 'tiger' where there was previously none.[3] Given this, if we disregard the meaning of those worlds and focus only on formal properties, we can construct a language where the one-sentence theories 'There is a tiger' and 'There is a lion' are inter-derivable, and where 'Everything is a lion if and only if it is a tiger' is an axiom. This would lead to the conclusion that the theories are equivalent, which is clearly unacceptable. So, whatever distinguishes the Sklar example from Dewar's case is not the formal properties of the languages involved. But then, what is it?

Proponents of content-based approaches have a straightforward answer: sentences that are inter-derivable in the joint language are such because *have the same content* as interpreted by their respective languages, and they preserve their content under expansion. This explanation is not available to proponents of formal approaches, as it would require admitting that content-based considerations underwrite the judgment of equivalence, or invoking a notion of empirical content, which would bring us back to the first difficulty. Instead, Dewar (2023) regards the rules of the joint language as the stopping point of the explanation, but this seems an odd place to stop: since we are dealing with the rules of a joint language, it is natural to seek an explanation of these rules in terms of the languages that are being joined.

Rather than stopping at the rules of the joint language, I propose an explanation that appeals to the *use* of the expressions in the original languages. The explanation will be provided in detail in Section 5.1, but I can sketch the general idea here. The use of the expressions in $\mathcal{L}_\rho$ grants us access to certain expansions of it, and similarly for $\mathcal{L}_\mu$. These possibilities of expansion are not determined by the formal properties of the languages alone, but by the ways in which we *use* the expressions of the relevant languages. In this case, it happens that any common expansion of the theories that respects the use of the expressions in $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$ will permit the required derivations. From here, the explanation can proceed analogously to Dewar's.

[2] Some may argue that we already have $\mathcal{L}^+$, but I find this implausible. It is not as if, whenever we want to integrate two theories, we always find ourserves already in possession of a joint framework. What actually happens is closer to how the current approach cashes it out: we find ourselves using the unintegrated theories, embedded in other practices that *constrain* how we can integrate them.

[3] Say, suppose two disjoint linguistic communities inhabited disjoint regions where there were only lions, and tigers, respectively, and that we are the first descendants of the "joint" community.

More generally, I propose that one can justify an equivalence claim between two formally equivalent languages by finding a joint language that is *accessible from both* and possesses the desired properties. However, this relation of accessibility should not be conceived as holding purely in virtue of the formal properties of the languages involved. Indeed, there will typically be many ways of embedding two uninterpreted theories into a joint language, and many of these will not yield the desired verdicts. In such cases, considering their formal properties alone will not provide any reason to think that the rules of the joint language will have the desired properties.

As we will see in Section 5.2, this point extends to formal notions based on definitional or Morita "extensions" of theories: many such extensions are possible, but the relevant ones are those that respect the uses of the expressions in the theories to be compared. Additionally, I will show that notions which are rarely explained in terms of formal properties, but which are central to the application of formal criteria—such as *taking a sentence as a definition*—can be understood in terms of the current framework. Even more, I will show that once we take these notions into account, theories that are (wrongly) deemed equivalent by formal criteria can be distinguished.

Having explored formal approaches, we now turn to content-based approaches to see if they fare any better. As we will see, content-based approaches face similar problems, and similar strategies are employed to address these issues. This will help show that the proposed framework is not arbitrary and that it is a natural step to take within the context of the current debate.

3 Content-Based Approaches

Content-based approaches to theoretical equivalence focus on the *content* of theories as construed in various metaphysical approaches to semantics. For two theories to be equivalent under this approach, they must (i) be able to express the same contents—and thus be *expressively equivalent*[4]—, and (ii) be committed to the truth of the same of these contents. These "contents" can be sets possible worlds (Kripke, 1959), sets of truthmakers and falsemakers (Fine, 2017), sets of possibilities (Holliday, 2021), or any other metaphysical construct that can be used to specify a theory's *demands on the world*.

What is common in such constructions is that the contents in question are coarse-grained enough to be expressed by syntactically distinct theories and that they are in some sense *worldly*. For instance, the *possible worlds* in possible world semantics have been taken to be concrete objects (Lewis, 1986) or properties that the world might have (Stalnaker, 2007). Similarly, the *truthmakers* in truthmaker

[4]Admittedly, one can think of a coarser notion of equivalence that does not require expressiveness, and only sameness of content of the theory—where the *theory's content* is understood as conjunction of all the contents of the sentences entailed by the theory. Nevertheless, I opt for including expressive equivalence since it has been explicitly invoked by proponents of content-based approaches (e.g. Hirsch, 2009).

semantics have been taken as parts or aspects of the world (Fine, 2020). Approaches like these have been adopted or implicitly assumed in works such as Hirsch (2005; 2009), Warren (2015), Rayo (2017), and Teitel (2021). Like formal criteria, this requires us to consider theories as sets of sentences in languages, but it also requires us to consider functions that assign contents to these sentences—let's call these *content assignments* or simply *assignments*. We should note that, since contents are assumed to have a natural *entailment* relation[5], one can reconstruct syntactic or model-theoretic consequence in a language using these contents.[6]

Content-based approaches attempt to sidestep syntactical structure and intend to directly capture the *demands on the world* that theories make. This allows them to distinguish between syntactically isomorphic theories which make different demands on the world, effectively addressing cases where formal approaches fail. For instance, the difference between 'All lions have stripes' and 'All tigers have stripes' is simply that they make different demands on the world as interpreted by the languages they belong to, which should be cashed out in terms of the contents these sentences express in their respective languages.

However, content-based approaches face their own challenges regarding the overgeneration of equivalence claims. For instance, coarse-grained content-based criteria distinguish theories based solely on (i) truth across the same possible worlds, and (ii) identical expressive power of their background languages in terms of coarse-grained contents. Consequently, such criteria unintuitively equate theories that are widely regarded as distinct—such as ZF vs PA, mereological nihilism vs universalism, or more generally different metaphysical theories which are necessarily true or false and are formulated in languages which can only express necessary truths or falsehoods. While one might accept some equivalences (e.g., Hirsch on specific versions of nihilism and universalism), counting all such theories as equivalent would trivialize substantial logical, mathematical, and metaphysical distinctions.

Faced with this problem, proponents of content-based approaches have a choice: they can either accept the equivalences, reject the received view about the content of the affected theories, or make adjustments to distinguish these theories within their framework. The first option carries a high theoretical cost, since it forces one to accept that a wide variety of theories that are considered inequivalent in their respective domains, and *across* domains, are actually equivalent. Thus, I will not consider this option further.

The second option seems promising at first, but it carries greater theoretical costs than one might initially think. One way of carrying it out would be to accept the received view about contents,

[5]For instance, if they are construed as sets of possible worlds, then entailment is the *subset* relation. If they are construed as sets of facts or truthmakers, then it is the *superset* relation. Other constructions require different relations, but every such construction has a natural entailment relation, as far as I can tell.

[6]Given an assignment I of contents to the sentences of L, a set of sentences $\Gamma \cup \{\phi\}$ of L, we say that $\Gamma \vdash_I \phi$ iff, for every content c, if c entails every $I(\psi)$ for $\psi \in \Gamma$, then c entails $I(\phi)$.

but to hold that mathematical theories are *contingently* true or false. This is the option taken by Field (2016) and Balaguer (2021), among others. The problem is that this solution does not apply across the board, since we also want to account for logical and metaphysical theories which are (presumably) necessarily true.

A third kind of approach is adopted by content-based theorists who maintain that logical and mathematical truths are necessary, but rely on finer-grained notions of *content*. These approaches hold that multiple distinct contents can be "necessary," or share broadly the same truth-conditions—in some suitable sense of "truth-conditions" or "necessary." There are several ways one might do this.

One relatively straightforward way of doing this involves distinguishing between different classes of possibility—logical, mathematical, metaphysical, and so forth—and evaluating theories not simply by the total set of possibilities they rule in or out, but rather by how each theory partitions these separate classes of possibilities.[7] Under this strategy, (true) logical theories exclude logically impossible worlds, mathematical theories exclude mathematically impossible worlds, and so forth. We can therefore say that both ZF and PA are *metaphysically* (or perhaps *physically*) necessary while still holding that they differ in other aspects of their content—e.g. maybe because they are true in different *logically* possible worlds. Another way of fine-graining contents, exemplified by philosophers like Sider (2009), holds that metaphysical equivalence requires equivalence not merely in the possibilities selected by each theory but also in how these possibilities are *structured* or "carved up" by each theory's constants, predicates, quantifiers, and so on.[8] Yet another variant involves appealing to the *grounding relations* that theories posit, distinguishing theories that otherwise select the very same possibilities (Schaffer, 2009; Fine, 2012).[9]

While I do not claim to have a decisive argument against these options,[10] I do have some concerns. First, all of these approaches employ finer-grained distinctions between *contents* compared to standard coarse-grained frameworks, and therefore require greater theoretical resources. The second, related concern is that using these contents to account for *metaphysical equivalence*[11] as currently articulated, commits one to there being *aspects of reality* that respond to these distinctions. For instance, if one employs distinct classes of possibility—logical, mathematical, metaphysical—then one commits oneself to the existence of distinctively logical, mathematical, or metaphysical worldly

[7]Thanks to an anonymous reviewer for suggesting this option.

[8]A similar strategy is pursued by North (2021) in the context of physical theories using the content of "perspicuous representation".

[9]Grounding distinctions depend roughly on the order of definitions within a theory, distinguishing theories that select the same possibilities while using a different order of definitions to do so.

[10]Indeed, the diversity within this family of strategies makes it challenging to provide a unified argument against all of them.

[11]As clarified in the first paragraph of the introduction, I use "metaphysical equivalence" not to mean equivalence concerning a specific philosophical domain (e.g., "equivalence concerning metaphysics"), but rather equivalence regarding *the aspects of the world* that theories capture. Two theories, on this understanding, are metaphysically equivalent just in case they capture exactly the same *worldly aspects*, even if they differ linguistically or representationally. This, of course, doesn't completely pin down metaphysical equivalence, but it gives one a rough idea of the target of this paper.

aspects. Such a commitment is significantly stronger than distinguishing between, say, epistemic and metaphysical possibility: the latter distinction does not entail that every pair of epistemically inequivalent theories capture different aspects of reality.[12] In contrast, making these finer-grained distinctions explicitly metaphysical (i.e. *worldly*) raises potentially problematic questions: what makes a particular worldly aspect distinctively "logical," rather than "mathematical" or "physical"? How many *kinds* of possibility there are and why? The proponent of such a approach will have pressure to articulate this further structure in worldly terms, or posit it as a primitive, which incurs in additional theoretical costs. Similar concerns apply to other fine-grained approaches. For example, relying on hyperintensional notions like joint-carvingness commits one to the world having an objective "carving structure" that theories might capture or fail to capture, and which can serve as a difference-maker for theories that otherwise capture the same possibilities. Similarly, grounding-based accounts must posit worldly features corresponding to "grounding structure", which can be captured or fail to be captured purely because of the order of definitions within theories. Furthermore, the more structure one imposes on *contents*, the greater the risk of conflicting with plausible accounts of theoretical equivalence in the philosophy of physics and mathematics that posit a *different* structure. For example, employing linguistically structured contents[13] to adjudicate metaphysical equivalence implicitly commits one to the idea that facts themselves are linguistically structured—a position difficult to reconcile with views of equivalence common in the philosophy of physics and mathematics, according to which differently structured sentences can represent the very same fact in a fully adequate and complete manner.[14]

Another last approach is to preserve both the received view about the content of logical and mathematical theories and the coarse-grained notion of content, but to locate the relevant difference not in the theories themselves but in certain *expansions* of them. This is done by Hirsch (2009). Following the line of the content-based approach, Hirsch argues that the endurantism vs. perdurantism debate is *merely verbal*, since there are alternative interpretations of the quantifiers in each theory that makes them express and commit to the same contents—call these *Hirschean reinterpretations*.[15] However, he also wants to say that the dispute between mathematical platonists and nominalists is a genuine one, which requires the theories to be *inequivalent*. But, as in the cases mentioned above, this is one where platonism is necessarily true, mathematical entities exist by necessity if at all, and their intrinsic properties hold necessarily. Thus, if both the nominalist and the platonist formulate their theories in minimal languages, then platonism will be equivalent to nominalism, if true. By

[12] Indeed, it is common to think of metaphysically equivalent but epistemically inequivalent theories as *capturing the same aspects of reality* but in ways that differ in their epistemic accessibility.

[13] Structured propositions like this are used in Soames (2009) and Salmon (1986). However, it is unclear whether these authors intended to use them as difference-makers for metaphysical equivalence.

[14] This point is emphasized by Wallace (2022) in philosophy of physics and Rayo (2013) in philosophy of mathematics.

[15] Hirschean reinterpretations are needed because, if interpreted using the same language, endurantism and perdurantism are incompatible. But this doesn't preclude the debate from being merely verbal. Hirsch tries to show that the debate is merely verbal by considering reinterpretations of the languages that respect the ways in which the endurantists and the perdurantists use their quantifiers, and considering the contents that these reinterpretations express.

considering the dispute genuine and thereby the theories inequivalent, Hirsch is not intending to endorse nominalism, so he must find a way to distinguish the theories even on the assumption that platonism holds.

Hirsch's way of tackling this problem is to point out that in "this dispute it is important not to focus exclusively on sentences of pure set (property, number) theory, which are either necessarily true or necessarily false" (2009, p. 253). Indeed, once we focus on sentences of *mixed* languages, we seem to be able to find coarse-grained propositions that the platonist can express and the nominalist cannot, making the positions intensionally inequivalent.[16] Hirsch presents the following sentence as an example:

(1) There are two [nondenumerably] infinite sets X and Y, whose members are [nondenumerably] infinite sets of angels, satisfying the condition that, for any set X' in X, there is a set Y' in Y such that all angels in X' love all and only angels in Y', and some angel in Y' loves some angel in some set in X other than X'.

Here, 'angel' and 'loves' could be replaced by any suitable predicates. But distinguishing platonism and nominalism this way requires that their theories be embedded in a larger language in which propositions like (1) can be expressed. Thus, Hirsch doesn't really answer the question of whether the theories are equivalent or not *when stated in their minimal languages*. Instead, he shows that close variants of these theories are inequivalent. But why should we care about these variants when we are interested in the original theories? Whether this appeal to mixed languages adequately addresses equivalence partly depends on his underlying metasemantic commitments, whose discussion would take us too far afield, but it is important to note the explanation is not simple or directly connected to *sentential contents*.[17]

At this point, we should notice a remarkable resemblance between Hirsch's and Dewar's strategies in dealing with the limitations of their respective frameworks: that they embed the theories to be compared into larger languages and focus on the properties of these embeddings rather than on the original theories themselves. It is also true, in both cases, that we cannot embed these theories in the larger language in any arbitrary way, and possibilities for embedding are not determined solely by the formal properties of the theories involved or by the (coarse-grained) contents of the sentences in the original theories.

[16]It is worth noting that this example is being evaluated in the context of a first-order formulation of both theories. Once we switch to second-order logic, the nominalist might be able to express the sentence (1). But notice that, when comparing second-order nominalism with second-order platonism, there might be *other* contingent propositions that the second-order platonist can express and the second-order nominalist cannot. This suggests looking at many different expansions or formulations of the same metaphysical thesis (e.g. first- vs. second-order), which supports the approach I am trying to advance (Section 4.1). I will come back to this in Sections 5.3 and 5.4.

[17]It should also be noted that this explanation is different than the one he considers for, e.g., perdurantism and endurantism in Hirsch (2009) since expansions are not explicitly mentioned in this latter case.

I maintain that this is not accidental, and that theories do the *work* they do precisely in virtue of these possibilities of embedding them in larger languages. Theories are *tools* that we use to represent the world, and they fulfill this function in several theoretical contexts, in which they are typically embedded in other theories. This is not to say that theories cannot represent the world in isolation, but rather that this phenomenon is a *limiting case*; it is not the main way in which theories represent the world. Since these embeddings can be thought of as expansions of the original theory, I will call a theory's possibilities for embedding its *expansion potential*. A theory's expansion potential depends on how we *use* its syntax and how we can extend that use into other theoretical contexts.

As before, I will provide a detailed explanation in Section 5.3, but I can sketch how platonism and nominalism can be distinguished in the proposed framework. Suppose $L_\in$ is the (uninterpreted) set-theoretic language and $T_\in$ some axiomatization of pure set theory. In virtue of using $L_\in$ the way she does, the platonist thereby acquires access to the *interpreted* language $\mathcal{L}_\in^P$. If our platonist is of a particular ilk (e.g., Rayo, 2009), she will use $\mathcal{L}_\in^P$ in such a way as to make all sentences of $T_\in$ *trivially* true. At this stage, one cannot distinguish her from the nominalist, since the nominalist can arguably interpret $L_\in$ in a "deflationary" way, making every sentence of $T_\in$ equivalent to a tautology.[18] Suppose this latter use of $L_\in$ corresponds to the interpreted language $\mathcal{L}_\in^N$. What really distinguishes the platonist from the nominalist—or at least one of the things that does—is that *if* the platonist succeeds in using $L_\in$ as she intends to, then $\mathcal{L}_\in^P$ has a possibility of expansion that is able to express the truth conditions of (1), whereas $\mathcal{L}_\in^N$ lacks this possibility. Notice that this account explains the difference between the platonist and the nominalist in terms of the use of the original languages, and the ability to express (1) is part of the *explanandum* instead of the *explanans*. Also, it is clear why we should care about the variants considered by Hirsch: they are accessible from the original theories and thus realize the theories' potential to represent the world.

It should also be clear how to extend this account to other cases. Indeed, there is a sense in which the possibilities of expansion of each theory are *prefigured* in the way we use the theory. For instance, platonists about mathematical entities like numbers (or sets) can expand their theories to include propositions about the number (or the set) of characters in this document; nihilists or universalists about composition can expand their minimal theories to talk about objects that (fail to) compose or are the result of composition, like mereological atoms and chairs; endurantists and perdurantists can expand their minimal theories to talk about particular objects that either endure the passage of time or extend through it, and so on. Even more, this phenomenon is not limited to logical, mathematical, or metaphysical theories, but extends to theories in physics. As Wallace (2020) points out, classical and quantum mechanics are not "fixed" theories, but rather *frameworks* that can *instantiate* different concrete theories. Naturally, one way to understand this instantiation relation is as a kind of expansion of the framework theory into a concrete theory.

[18] Notice that this holds for the *original* theories, and not for Hirschean reinterpretations of them, which makes the problem even more acute.

In the next section, I will provide a framework into which these considerations can be formalized. First, I will motivate the framework by considering a simple example in which part of the expansion potential of a theory is represented. Next, I will show how the framework can be used to provide the explanations sketched above.

4 Expansion

On the face of it, the idea that the use of a theory's expressions constrains the interpretation of its expansions might strike one as odd. After all, the expansions of a theory are distinct from the theory itself, and they can be used in ways that are not determined by the original theory. To motivate the idea that the expansions of a theory are determined by the use of the theory's expressions, I will consider a simple example. Suppose a community κ uses a first-order language L with an expression '†' in the following way:

(U†E) Anyone accepting $\ulcorner\phi \dagger \psi\urcorner$ is committed to accepting both ϕ and ψ.

(U†I) Anyone accepting both ϕ and ψ is committed to accepting $\ulcorner\phi \dagger \psi\urcorner$.

Let's call these *use facts*. These are not (mainly) about the content of the sentences, but about the *use* of the expressions in them—where 'use' is being used in a very broad sense—, and they require L to have a certain syntactical structure. In particular, they require L to have a two-place sentential connective '†'. But more interestingly, one can see how these facts constrain which assignments of L are *compatible* with κ's use of it in content-based approaches. The problem of deriving constraints from use-facts is a thorny issue.[19] However, in this case it is reasonable to think that the following conditions are required for an assignment I of L to be compatible with κ's use of L:

(I†E) For any sentences ϕ and ψ of L, $I(\ulcorner\phi \dagger \psi\urcorner)$ entails both $I(\phi)$ and $I(\psi)$.

(I†I) For any sentences ϕ and ψ of L, any content entailing both $I(\phi)$ and $I(\psi)$ also entails $I(\ulcorner\phi \dagger \psi\urcorner)$.

Here, *entailment* in (I†E) and (I†I) is a relation between contents, and not sentences, and it will be construed differently depending on the theory of content one adopts. These latter constraints are not use facts, but rather *depend* on them. In particular, (I†E) depends on (U†E), and (I†I) depends on (U†I). Together, these constraints fully determine that '†' functions *as conjunction* in L,

[19] There is significant literature on different aspects of this topic. See Lewis (1975), Skyrms (2010) and Hlobil and Brandom (2025) for different perspectives on it.

regardless of how we choose to describe this role.[20] In this sense, (U†E) and (U†I) jointly determine the meaning of '†' in L as used by κ.

Thus, the framework pushes one to some kind of semantic holism—since the meaning of expressions like '†' is determined by restrictions over *entire* assignments like (I†E) and (I†I). However, two things are worth noting. First, that the framework is perfectly capable of specifying "atomistic" restrictions—e.g. the restriction that $I(A) = c$ for any compatible assignment I, where A is an atomic sentence of the language and c a particular content. Second, that the framework goes beyond particular restrictions like (I†E) and (I†I). Indeed, (U†E) and (U†I) don't only constrain the interpretation of L. They also constrain the ways in which L can be embedded in other languages.[21] For suppose that κ decides to extend L into a larger language L^+ containing some new sentence ξ. Naturally, κ will want to use L^+ in such a way that it "goes on as before" with respect to L.[22] This means that κ will want to use '†' in L^+ in such a way that the analogues of the constraints (I†E) and (I†I) hold for L^+. Notice, however, that (I†E) and (I†I) concern only the sentences of L, and thus entail nothing about how an assignment I^+ of L^+ should behave with respect to the new sentence ξ. This means that the constraints on I^+ will be underdetermined by the constraints on I. Regardless of this, it is clear that (U†E) and (U†I) will constrain the way in which κ can extend L into L^+, and that the constraints on I^+ will be affected by these use facts.

This example also shows that so-called "purely formal" accounts are not entirely divorced from interpretive practice. One could consider constraints on the consequence relations of L and L^+, determined by relevant use-facts, that are analogous to (I†E) and (I†I). These constraints are often treated as part of the formal structure of the language—for example, via syntactically specified inference rules governing logical constants like '$\wedge$'. But what makes these rules correctly describe an interpreted language is not the syntax alone. The fact that '$\wedge$' is treated as conjunction is not fixed by its being a two-place connective, or by a *merely* syntactical relation $\vdash$. It is fixed by what $\vdash$ *tracks*: that, for the users of the language, commitment to a set Γ of sentences such that $\Gamma \vdash \phi$ involves commitment to ϕ. So formal features like consequence relations already encode certain facts about use. This casts doubt on critiques of formal approaches based on "trivial semantic conventionality" (e.g., Teitel 2021): while formal accounts abstract away from content-assignments, they still presuppose norms of usage, and these norms constrain interpretation.

[20] We might describe this role using a function like $F_\wedge$ that takes two contents and returns the weakest content entailing both; we would then need to say that $F_\wedge$ is the "semantic value" of '†'. But nothing so far precludes us from assigning '†' some *other* semantic value while preserving the constraints (I†E) and (I†I), by changing the semantic values of other parts of the language. What matters here is capturing the *overall semantic behavior* of '†', not its *semantic value*, and the point generalizes to other expressions, both logical and non-logical.

[21] Thus, in a sense, it goes beyond Quine-style holism, considering also every bigger language that our language can be embedded in. The difference will be rendered salient in Sections 5.3 and 5.4.

[22] The connection to the rule-following problems in Kripke (1982) is evident. The present paper doesn't require a commitment to any a particular solution of them, but only to the fact that in particular instances of expansion *we do know* how to go on as before, thus rejecting Kripkenstein's "skeptical solution".

This fact might drive proponents of formal approaches to think that contents are not needed: after all, perhaps all constrains on interpretation given by use can be encoded by formal properties of the languages involved. Indeed, the constraints (I†E) and (I†I) could be captured equally well by a syntactically specified consequence relation $\vdash$ over L. But our languages and theories also contain expressions whose usage cannot be specified merely by relating contents to other contents (like 'chair').[23] Thus, consequence relations, model-theoretic semantics, or other formal properties of languages are not enough to capture constraints on interpretation in general.[24] As opposed to these, the current approach deals with "logical" and "non-logical" constraints *in exactly the same way*: both are determined by the use of the relevant expressions, and both determine a set of admissible assignments.

Proponents of content-based approaches might have a different reason for suspicion. In many cases, we can obtain the relevant restrictions in content-based approaches by assigning semantic values to *parts* of a sentence and stipulating how these parts are combined to yield contents. For instance, if 'F' in a first-order language is used as we use 'chair' in English, then content-based approaches like possible-world semantics can assign the semantic value $[\![\text{`}F\text{'}]\!]^w$ of '$F$' at world $w$ to be the set of chairs at $w$. We could then set restrictions on assignments using these sub-sentential semantic values. For example, we could stipulate that $[\![\text{`}Fa\text{'}]\!]^w = 1$ iff $[\![\text{`}F\text{'}]\!]^w$ contains the object denoted by '$a$' at $w$, and 0 otherwise, and have the content $I(\text{`}Fa\text{'})$ be $\{w : [\![\text{`}Fa\text{'}]\!]^w = 1\}$, and we could have analogous conditions for more complex sentences containing '$F$'. The case of '†' can be similarly handled by assigning it a semantic value $[\![\text{`}\dagger\text{'}]\!]^w = \mathsf{min}$ in every world w, and defining $[\![\text{`}\phi \dagger \psi\text{'}]\!]^w$ as $[\![\text{`}\dagger\text{'}]\!]^w([\![\text{`}\phi\text{'}]\!]^w, [\![\text{`}\psi\text{'}]\!]^w)$, where min is a function that returns the minimum of its arguments. This way, we can obtain the relevant restrictions on assignments by assigning semantic values, stipulating how these values combine to yield contents, and set the restriction that expansions must assign the same values and combine them in the same way. So why not simply use this strategy to obtain these restrictions?

One reason is that the specification of such restrictions is mainly a semanticist's job, and for the present purposes I only need to assume that such restrictions *exist*.[25] Furthermore, it is not true even in these semantic approaches that restrictions on assignments come exclusively from assignments of sub-sentential semantic values. Indeed, expressions like '$\exists x$' are treated as *syncategorematic* in many

[23]Whether this can be specified by "empirical content" or not depends on whether one can specify these situations empirically or whether more than just empirical information to specify the relevant situations. The current framework is not committed either view. Thus, it works with a general notion of content, not of *empirical* content.

[24]Of course, we could use a semantics in which 'F' is interpreted as the set of chairs, but most model-theoretic semantics are defined over a universe of pure sets (Halvorson, 2019, p. 22), and even in the case they don't explicitly limit themselves to such a universe, they make no use of any particular property of the objects in the domain (Manzano, 1999).

[25]Arguably, proponents of content-based approaches also only need to assume such restrictions exist, so the view isn't less committal than content-based approaches. Nevertheless, it achieves more by operating with the restrictions themselves, rather than with the mappings they generate.

contexts, and their "meaning" is specified by specifying the semantic values of sentences containing them (Heim and Kratzer, 1998).

But these are not the only reasons to avoid committing to the thesis that all restrictions are obtained via assignments of sub-sentential semantic values. For one, it is an open question whether theories that 'refer to different things' or 'have a different ontology' at the object-property level could nevertheless be metaphysically equivalent. Warren (2015), Hirsch (2002) and Barrett and Halvorson (2017) seem to think that this is a genuine possibility, while Sider (2009) seems to disagree. For another, the strategy of using sub-sentential semantic values might open the door to fine-grained notions of content, with all the costs that this entails as discussed in Section 3. The current approach avoids these issues altogether by sidestepping the assignment of sub-sentential semantic values.

4.1 The Framework

The resulting picture follows content-based approaches in that both consider languages as uninterpreted syntax plus *assignments*—mappings from sentences to contents—, but differs from most content-based approaches in that it considers an interpreted language as a *condition* defined over assignments of the language and its syntactic expansions, rather than as a pair of a language and an assignment.[26] More formally, we can define an *interpreted language* as follows:

> ***Interpreted Language***. An *interpreted language* $\mathcal{L}$ is a condition defined over assignments of some (uninterpreted) language L and its syntactic expansions, such that
>
> (i) at least one assignment of L satisfies $\mathcal{L}$, and
>
> (ii) for every assignment I' of some syntactic expansion L' of L, if I' satisfies $\mathcal{L}$, then the restriction of I' to L (denoted $I'|_L$) satisfies $\mathcal{L}$.
>
> We will say that L is the *base* of $\mathcal{L}$, and that an assignment I of L (or some syntactic expansion of it) is *$\mathcal{L}$-compatible* or that it is a *point* of $\mathcal{L}$ just in case it satisfies $\mathcal{L}$.

I will not assume anything in particular about the nature of the contents except the structure of a complete lattice with generalized disjunction and conjunction operations, $\bigvee$ and $\bigwedge$ respectively, an absurd content $\bot$, and a trivial content $\top$.[27] One can think of $\mathcal{L}$ as encoding the interpretive

[26] This framework closely resembles one developed in Przelecki (1969). Although it differs in that the current framework (i) is not only compatible with first-order languages but any language containing sentences, (ii) doesn't require the base language to be an empirical language and honors the Quinean idea of not separating "empirical" from "non-empirical" content, and (iii) works with contents as sketched in footnote 27, instead of the model-theoretic notion of model (or set of models). Thanks to an anonymous reviewer for pointing out this similarity.

[27] A complete lattice is a set C with a partial order $\leq$ on it, where every subset $S \subseteq C$ has a least upper bound (denoted $\bigvee S$) and a greatest lower bound (denoted $\bigwedge S$) with respect to $\leq$. In this case, the elements of the lattice are

constraints of some community's use of L. The first condition in ***Interpreted Language*** ensures that the community's use is compatible with at least one assignment of L, and the second condition ensures that expansions of the language used by the community are also compatible with the community's current use. Examples like the case of '†' can be accommodated by this definition by picking a condition $\mathcal{L}$ that requires every assignment of an expansion L' to satisfy the analogues of (I†E) and (I†I) for L'.

We also can recover the original representation of interpreted languages as pairs of a language and an assignment if we assume that there is exactly one assignment I of the base language L satisfying $\mathcal{L}$. But this assumption is not necessary for the general account: we can consider cases where more than one assignment of the base language is compatible. The possibility of multiple assignments for the base language is not only a feature of the present account, but it is independently motivated by many semantic approaches to phenomena like vagueness (Fine, 1975), indeterminacy of reference (Van Fraassen, 1966) and self-referential paradoxes (Field, forthcoming). It is useful to think of $\mathcal{L}$ as an *inverted cone*, as in Figure 1.

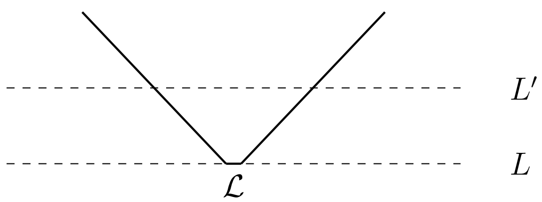

Figure 1: Representation of an interpreted language $\mathcal{L}$.

Here, difference in the vertical dimension represents different languages—with higher points representing syntactic expansions—and difference in the horizontal dimension represents different assignments of the same languages. Points in the dashed line represent assignments of L', some of which are $\mathcal{L}$-compatible (inside the cone), and some of which are not (outside the cone).

Of course, there are limits to diagrammatic representations like the one above. For one, the structure of syntactic expansions cannot be represented by a single axis, since it resembles more that of partially ordered sets with incomparable elements than that of a linear order. However, the diagrammatic representation captures an important fact: that there are relatively few assignments at the base, and there are more degrees of freedom as we expand the language. This is due to the fact that condition (ii) in ***Interpreted Language*** constrains the assignments only in terms of the contents that they assign to sentences at the base—it states that expansions might not be free to reinterpret sentences at the base, but they might be free to interpret new sentences, to an extent.

the contents, and the ordering relation is the relation of *entailment*, the least upper bound is (infinitary) disjunction, and the greatest lower bound is (infinitary) conjunction. Additionally, we define the *trivial* content as $\top := \bigvee C$, which is entailed by every other content, and the *absurd* content as the $\bot := \bigwedge C$, which entails every other content.

When we expand a language $\mathcal{L}$, we first syntactically expand its base into a larger language, and then choose among the $\mathcal{L}$-points that interpret this language and its syntactic expansions. This suggest the following definition of expansion:

> ***Language Expansion.*** An interpreted language $\mathcal{L}_2$ is an *expansion* of another interpreted language $\mathcal{L}_1$ just in case every point of $\mathcal{L}_2$ is a point of $\mathcal{L}_1$.

Geometrically, this means that the cone of $\mathcal{L}_2$ is contained in the cone of $\mathcal{L}_1$, as represented as in Figure 2, where the horizontal lines correspond to the bases L_1 and L_2 and to a language L' that syntactically expands both.

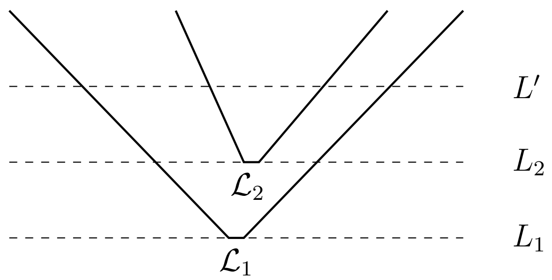

Figure 2: Representation of the expansion $\mathcal{L}_2$ of $\mathcal{L}_1$.

By choosing among the points of $\mathcal{L}_1$ to produce $\mathcal{L}_2$, we discard some assignments of L_2 and its syntactic expansions. When we do so, it is useful to "retain" our ability to interpret the base language L_1 in the same way as before. This is captured by the following definition:

> ***Preservation.*** An expansion $\mathcal{L}_2$ of $\mathcal{L}_1$ *preserves* an $\mathcal{L}_1$-point I_1 just in case there is an $\mathcal{L}_2$-point I_2 whose restriction to I_1's domain is I_1 itself. We will say that the expansion $\mathcal{L}_2$ is *lossless* just in case it preserves every $\mathcal{L}_1$-compatible assignment of $\mathcal{L}_1$'s base.[28]

Here, $\mathcal{L}_2$ "preserves" I_1 not because it contains I_1 itself, but rather because it contains an assignment from which I_1 can be *recovered* via domain restriction. Thus, in choosing to expand $\mathcal{L}_1$ to $\mathcal{L}_2$, we don't necessarily lose the possibility of interpreting the base language using I_1. But this isn't to say that we don't lose anything in a lossless expansion. In fact, in choosing a particular cone for $\mathcal{L}_2$'s base language L_2, we lose some of the ways in which we *could have extended* the base language,

[28] Losslessness is related to, but distinct from, the logical notion of a *conservative extension*. For one, conservative extensions are defined between theory-language pairs $\langle T, L\rangle$, where L is an *uninterpreted* language. On the other hand, losslessness is defined between *interpreted languages*, without reference to any set of sentences. But given a notion of an *interpreted* theory $\langle T, \mathcal{L}\rangle$, we can define a corresponding notion of a *conservative expansion* that (i) entails the formal/syntactic notion, and (ii) can be shown to hold between pairs $\langle T, \mathcal{L}\rangle$ and $\langle T, \mathcal{L}^+\rangle$ when $\mathcal{L}^+$ is a lossless expansion of $\mathcal{L}$. The definition is simple: if $\mathcal{L}'$ expands $\mathcal{L}$, then $\langle T', \mathcal{L}'\rangle$ is a conservative expansion of $\langle T, \mathcal{L}\rangle$ iff for any point I' of $\mathcal{L}'$, $\bigwedge_{\phi\in T} I'|_L(\phi) = \bigwedge_{\psi\in T'} I'(\psi)$.

and we preserve others. This means that there is a trade-off between *expressiveness* and *flexibility* whenever we expand a language.

Using this notion of an interpreted language and of language expansion, we can define a notion of interpreted theory and of theory expansion.

Interpreted Theory. An *interpreted theory* is a tuple $\langle T, \mathcal{L}\rangle$, where $\mathcal{L}$ is an interpreted language and T is a set of sentences in $\mathcal{L}$'s base.

Theory Expansion. A theory $\langle T_2, \mathcal{L}_2\rangle$ is an *expansion* of another theory $\langle T_1, \mathcal{L}_1\rangle$ iff $\mathcal{L}_2$ is an expansion of $\mathcal{L}_1$, and $T_1 \subseteq T_2$.

Assuming lossless expansions, $\langle T_1, \mathcal{L}_1\rangle$ being expanded by $\langle T_2, \mathcal{L}_2\rangle$ ensures that the content of the sentences of T_1 is preserved in T_2, and thus that T_2 "says as much about the world" as T_1. More precisely, we have the following theorem, which is proved in the Appendix:

Theorem 1. *Let $\langle T_1, \mathcal{L}_1\rangle$ and $\langle T_2, \mathcal{L}_2\rangle$ be interpreted theories such that $\langle T_2, \mathcal{L}_2\rangle$ is an expansion of $\langle T_1, \mathcal{L}_1\rangle$. If $\mathcal{L}_2$ is a lossless expansion of $\mathcal{L}_1$, then for every $\mathcal{L}_1$-compatible assignment I_1 of $\mathcal{L}_1$'s base, there is an $\mathcal{L}_2$-compatible assignment I_2 of $\mathcal{L}_2$'s base such that $\bigwedge_{\phi\in T_2} I_2(\phi)$ entails $\bigwedge_{\psi\in T_1} I_1(\psi)$.*

5 Using the Framework

In this section, I show how the current framework can be used to explain the equivalence or inequivalence of formally identical theories in formal approaches (discussed in Section 2), as well as to distinguish theories that are equivalent under coarse-grained content-based approaches (discussed in Section 3). Thus, it complements and improves upon its counterparts.

5.1 Expansion in Formal Approaches

Recall that in Section 2, Dewar justifies the equivalence between two notational variants M_ρ and M_μ of Maxwell's equations by embedding the languages $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$ in a joint language $\mathcal{L}^+$, and showing that the sentences of $\mathcal{L}_\rho$ are inter-derivable with their counterparts in $\mathcal{L}_\mu$. The present framework can be used to explain this equivalence.

More precisely, let R be the relation two sentences ϕ and ψ in the joint language $\mathcal{L}^+$ satisfy just in case one of them is the result of replacing some instances of 'ρ' with 'μ' or vice-versa in the other. Dewar's explanation depends on the fact that sentences bearing R in $\mathcal{L}^+$ are inter-derivable. But, as I have argued, this fact is not explained by the mere existence a joint language that supports said derivations, since there are many ways to embed the languages in which the required derivations are not supported. For instance, one can consider a joint language in which the sentence '$\rho \neq \mu$' is an axiom. On the other hand, the present framework can be used to explain this inter-derivability.[29] What we need, is to specify some of the compatibility conditions of the languages. In particular, we need to specify the following:

(5.1.i) Any assignment which is either $\mathcal{L}_\rho$-compatible or $\mathcal{L}_\mu$-compatible must respect first-order entailment.[30]

(5.1.ii) $\mathcal{L}^+$ is a lossless expansion of $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$.

(5.1.iii) Any assignment I which is both $\mathcal{L}_\rho$-compatible and $\mathcal{L}_\mu$-compatible must be such that, if $R(\phi, \psi)$, then $I(\phi) = I(\psi)$.[31]

Condition (5.1.i) is a plausible condition for the languages $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$ assuming these are at least as expressive as first-order logic. Condition (5.1.ii) is a plausible assumption for the joint language, since it is obtained by expanding the languages $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$ without discarding assignments. Condition (5.1.iii) is obtained by Dewar's stipulation that 'ρ' and 'μ' are used for charge density in the original languages—since they are used *in the same way*, the result of replacing 'ρ' with 'μ' or vice-versa should not change the content of the sentences under compatible assignments. It is then straightforward to show that, under these conditions, M_ρ and M_μ are equivalent in the joint language, and even more, that they are equivalent to the theories formulated in the *original* languages. This entails, in turn, that the original theories are equivalent, as Dewar claims. The result is proved in the Appendix.

It may be useful to consider this explanation alongside Dewar's. In Dewar (2023), the starting point is the fact that we have inter-derivability in the joint language $\mathcal{L}^+$, and the explanation of this inter-derivability is *implicitly* attributed to use.[32] In the current framework, the inter-derivability is *explicitly* obtained as a consequence of assumptions (5.1.i), (5.1.ii), and (5.1.iii), which encode interpretive constraints determined by the use of the languages.

[29] Admittedly, content-based approaches can also explain this case by saying that the theories are equivalent because they are mapped to the same contents. However, the explanation given by the present framework is more comprehensive, since it ties the equivalence to the use of the languages.

[30] An assignment I of a language L respects first-order entailment iff the content $\bigwedge_{\psi \in \Gamma} I(\psi)$ entails the content $I(\phi)$ whenever ϕ is a first-order consequence of Γ.

[31] An account that uses fine-grained contents might want to replace this latter condition by saying that $I(\phi)$ and $I(\psi)$ are *mutually entailing*. This makes the explanation compatible with fine-grained content-based approaches, but it is not necessary for the present purposes.

[32] "Suppose that you and I both write down Maxwell's equations—but whereas I use ρ to indicate charge density, you use μ. It seems clear that we should judge the two theories to be equivalent." (Dewar, 2023, p. 16)

5.2 Definitional and Morita Extensions

The explanation of Dewar's case suggests that the present framework can be used to complement explanations in formal approaches to equivalence which rely on placing the theories in a common language. In particular, we can use the framework to specify conditions under which the syntactic expansions are relevant to the equivalence of theories. To see how, it is best to consider concrete criteria, such as *definitional equivalence* or *Morita equivalence* (Barrett and Halvorson, 2016b).

Seeing how the present framework interacts with these criteria does not require being too involved with the specific formal properties of either notion of equivalence, so a brief explanation of them will suffice. The reader is encouraged to consult Barrett and Halvorson (2016b) for more details. Both definitional and Morita equivalence are formulated in the context of *many-sorted first-order languages*, and a theory T in a language L—which we will denote $\langle T, L\rangle$—is simply a set of sentences of L. Briefly, a many-sorted first-order language is a language L built from a *signature* Σ of different primitive symbols: constants, predicates, function symbols, and *sorts*. Here, each constant is assigned a sort symbol σ, each n-place predicate symbol is assigned an "arity" $\sigma_1 \times \cdots \times \sigma_n$—indicating that it is combined with n-sequences of symbols of sorts $\sigma_1, \ldots, \sigma_n$ to form a sentence—, and each n-place function symbol is assigned an arity $\sigma_1 \times \cdots \times \sigma_n \to \sigma_0$—indicating that it is combined with n-sequences of symbols of sorts $\sigma_1, \ldots, \sigma_n$ to form a symbol of sort σ_0. We also have the identity predicate '$=$', and quantifiers $\ulcorner \exists_\sigma \urcorner$ and $\ulcorner \forall_\sigma \urcorner$ for each sort σ. From here, sentences are built much in the same way as in first-order logic, but ensuring that functions and predicates are applied to terms of the right sort.

In this context, an *extension* of a theory $\langle T, L\rangle$ is a theory $\langle T^+, L^+\rangle$, such that L^+'s signature extends L's, $T \subseteq T^+$, and the remaining sentences in $T^+ \backslash T$ are *definitions* of the new vocabulary in L^+ in terms of the vocabulary in L. Beyond this point, definitional and Morita extensions differ on which kind of vocabulary additions are allowed, with only Morita extensions allowing for the addition of new sorts, but this difference is not relevant for our purposes. Two theories are *definitionally* or *Morita equivalent* if and only if they have a common definitional or iterated Morita extension, respectively, up to logical equivalence. Barret and Halvorson consider definitionally or Morita equivalent theories as *saying the same* about the world. This is because definitional or Morita extensions "say no more" than the original theories that they extend (pp. 560, 565), and arguably, also "no less". Thus, the existence of a common extension $\langle T^+, L^+\rangle$ of two theories $\langle T_1, L_1\rangle$ and $\langle T_2, L_2\rangle$ should be taken as saying no more and no less than the original theories themselves, proving that they are equivalent.

But nothing in this reasoning depends on the *content* of these theories, and the fact that the Sklar example (or close analogues) can fit this mold gives us reasons for suspicion. In fact, considering a simple example with first-order analogues of 'lion' ('l') and 'tiger' ('t') will help find some gaps in this reasoning. Let L_1 be built from $\Sigma_1 = \{\sigma, l\}$ and L_2 from $\Sigma_2 = \{\sigma, t\}$, where 'σ' is a sort symbol and 'l','t' are predicates of arity σ.[33] Consider the theories $T_1 = \{\exists_\sigma x\, l(x)\}$ and $T_2 = \{\exists_\sigma x\, t(x)\}$ in L_1 and L_2, respectively. Now, consider the expansion L^+ of L_1 and L_2 built from $\Sigma^+ = \{\sigma, l, t\}$, and the theory $T^+ = \{\exists_\sigma x\, l(x), \forall_\sigma x(t(x) \leftrightarrow l(x))\}$ in L^+. Here, we have that $\langle T^+, L^+\rangle$ is a definitional extension of both $\langle T_1, L_1\rangle$ and $\langle T_2, L_2\rangle$ (up to logical equivalence). Therefore, $\langle T_1, L_1\rangle$ and $\langle T_2, L_2\rangle$ are definitionally equivalent (and thereby Morita equivalent).

Of course, these theories are *not* equivalent. One says that there is a lion, and the other that there is a tiger. So what is the problem? The problem, as diagnosed by the current approach, is that the joint language poses *incompatible constraints on content assignments*. To see this, we should think of L_1 and L_2 as the syntactic bases of *interpreted* languages $\mathcal{L}_1$ and $\mathcal{L}_2$ in which 'l' and 't' are used as we use 'lion' and 'tiger' in English, respectively. Now, suppose there is an expansion $\mathcal{L}^+$ of $\mathcal{L}_1$ and $\mathcal{L}_2$. From ***Language Expansion***, we know that every point I^+ of $\mathcal{L}^+$ is both $\mathcal{L}_1$- and $\mathcal{L}_2$-compatible. Since $\mathcal{L}_1$ demands that 'l' be interpreted as 'lion', it follows that every $\mathcal{L}_1$-compatible assignment assigns the proposition that there is a lion to '$\exists_\sigma x\, l(x)$'; and analogously, every $\mathcal{L}_2$-compatible assignment assigns the proposition that there is a tiger to '$\exists_\sigma x\, t(x)$'. Thus, I^+ maps T_1 to the proposition that there is a lion, and T_2 to the proposition that there is a tiger. Clearly, these propositions are not the same, and so the theories are not equivalent under I^+.

But notice that here we are evaluating T_1 and T_2 in $\mathcal{L}^+$, and not their expansion T^+. So what happens if we evaluate T^+? What happens is that T^+ will be found to be *stronger* than both T_1 and T_2, because the sentence δ_1 = '$\forall_\sigma x(t(x) \leftrightarrow l(x))$' will be assigned a *non-trivial* content in $\mathcal{L}^+$—namely, that everything is a tiger if and only if it is a lion, a proposition that is not true in the actual world. The mistake here was to assume that taking δ_1 as a definition of 't' is an *innocent* move: 't' already occurs in the opposite interpreted theory, $\langle T_2, \mathcal{L}_2\rangle$, and the interpretative constraints of $\mathcal{L}_2$ demand that 't' be taken as synonymous with 'tiger', which is *incompatible* with taking 'l' as synonymous with 'lion' and δ_1 as a definition of 't'. More generally, the current analysis reveals that there is a tension between (i) interpreting the original vocabulary in the expanded language in the same way as in the original languages, and (ii) taking the 'definitional' sentences in the expanded language as definitions of the new vocabulary. These conditions can be formalized as follows, for any common expansion $\langle T^+, \mathcal{L}^+\rangle$ of a pair of interpreted theories $\langle T_1, \mathcal{L}_1\rangle$ and $\langle T_2, \mathcal{L}_2\rangle$:

(5.2.i) $\mathcal{L}^+$ is a lossless expansion of $\mathcal{L}_1$ and $\mathcal{L}_2$.

(5.2.ii) Every point I^+ of $\mathcal{L}^+$ assigns a *trivial* content to any sentence in $T^+ \backslash T_1$ and $T^+ \backslash T_2$.

Fortunately, we can show that the conditions above, together with respecting many-sorted first-order entailment, are sufficient to guarantee equivalence of the original theories. This result is

[33] I am taking 'σ' to be a shared symbol between Σ_1 and Σ_2 for simplicity. The example can be recreated with completely disjoint signatures, but this would make the discussion more cumbersome. Also, I am omitting quotations in sentences inside set-builder notation for simplicity.

proved in the Appendix. Notice that these constraints are not tangential to the formal approaches developed in Barrett and Halvorson (2016b), but rather are *necessary* components of any sound chain of reasoning from definitional or Morita equivalence to real equivalence—sentences like δ_1 are called "explicit definitions" for a reason (p. 560). The current framework can be used to account for this reasoning, and to show that it poses substantive constraints on equivalence.

5.3 Expansions in Content-Based Approaches

The present framework can also be used to explain the distinction between some theories that are equivalent in coarse-grained content-based approaches, but that are widely considered to be distinct. Recall the case of the debate between the platonist and the nominalist about the existence of abstract entities, as described in Hirsch (2009). This can be regimented in Quinean terms by saying that the platonist is *committed* to the existence of abstract entities while the nominalist is *not*, which suggests that the nominalist's theory should be *weaker* than the platonist's, and thus that the theories should not be equivalent.

However, we can find settings in which the theories would be (wrongly) considered as equivalent by coarse-grained content-based approaches. To do so in the present framework, let's consider how the nominalist and the platonist would interpret set-theoretic statements in a(n uninterpreted) language $L_\in$ containing only set-theoretic and logical vocabulary. While the platonist about sets would have no trouble directly specifying the truth-conditions for sentences in $L_\in$, the nominalist would have to specify these truth-conditions using paraphrases that do not involve appeal to a "set-theoretic universe". For instance, if we assume that $L_\in$ contains bracket notation, the nominalist might paraphrase '$x \in \{y, z\}$' as '$x = y \vee x = z$'. In the current framework, this amounts to demanding that admissible assignments I satisfy $I(`x \in \{y, z\}') = I(`x = y \vee x = z')$. We can then consider the interpreted nominalist language $\mathcal{L}_\in^N$ as the result of imposing all constraints on assignments of $L_\in$ that derive from the nominalist's use of the language, and analogously for the interpreted platonist language $\mathcal{L}_\in^P$.

Now, assume that the platonist is right, and that pure sets necessarily exist, with their intrinsic features being necessary. Thus, any set of axioms $T_\in$ in $L_\in$ which "describes" the realm of sets should be taken as containing only *trivial* sentences if interpreted in $\mathcal{L}_\in^P$.[34] Furthermore, suppose our nominalist has succeeded in paraphrasing all sentences in $T_\in$ in her language $\mathcal{L}_\in^N$, so that they are *also* trivially true. At this point, the interpreted theories $\langle T_\in^N, \mathcal{L}_\in^N \rangle$ and $\langle T_\in^P, \mathcal{L}_\in^P \rangle$ should be considered equivalent by coarse-grained content-based criteria along the lines of Section 3, since

[34] Here, I am using coarse-grained contents, for the reasons stated in Section 3.

they are both (i) trivially true, and (ii) formulated in background languages able to express only trivially true or trivially false sentences.[35]

Of course, this verdict is unintuitive, and arguably wrong. Evidence of this is that Hirsch himself tries to avoid it by appealing to the content of *mixed* mathematical statements like (1). But notice that this explanation is not available to proponents of standard content-based approaches, since it concerns *expansions* of the languages $\mathcal{L}_\in^P$ and $\mathcal{L}_\in^N$, and not the languages themselves, which can express nothing about angels or love. On the other hand, the present framework allows us to make sense of this explanation. Suppose, again, that the platonist is right. Given this, we can take $\mathcal{L}_\in^P$ to be a strictly stronger condition on assignments of $L_\in$ than $\mathcal{L}_\in^N$, since the platonist has more resources to articulate constraints on assignments. Thus, in the current sense, these two theories would be inequivalent, since they allow for different assignments of $L_\in$ to be admissible.

The current picture can also account for Hirsch's purported explanation of the inequivalence. Let $L_{A,L}$ be a minimal (first-order) language containing (analogues of) the predicates 'angel' and 'loves', and $\mathcal{L}_{A,L}$ be the condition on assignments of $L_{A,L}$ and its expansions that results from interpreting these predicates as we do in English. We can then take a minimal syntactic expansion L^+ of both $L_{A,L}$ and $L_\in$, and consider the interpreted languages $\mathcal{L}_\in^{P+}$ and $\mathcal{L}_\in^{N+}$ as the result of imposing the constraints of jointly satisfying either $\mathcal{L}_\in^P$ and $\mathcal{L}_{A,L}$, or $\mathcal{L}_\in^N$ and $\mathcal{L}_{A,L}$, in each case. The situation can be represented as in Figure 3, where $\mathcal{L}_\in$ can be taken as $\mathcal{L}_\in^P$ or $\mathcal{L}_\in^N$ in each case, and the new cone represents the corresponding expanded language. Now, since the platonist has more resources at her disposal, $\mathcal{L}_\in^P$ is constrained enough that $\mathcal{L}_\in^{P+}$ already rules out pairs of assignments which disagree on which coarse-grained content to assign (1). On the other hand, since the nominalist has less resources, $\mathcal{L}_\in^N$ is not strong enough to allow $\mathcal{L}_\in^{N+}$ to rule out pairs of assignments which disagree on which coarse-grained content to assign (1) (or an analogue).[36] This means that, even though the original languages $\mathcal{L}_\in^P$ and $\mathcal{L}_\in^N$ could be deemed as expressively equivalent, their *potential* for representing the world is different, since the platonist's theory has expressive possibilities that the nominalist's theory does not, under this particular expansion.[37]

It is clear that $\mathcal{L}_{A,L}$ is not special, and that the same reasoning can be applied to other ways of expanding the languages. But it is worth noting that not *every* expansion should count. For instance,

[35]Here, I am assuming that facts purely about pure sets hold of necessity. This, $\mathcal{L}_\in^P$ should only be able to express trivial truths or falsehoods, and provided that the nominalistic paraphrase succeeds, the same should hold for $\mathcal{L}_\in^N$.

[36]Formally, this means that there is a sentence $\phi_{(1)}$ in L^+ such that any $\mathcal{L}_\in^{P+}$-point assigns $\phi_{(1)}$ the content we assign to (1) in English, but no such sentence exists for $\mathcal{L}_\in^{N+}$. This happens not because $\mathcal{L}_\in^{N+}$'s syntactic base doesn't have "enough" sentences (indeed, it has the same as $\mathcal{L}_\in^{P+}$), but because $\mathcal{L}_\in^{N+}$ doesn't constrain content assignments enough that they would all assign the same content to sentences like $\phi_{(1)}$.

[37]The attentive reader could note, again, that the nominalist should also be able to express (1) if her background logic is second-order. This is true, but does not entail that second-order nominalism and second-order platonism are equivalent, as discussed in footnote 16: while (1) is not a difference-maker for these positions, there might be other sentences that only the latter is able to express. Also, notice that *reformulating a theory in second-order logic* might be a way of expanding it, so when we consider platonism vs. nominalism in a first-order formulation that is *open to expansions*, we might also be explicitly considering second-order reformulations (if they are indeed reformulations).

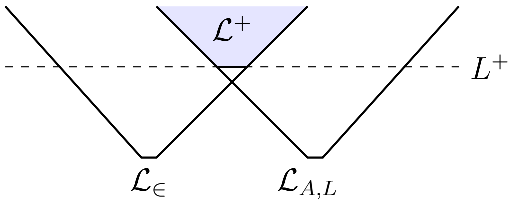

Figure 3: Expansion of set-theoretic languages by non-set-theoretic vocabulary.

we shouldn't count expansions of the nominalist's theory that simply consist in adding set-like vocabulary interpreted as the platonist would do. For example, claiming that the nominalist has access to relevant possibilities by expanding her language with a predicate $\in^*$ ("schmembership") such that '$x \in^* y$' is interpreted as the platonist would interpret '$x \in y$', will not do. This is because, just as we assumed the nominalist has no resources to constrain the interpretation of $L_\in$ to make the content of (1) determinate, we can assume that she also has no resources to constrain the interpretation of $\in^*$ in the required way.

Thus, we need independent considerations of which expansions are relevant for the explanation to work, and the explanation cannot be straightforwardly generalized comparisons of equivalence. But this is a feature, not a bug, of the present framework, since it diverts attention to how the original theories are *used* and *understood* by the theorists involved.

5.4 Further Applications

While a general criterion to handle expansions in content-based approaches is not provided—and indeed, since considerations about use are necessary, it is unlikely that such criterion can be provided—the framework can be used to explain other cases mentioned in Section 3 where expansions are relevant to the equivalence of theories.

For instance, we can explain why, even under a coarse-grained "platonist" approach, the Peano Axioms for arithmetic (PA) are considered to be weaker than Zermelo-Fraenkel set theory (ZF). Suppose there are pure (interpreted) languages $\mathcal{L}_\mathbb{N}$ and $\mathcal{L}_\in$ in which PA and ZF are formulated, respectively. In the usual axiomatizations, $\mathcal{L}_\mathbb{N}$ contains only the non-logical expressions '0' (constant, to be read as 'zero') and 'S' (2-place predicate, to be read as 'immediately succeeds'), while $\mathcal{L}_\in$ contains only the non-logical predicate '$\in$' (2-place, to be read as 'is a member of'). If the PA and the ZF axioms are taken to be *necessarily* true, then the theories $\mathbf{PA} = \langle \text{PA}, \mathcal{L}_\mathbb{N} \rangle$ and $\mathbf{ZF} = \langle \text{ZF}, \mathcal{L}_\in \rangle$ should be considered equivalent by coarse-grained content-based approaches. But the ZF axioms are more expressive than the PA axioms by many 'formal' criteria, such as the fact that the consistency of PA can be proved in ZF, but not the other way around.

In the current framework, the inequivalence can be captured by the fact that **ZF** and **PA** have different potentials to describe possibilities. For instance, it is already implicit in our use of arithmetic language $\mathcal{L}_{\mathbb{N}}$, that we can equip it with extra vocabulary for *counting* things—and thus add extra non-mathematical vocabulary plus the symbol '#', where '$\#x\Phi(x)$' is to be understood as an expression denoting the number of values of x satisfying $\Phi(x)$. It is also already implicit in our use of the set-theoretic language $\mathcal{L}_{\in}$ that we can equip it with extra vocabulary for talking about *urelements* (non-sets). Thus, we can expand $\mathcal{L}_{\in}$ with a new predicate '$\mathcal{A}$', where '$\mathcal{A}x$' is to be understood as 'x is a urelement'. If we do so, we will find that, *if expanded in the same way*, **PA** and **ZF** have different potentials for representing possibilities. For instance, if we add a contingent predicate 'F' to both theories, we will be able to count the Fs in **PA**, to express that there are finitely many Fs—e.g. by the sentence '$\exists x(Nx \wedge \#yFy = x)$'—, and so on. **ZF** has these possibilities, but it can also express more. For example, by expanding **ZF** with '$\mathcal{A}$' and the predicate 'F', we can express that there are exactly κ Fs, for any cardinal κ which can be singled out with set-theoretic language. These possibilities are not available to **PA** without *also* using set-theoretic or analogous vocabulary, which points to the inequivalence of the theories *purely in terms of coarse-grained content*.

I take it that examples like these at least point towards the possibility of using the present framework to explain differences in other coarse-grained equivalent theories mentioned in Section 3. Again, what exactly counts as a relevant expansion in these cases will not be a straightforward matter, and it will depend on the specific context in which the theories are being compared. But the framework points to the kind of considerations that are relevant to the question of equivalence, and how these considerations can be used to explain why certain theories are considered equivalent or not.

6 Conclusion

The concept of *expansion potential* introduced in this paper fulfills two functions. First, it narrows the gap between formal and content-based approaches to theoretical equivalence by providing a framework in which formal and content-based considerations can be integrated. Second, it provides a way to justify strategies used by proponents in both approaches when the relevant approach delivers an unexpected verdict. It is also worth noting that the framework can be implemented using a reasonably coarse-grained (and even 'deflationary') notion of content, and so makes few assumptions about the metaphysical structure of the world.

Moreover, the framework sheds light on why certain debates in philosophy persist despite apparent equivalences, and how one might argue for the equivalence or inequivalence of certain positions based on the way they can be expanded. Indeed, arguments like these already exist in the literature—e.g. Uzquiano (2004)—, although they are not always recognized as such. The framework also diverts attention to the way theories are used and understood by the theorists involved, which can be crucial for understanding the significance of the theories themselves. It suggests that the potential for a theory to be embedded in larger languages is not merely a technical consideration but is fundamental to understanding its role in our conceptual framework.

References

Balaguer, Mark (2021). *Metaphysics, Sophistry, and Illusion: Toward a Widespread Non-Factualism*. First edition. Oxford: Oxford University Press. 295 pp.

Barrett, Thomas William and Hans Halvorson (2016a). "Glymour and Quine on Theoretical Equivalence". In: *Journal of Philosophical Logic* 45.5, pp. 467–483.

— (2016b). "Morita Equivalence". In: *The Review of Symbolic Logic* 9.3, pp. 556–582.

— (2017). "From Geometry to Conceptual Relativity". In: *Erkenntnis* 82.5, pp. 1043–1063.

— (2022). "Mutual Translatability, Equivalence, and the Structure of Theories". In: *Synthese* 200.3, p. 240.

Coffey, Kevin (2014). "Theoretical Equivalence as Interpretative Equivalence". In: *The British Journal for the Philosophy of Science* 65.4, pp. 821–844.

Dewar, Neil (2023). "Interpretation and Equivalence; or, Equivalence and Interpretation". In: *Synthese. An International Journal for Epistemology, Methodology and Philosophy of Science* 201.4, pp. 1–24.

Field, Hartry (2016). *Science without Numbers: A Defense of Nominalism*. Second edition. Oxford, United Kingdom: Oxford University Press. 53 pp.

— (forthcoming). "Nonclassical Solutions to the Liar Paradox: Supervaluationist Logic". In: *The Liar Paradox*. Ed. by Lorenzo Rossi. Cambridge University Press.

Fine, Kit (1975). "Vagueness, Truth and Logic". In: *Synthese. An International Journal for Epistemology, Methodology and Philosophy of Science* 30.3–4, pp. 265–300.

— (2012). "Guide to Ground". In: *Metaphysical Grounding: Understanding the Structure of Reality*. Ed. by Fabrice Correia and Benjamin Schnieder. Cambridge University Press, pp. 37–80.

— (2017). "A Theory of Truthmaker Content I: Conjunction, Disjunction and Negation". In: *Journal of Philosophical Logic* 46.6, pp. 625–674.

— (2020). "Yablo on Subject-Matter". In: *Philosophical Studies* 177.1, pp. 129–171.

Glymour, Clark (1977). "The Epistemology of Geometry". In: *Noûs* 11.3, p. 227.

— (2013). "Theoretical Equivalence and the Semantic View of Theories". In: *Philosophy of Science* 80.2, pp. 286–297.

Halvorson, Hans (2019). *The Logic in Philosophy of Science*. Cambridge and New York: Cambridge University Press.

Heim, Irene and Angelika Kratzer (1998). *Semantics in Generative Grammar*. Ed. by Angelika Kratzer. Malden, MA: Blackwell.

Hirsch, Eli (2002). "Quantifier Variance and Realism". In: *Noûs (Detroit, Mich.)* 36.s1, pp. 51–73.

Hirsch, Eli (2005). “Physical-Object Ontology, Verbal Disputes, and Common Sense”. In: *Philosophy and Phenomenological Research* 70.1, pp. 67–97.

— (2009). “Ontology and Alternative Languages”. In: *Metametaphysics: New Essays on the Foundations of Ontology*. Ed. by David Chalmers, David Manley, and Ryan Wasserman. Oxford University Press, pp. 231–58.

Hlobil, Ulf and Robert Brandom (2025). *Reasons for Logic, Logic for Reasons: Pragmatics, Semantics, and Conceptual Roles*. New York, NY: Routledge, Taylor & Francis Group. 342 pp.

Holliday, Wesley H. (2021). “Possibility Semantics”. In: *Selected Topics from Contemporary Logics*. Ed. by Melvin Fitting. College Publications, pp. 363–476.

Kripke, Saul A. (1959). “A Completeness Theorem in Modal Logic”. In: *Journal of Symbolic Logic* 24.1, pp. 1–14.

— (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge: Harvard University Press.

Lewis, David (1975). “Languages and Language”. In: *Minnesota Studies in the Philosophy of Science*. Ed. by Keith Gunderson. Vol. VII. Minneapolis: University of Minnesota Press, pp. 3–35.

— (1986). *On the Plurality of Worlds*. Malden, Mass.: Wiley-Blackwell.

Linnebo, Øystein and Agustín Rayo (2012). “Hierarchies Ontological and Ideological”. In: *Mind; a quarterly review of psychology and philosophy* 121.482, pp. 269–308.

Manzano, María (1999). *Model Theory*. Oxford Logic Guides 37. Oxford : New York: Clarendon Press ; Oxford University Press. 239 pp.

North, Jill (2021). *Physics, Structure, and Reality*. 1st ed. New York: Oxford University Press.

Przelecki, Marian (1969). *The Logic of Empirical Theories*. London, UK: Routledge & Kegan Paul.

Quine, W. V. (1975). “On Empirically Equivalent Systems of the World”. In: *Erkenntnis. An International Journal of Scientific Philosophy (1975-)* 9.3, pp. 313–328.

Rayo, Agustín (2009). “Toward a Trivialist Account of Mathematics”. In: *New Waves in Philosophy of Mathematics*. Ed. by Otávio Bueno and Øystein Linnebo. London: Palgrave Macmillan UK, pp. 239–260.

— (2013). *The Construction of Logical Space*. Oxford University Press.

— (2017). “The World Is the Totality of Facts, Not of Things”. In: *Philosophical Issues* 27.1, pp. 250–278.

Rynasiewicz, Robert (1996). “Absolute versus Relational Space-Time: An Outmoded Debate”. In: *Journal of Philosophy* 93.6, pp. 279–306.

Salmon, Nathan U. (1986). *Frege's Puzzle (2nd Edition)*. Atascadero, CA: Ridgeview Publishing Company.

Schaffer, Jonathan (2009). “On What Grounds What”. In: *Metametaphysics: New Essays on the Foundations of Ontology*. Ed. by Ryan Wasserman, David Manley, and David Chalmers. Oxford University Press, pp. 347–383.

Sider, Theodore (2009). “Ontological Realism”. In: *Metametaphysics: New Essays on the Foundations of Ontology*. Ed. by David Chalmers, David Manley, and Ryan Wasserman. Oxford University Press, pp. 384–423.

— (2020). *The Tools of Metaphysics and the Metaphysics of Science*. First edition. Oxford ; New York, NY: Oxford University Press. 233 pp.

Sklar, Lawrence (1982). “Saving the Noumena”. In: *Philosophical Topics* 13.1, pp. 89–110.

Skyrms, Brian, ed. (2010). *Signals: Evolution, Learning, & Information*. Oxford New York: Oxford University Press. 199 pp.

Soames, Scott (2009). “Direct Reference, Propositional Attitudes, and Semantic Content”. In: *Philosophical Essays, Volume 2: The Philosophical Significance of Language*. Ed. by Scott Soames. Princeton University Press, pp. 33–71.

Stalnaker, Robert (2007). “Ways a World Might Be”. In: *Philosophical Studies* 133.3, pp. 439–441.

Teitel, Trevor (2021). “What Theoretical Equivalence Could Not Be”. In: *Philosophical Studies* 178.12, pp. 4119–4149.

Uzquiano, Gabriel (2004). “Plurals and Simples”. In: *The Monist* 87.3, pp. 429–451.

Van Fraassen, Bas C. (1966). “Singular Terms, Truth-Value Gaps, and Free Logic”. In: *Journal of Philosophy* 63.17, pp. 481–495.

Wallace, David (2020). “On the Plurality of Quantum Theories: Quantum Theory as a Framework and Its Implications for the Quantum Measurement Problem”. In: *Scientific Realism and the Quantum*. Ed. by Steven French and Juha Saatsi. Oxford University Press.

— (2022). “Stating Structural Realism: Mathematics-first Approaches to Physics and Metaphysics”. In: *Philosophical Perspectives*, phpe.12172.

Wallace, David and Christopher G Timpson (2009). “Quantum Mechanics on Spacetime I: Spacetime State Realism”. Version 1. In.

Warren, Jared (2015). “Quantifier Variance and the Collapse Argument”. In: *The Philosophical Quarterly* 65.259, pp. 241–253.

Weatherall, James Owen (2014). “Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent?” Version 3. In.

Appendix

If F is a function from A to B ($F : A \to B$) and $C \subseteq A$, $F|_C$ will denote the restriction of F to C. Fix a space $\langle C, \leq, \bigwedge, \bigvee \rangle$ of contents. For any language L, $\mathfrak{I}(L)$ will denote the set of all mappings from sentences of L to C, and L_i will always denote the base of $\mathcal{L}_i$. We will consider each interpreted language $\mathcal{L}$ based on L as defined in ***Interpreted Language*** to simply the set of all mappings from sentences of syntactic expansions of L to C. For a set S of sentences of an interpreted language $\mathcal{L}$ and a point $I \in \mathcal{L}$, we define $I(S) = \bigwedge_{\phi \in S} I(\phi)$.

Proof of Theorem 1. Let $\langle T_2, \mathcal{L}_2 \rangle$ be an expansion of $\langle T_1, \mathcal{L}_1 \rangle$. Suppose that $\mathcal{L}_2$ is a lossless expansion of $\mathcal{L}_1$. Let $I_1 \in \mathfrak{I}(L_1) \cap \mathcal{L}_1$. We need to show that there is a $I_2 \in \mathfrak{I}(L_2) \cap \mathcal{L}_2$ such that $\bigwedge_{\phi \in T_2} I_2(\phi)$ entails $\bigwedge_{\phi \in T_1} I_1(\phi)$. Since $\mathcal{L}_2$ is a lossless expansion of $\mathcal{L}_1$, it preserves I_1, meaning that there is an assignment $I_2 \in \mathcal{L}_2$ such that $I_2|_{L_1} = I_1$. Because $T_1 \subseteq T_2$, $I_2(T_2)$ entails $I_2(T_1)$. But since $I_2|_{L_1} = I_1$, and T_1 contains only sentences in L_1, we have that $I_2(T_1) = I_1(T_1)$, which means that $I_2(T_2)$ entails $I_1(T_1)$, the desired result. □

Theorem 2 (Dewar's Notational Variants). *Let $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$ be interpreted languages with bases L_ρ and L_μ, such that any sentence of L_ρ can be obtained from a sentence of L_μ by replacing occurrences of 'μ' with 'ρ', and vice-versa. Suppose L^+ is a syntactic expansion of both, containing every sentence of both but also mixed sentences (with both 'ρ' and 'μ'), and that $\mathcal{L}^+$ is an interpreted language with base L^+. Let R be a relation between sentences of L^+ that holds between two sentences when one can be obtained from the other by replacing some occurrences of 'ρ' with 'μ' or vice-versa. Suppose (5.1.i) that all languages respect first-order entailment, (5.1.ii) that $\mathcal{L}^+$ is a lossless expansion of $\mathcal{L}_\rho$ and $\mathcal{L}_\mu$, and (5.1.iii) that any assignment which is either $\mathcal{L}_\rho$-compatible or $\mathcal{L}_\mu$-compatible must assign the same content to sentences related by R. Then, the following results hold:*

(a) For any $\mathcal{L}^+$-compatible assignment I^+, $I^+(M_\rho) = I^+(M_\mu)$, and $I^+(\text{'}\rho = \mu\text{'}) = \top$.

(b) For any $\mathcal{L}_\rho$-compatible assignment I_ρ there is an $\mathcal{L}_\mu$-compatible assignment I_μ such that $I_\rho(M_\rho) = I_\mu(M_\mu)$, and vice-versa.

Proof. For (a), let I^+ be an $\mathcal{L}^+$-compatible assignment. By (5.1.ii), I^+ must be both $\mathcal{L}_\rho$-compatible and $\mathcal{L}_\mu$-compatible. By this and (5.1.iii), I^+ must assign the same content to sentences related by R. Since M_ρ and M_μ only differ by sentences so related, we must have $I^+(M_\rho) = I^+(M_\mu)$. Also, since '$\rho = \rho$' and '$\rho = \mu$' are related by R, we must have $I^+(\rho = \mu) = I^+(\rho = \rho)$, which by (5.1.i) is $\top$. For (b), take any $\mathcal{L}_\rho$-compatible assignment I_ρ. Since $\mathcal{L}^+$ is a lossless expansion, ***Preservation*** entails that there is an $\mathcal{L}^+$-compatible interpretation I^+_ρ such that $I^+_\rho|_{L_\rho} = I_\rho$. Using (a), we have that $I^+_\rho(M_\rho) = I^+_\rho(M_\mu)$. But $\mathcal{L}^+$ is also an expansion of $\mathcal{L}_\mu$, so by clause (ii) of ***Interpreted Language***, we have that $I^+_\rho|_{L_\mu}$ is an $\mathcal{L}_\mu$-compatible assignment. Thus, we can take $I_\mu = I^+_\rho|_{L_\mu}$, and we have that $I_\rho(M_\rho) = I_\mu(M_\mu)$. The other direction is analogous. □

Theorem 3 (Adequacy of Formal Extensions). *Let $\langle T_1, \mathcal{L}_1 \rangle$, $\langle T_2, \mathcal{L}_2 \rangle$, and $\langle T^+, \mathcal{L}^+ \rangle$ be interpreted theories which satisfy (5.2.i) and (5.2.ii), and respect many-sorted first-order entailment. Then for any $I_1 \in \mathcal{L}_1$ there is an $I_2 \in \mathcal{L}_2$ such that $I_1(T_1) = I_2(T_2)$, and vice-versa.*

Proof. Let $I_1 \in \mathcal{L}_1$. By (5.2.i) and ***Preservation***, there is an $I^+ \in \mathcal{L}^+$ such that $I^+|_{L_1} = I_1$. By (5.2.ii), $I^+(T^+ \backslash T_1) = \top$, which means that $I^+(T^+) = I^+(T_1)$. Since $I^+|_{L_1} = I_1$ and T_1 contains only sentences in L_1, we have that $I^+(T_1) = I_1(T_1)$. Thus, $I^+(T^+) = I_1(T_1)$. Now, by (5.2.ii), we also have that $I^+(T^+ \backslash T_2) = \top$, which means that $I^+(T^+) = I^+(T_2)$. Now, define

$I_2 = I^+|_{L_2}$. By (5.2.i), $I_2 \in \mathcal{L}_2$, and by (5.2.ii), and since T_2 contains only sentences in L_2, we have that $I_2(T_2) = I^+(T_2) = I^+(T^+)$. Thus, $I_2(T_2) = I^+(T^+)$. By the two identity results, we have that $I_1(T_1) = I_2(T_2)$, as desired. □