Bruno Jacinto
Javier Belastegui

Quineanism, Noneism and Metaphysical Equivalence

Abstract. In this paper we propose and defend the *Synonymy account*, a novel account of metaphysical equivalence which draws on the idea (Rayo in *The Construction of Logical Space*, Oxford University Press, Oxford, 2013) that part of what it is to formulate a theory is to lay down a theoretical hypothesis concerning logical space. Roughly, two theories are synonymous—and so, in our view, equivalent—just in case (i) they take the same propositions to stand in the same entailment relations, and (ii) they are committed to the truth of the same propositions. Furthermore, we put our proposal to work by showing that it affords a better and more nuanced understanding of the debate between Quineans and noneists. Finally we show how the *Synonymy account* fares better than some of its competitors, specifically, McSweeney's (Philosophical Perspectives 30(1):270–293, 2016) epistemic account and Miller's (Philosophical Quarterly 67(269):772–793, 2017) hyperintensional account.

Keywords: Metaphysical equivalence, Logical space, Verbal disputes, Quineanism, Allism.

1. Introduction

Metaphysical equivalence is a notion of equivalence between theories concerned with what theories *say*, i.e., it is concerned with 'the relationship between theory and world'.[1] To a first approximation, when theories are metaphysically equivalent they require the same of the world for their truth.

McSweeney [41] usefully distinguishes metaphysical equivalence from both *empirical equivalence* and *meaning equivalence*. Roughly, theories are empirically equivalent just in case they share all their observational/empirical commitments.[2] But one might think that theories that agree on their observational commitments may still say different things about what the world is

[1]Miller [43] singles out the following theorists as having defended, of interesting pairs of prima facie incompatible theories in metaphysics, that they turn out to be metaphysically equivalent: Carnap [8], Putnam [50], Hirsch [21,22,25], McCall and Lowe [40], Miller [42], and Benovsky [4]. See also [60, Ch. 4, §4].

[2]See, e.g., Worrall [65]. See also Carnap [7].

Presented by **Walter Carnielli**; *Received* September 8, 2023

Studia Logica
https://doi.org/10.1007/s11225-023-10085-y

Published online: 09 December 2023

like. In such case there can be *empirically equivalent* theories which nonetheless are *metaphysically inequivalent*.

Theories are meaning equivalent just in case their commitments have exactly the same meanings. But one might think that meanings are relatively fine-grained, with the consequence that sentences such as ‘Cambridge is north of NY’ and ‘NY is south of Cambridge’ have different meanings, even though they “require the same of the world for their truth”. In such case there can be *meaning inequivalent* theories which nonetheless are *metaphysically equivalent*.

The primary aim of the present paper is to propose and defend the *Synonymy Account* of *metaphysical equivalence*.[3] The account has the following main components, one semantical-cum-metaphysical, the other epistemological:

1. An *explication* of metaphysical equivalence as *theory synonymy*. Roughly, two theories are synonymous just in case (i) they take the same propositions to stand in the same entailment relations, and (ii) they are committed to the truth of the same propositions.
2. The specification of criteria, on the basis of the explication of theory equivalence as theory synonymy, for *ascertaining* when two theories are equivalent.

As shall be seen, our proposed explication of metaphysical equivalence as theory synonymy owes much to the formal work developed in Kuhn [31]. In addition, it is heavily inspired by the view [52,63] that part of what it is to formulate a theory is to lay down some theoretical hypothesis concerning logical space. Indeed, a different way of glossing *theory synonymy* is that theories are synonymous just in case they have the same (or “isomorphic”, as Rayo puts it) conceptions of logical space and are committed to the truth of the same propositions.

The paper’s subsidiary aim is to apply the *Synonymy Account* to the debate between Quineans and noneists, where *Quineanism* and *noneism* consist of the following views (or “slogans”):

Quineanism: To be is to be the value of a variable [51]. Or in the object language, as we will be putting it throughout the text: To exist is to be some thing.

[3] The *Synonymy Account* is applicable, in the first instance, to theories in metaphysics. While it is expected to constitute a correct account of metaphysical equivalence also between theories in other areas of inquiry, showing how this is so will have to be left for another occasion.

Noneism: Some things do not exist.

Given how some Quineans have dismissed noneism as being flat-out absurd, there has been considerable interest in the debate between Quineans and noneists. In the paper it will be shown that the *Synonymy account* possesses the resources for affording a better understanding of what is involved in this debate. Specifically, it will be shown how the *Synonymy account* construes the debate as concerning what distinctions there are between ways things could (or could not) have been.

1.1. Why Care?

There are at least three reasons why metaphysicians should be interested in metaphysical equivalence and the *Synonymy account*. The first reason has to do with debates in metametaphysics concerning whether metaphysical disputes are insubstantial—in particular, whether they are merely verbal—, and, if so, why.[4] Arguably, metaphysical equivalence offers a sufficient reason for a metaphysical dispute's insubstantiality, at least on one way of understanding 'insubstantial'. If two theories turn out to be metaphysically equivalent, then the debate as to which one is true is, at least in this sense, insubstantial. For then the two theories "require the same of the world in order to be true". Thus, if the *Synonymy account* is correct, then it should prove useful to those interested on whether metaphysical debates are substantial and, if so, why.

A different reason why accounts of metaphysical equivalence, and in particular the *Synonymy account*, should be of interest to metaphysicians is that an improved comprehension of metaphysical equivalence promises to afford a better understanding of certain debates and of what is, or should be, at stake in them. As will be shown, the *Synonymy account* delivers the result that it is often more illuminating to understand what is at stake in some metaphysical debates, such as the one between noneists and Quineans, as concerning whether certain expressive resources are required for appropriately describing the world (in particular, whether the noneists distinction between existence and being identical to something is required for an appropriate description of the world). In this regard, the application of the *Synonymy Account* to the debate between Quineans and noneists provides interesting lessons from a metaphysical point of view. For instance, whether two theories are metaphysically equivalent or not largely depends on the

[4] See, for instance, Hirsch [22–25], Sidelle [55], Lewis [35], Chalmers [10], Chalmers et al. [11].

expressive resources of the languages used to formulate them. As expected, once the previous slogans are precisified, the resulting theories will turn out not to be equivalent. Roughly speaking, due to their distinction between being and existing, the noneists can express more distinctions than the Quineans can and so can make claims, such as ‘there are things that do not exist’, which the Quineans cannot make sense of. However, by enriching the expressive resources of the Quinean language, we can obtain new theories that are equivalent. For instance, one such extension affords Quineans with the means for distinguishing between concrete entities and non-concrete entities. This distinction affords Quineans enough resources to make understandable, in his their own terms, the noneist distinction between existing and non-existing entities.

Also, the *Synonymy account* predicts that certain debates in metaphysics are better construed as concerning whether certain theories are true and should be accepted, instead of having to do with the truth of the particular slogans used to provide initial characterisations of those theories. The labelling of a certain theory as, for instance, “Quinean” or “noneist”, can be misleading. For instance, the theses of Quineanism and noneism are, on the face of it, contradictory. Yet, theories initially labelled as “noneist” may turn out to be equivalent to theories initially labelled as “Quinean” (in which case what proponents of a theory mean by, e.g., ‘some things do not exist’ is not what the proponents of the other theory mean by this sentence). Along the paper we offer some considerations as to why theorists may end up meaning different things with the slogans initially used as labels for their theories.

A third reason why metaphysicians should be interested in metaphysical equivalence concerns progress in metaphysics. A direct way of achieving progress is by ascertaining the truth or falsehood of particular theories. A more indirect way of achieving progress is by ascertaining the metaphysical equivalence between certain theories. Since the success of a theory typically depends on how well it fares in comparison with its rivals, metaphysical equivalence makes it possible to avoid double counting. The reason is that, in general, the merits and shortcomings of a theory are also merits and shortcomings of the theories that are metaphysically equivalent to it, since they all bear the same relationship to the world.

To put it differently, since metaphysically equivalent theories require the same of the world to be true, the choice between metaphysically equivalent theories is akin to the choice between two sentences requiring the same of the world in order to be true. There may be reasons for choosing one or another sentence, but these reasons will not be of relevance vis-à-vis the truth or falsehood of the sentence that turns out to be chosen. Mutatis mutandis for theories.

1.2. Is it Necessary?

One may worry whether an account of *metaphysical equivalence* is even necessary. Isn't such an account a byproduct of an account of the *nature* and *identity* of theories? That is, whatever it takes for theories to be the same, isn't *that* what it takes for theories to be metaphysically equivalent?

For instance, two of the main views on the nature of theories are, respectively, the *syntactic view* and the *semantic view*. According to the syntactic view a theory consists in (or is adequately represented by) a set of sentences of some formal language,[5] whereas on the semantic view a theory consists in nothing but a collection of models, where these are understood as nonlinguistic entities.[6] Together with the view that *for theories to be metaphysically equivalent is for them to be identical*, the syntactic view gives rise to an account of theory equivalence according to which two theories are equivalent just in case they consist of the same set of sentences of some formal language. The semantic view gives rise to an account of theory equivalence according to which two theories are equivalent just in case they consist in the same set of models. Why not stick to one of those accounts?

The problem is that, independently of the syntactic and semantic views' corresponding merits, the conjunction of each one of these views with a conception of metaphysical equivalence as theory-identity gives rise to a problematic account of metaphysical equivalence. The account that results from the syntactic view implies that only theories that are *trivial notational variants* of each other are equivalent.[7] However, this is not right. It is not

---

[5]The *received view* put forward by Carnap [9], Feigl [14] and Hempel [20], imposes the stronger constraint according to which theories contain only *theoretical* terms, which are connected to *observational* terms via *correspondence rules*, which link the two kinds of terms. Here, our interest is not in the received view but just in the weaker, syntactic view. For a recent defence of the received view and its history, see Lutz [38].

[6]Among the proponents of the semantic view are van Fraassen [60], Giere [15], Suppe [58] and Suppes [59]. Some of these theorists take theories to be set-theoretic predicates, whereas others take theories to be collections of state spaces, and even others allow models to be "built" out of somewhat more concrete entities, such as planets and animals. We are using the labels 'syntactic view' and 'semantic view' as these are used in the literature in the philosophy of science on the nature of scientific theories.

[7]Theory T_2 is a *notational variant* of theory T_1 just in case there is a 1–1 and onto function f from the language of T_1 to the language of T_2 such that f maps atomic expressions to like atomic expressions (constants to constants, n-ary predicates to n-ary predicates, connectives to connectives, quantifiers to quantifiers, etc.), f is *compositional*,

because '¬' is used for negation instead of '∼' and '∧' is used for conjunction instead of '&' that we thereby happen to have non-equivalent theories.[8]

The account of metaphysical equivalence that results from the semantic view implies that theories are equivalent only if they consist in the same collection of models. But consider the collection of models consisting in all partially ordered sets such that every pair of elements has both a least upper bound and a greatest lower bound and the collection of models consisting in all algebraic structures that satisfy the commutative, associative and absorption laws. The models in the first class consist in *pairs* of a domain and a relation on that domain. Models in the second class consist of n-tuples with at least a domain and the join and meet operations on that domain, and so all such models are sequences of three or more elements. Thus, the two collections of models are different. Yet the theories that correspond to the two collections of models are equivalent, corresponding to the theory of lattices.[9]

Thus, metaphysical equivalence consists in something over and above the relation of being the same theory that arises from the syntactic view or the semantic view. Hence, even if one of these views on the nature of theories is correct, an account of metaphysical equivalence is still required.[10] We offer here the *Synonymy account* of theory equivalence, and argue for its adequacy.

The syntactic and semantic views provide different conceptions about what *scientific* theories are. But one may object that *metaphysical* theories differ radically from scientific ones and therefore that different criteria

---

and $\{f(\varphi) : \varphi \text{ is a commitment of } T_1\}$ is the set of commitments of T_2. Theory T_2 is a *trivial* notational variant of theory T_1 if and only if f the identity mapping witnesses the notational variance of T_1 and T_2.

[8]Proponents of the syntactic view may reply by appealing to a coarser notion of metaphysical equivalence, for instance, by requiring two theories to have a common definitional extension in order to count as equivalent. Such a reaction would underscore the present worry with the equation of metaphysical equivalence with theory identity. Furthermore, we argue in §7 that the view that the existence of a common definitional extension is a necessary condition for metaphysical equivalence, recently proposed by McSweeney [41], is also problematic.

[9]For the theory of lattices, see, e.g., Davey and Priestley [12].

[10]In effect, Halvorson [17] surveys three accounts of theory equivalence that would fit naturally with the semantic view and shows the inadequacy of each one of them. For a recent exchange concerning the adequacy of the semantic view, see also Glymour [16] and Halvorson [18]. For a different sort of objection to the adequacy of the semantic view, see Azzouni [1].

should apply to them. For instance, one may object that, contrary to scientific theories, metaphysical theories are often stated in an informal and less precise fashion. If so, formal criteria as the ones that will be proposed and discussed in this paper have no place in discussions about metaphysical theories.

We think that this is not the case. To begin with, scientific theories are often formulated informally, using a mixture of ordinary language, mathematics (if at all), and technical vocabulary specific to the discipline concerned. More importantly (for the present purposes), standard applications of mereology and modal logic show that there are metaphysical theories that are explicitly formally formulated; moreover, even if many metaphysical claims are stated informally, philosophers are usually very explicit about what they take to be their theory's commitments (i.e., those sentences that the theorists take to be true) and about which of these commitments they take to entail which consequences. And when philosophers aren't so explicit, informally stated claims and the properties of the entailment relations can be further precisified by using formal tools (be it by philosophers of science or by metaphysicians). Doing so leads in any case to further fruitful theorising.

1.3. Structure

The paper is structured as follows. In Section 2 the reception of noneism by Quineans is considered with the purpose of extracting some desiderata that should be satisfied by an account of metaphysical equivalence if it is to be correct. In Section 3, the paper's key section, the *Synonymy account* is developed in full detail. First, an explication of theory equivalence as *theory synonymy* is offered, as well as explications of related notions. Second, the notion of a *deeply correct translation* (a notion employed in the characterisation of *theory synonymy*) is characterised, and some principles for determining when translations are deeply correct are presented. In particular, we formulate a defeasible rule of thumb for determining when theories are synonymous, and so metaphysically equivalent.

Then, in Section 4, the *Synonymy account* is applied to the debate between noneists and Quineans. It is first shown that the account satisfies the desiderata laid out in §2. Afterwards, the *Synonymy account* is shown to afford a better understanding of the dialectic between noneists and Quineans (being expected to shed light also on other debates in metaphysics). Finally, in Section 5 we show that the *Synonymy account* of metaphysical equivalence fares better than both McSweeney's epistemic account and Miller's hyperintensional account.

2. Quineanism, Noneism, and Some Desiderata

In this section we will consider the dispute between Quineans and noneists in order to extract data points that a correct account of metaphysical equivalence should be able to accommodate, explain or predict. Recall the "slogans" of quineanism and noneism[11]:

Quineanism: To exist is to be some thing.

Noneism: Some things do not exist.

Typical examples given by noneists of things that do not exist are *fictional entities*, *possibilia* and *mathematical entities*.[12] That is, noneists hold that every fictional entity, *possibile* and mathematical entity is something, even though no fictional entity, *possibile* and mathematical entity exists. According to them, while Santa Claus, the possible seventh son of Kripke and the number π are all something, none of them exists.

Noneism has been found to be unintelligible by many philosophers, in particular by supporters of Quineanism.[13] Quineans claim an inability to make sense of the noneists' distinction between existence and being something. According to Quineans, to exist is to be something. So, for Quineans, for some thing not to exist is for some thing not to be some thing. Since, from their standpoint, the claim that some thing is not something is not only *false* but also *absurd*, several Quineans find noneism *unintelligible*.

There are five aspects concerning how Quineans should understand and engage with noneism that may also be understood as data points for a correct account of theory equivalence. That is, a correct account of theory equivalence should be able to accommodate, explain or predict those aspects. In this section we introduce these data points, which can be found in, e.g., Lewis [35], Priest [49] and Woodward [64].

The first data point has already been alluded to. It concerns the fact that sometimes a theory will be understood as being *absurd* or *unintelligible*, and so not just as *false*, by proponents of another theory.

The second data point concerns the status of a common language, such as English, as the means through which proponents of two theories should

[11]The label 'quineanism' stems from the fact that this view is quite close to Quine's view according to which 'to be is to be the value of a variable' [51] (the difference being that we are expressing the view in the object language rather than metalinguistically).

[12]Noneist frameworks are developed in, e.g., Routley [53] and Priest [48].

[13]See, e.g., Lycan [39] and van Inwagen [61].

interpret each other. In order to flesh out what is at stake, consider the question whether Quineans should interpret noneists as meaning by 'some things do not exist' the literal meaning of the English sentence 'some things do not exist'. If Quineans interpret noneists in this way, then, given Quineans' views on the meaning of English's 'some things do not exist', they will take noneists to be advocating something absurd or unintelligible. For this reason, Lewis claims that such interpretation of noneists is a *misinterpretation*, since, according to him, 'to impute contradiction gratuitously is to mistranslate' [35, p. 26].

Call two words *homonymous*, in the context of the present paper, just in case they have the same spelling and pronunciation.[14] Furthermore, call an interpretation *homonymous* just in case any word or sentence used by a speaker is interpreted by the hearer as having the same meaning as an *homonymous* word or sentence of the hearer's language. Lewis draws attention to an aspect of theorising which reveals that homonymous interpretation based on the assumption that proponents of different theories share a common language may lead to misinterpretation, even when the theorists in fact are members of the same linguistic community. This aspect concerns the fact that theorists also entertain views on the meaning of the expressions of their language and that these views influence the words they choose to express their commitments.

If proponents of different theories have different views on the meanings of certain expressions of their common language, and one of them chooses to express his position by appealing to some of these expressions, then homonymous interpretation is not guaranteed to lead to correct interpretation. For then the hearer will interpret the speaker according to *his* own views on the meanings of the expressions of their common language, and thus will miss out on what is said by the speaker.

To use one of Lewis's examples, when Berkeley uses the sentence 'the tree in the quad exists' to report one of his commitments, Berkeley should not be understood as claiming that the tree in the quad exists, unless we believe, as Berkeley does, that 'the tree in the quad' denotes an idea. Since Berkeley holds that everything is mental, if he were to be interpreted homonymously, then he would be understood as contradicting himself, holding at the same time that something non-mental exists (namely, the denotation of 'the tree

[14] Thus, according to the way 'homonymous' will be used, homonymous words may (but need not) have the same meaning.

in the quad') and that everything is mental. But since Berkeley is not contradicting himself he should not be interpreted homonymously, regardless of the fact that he is stating his view in a common language.

Thus, this second data point can be captured by the following slogan: *homonymous interpretation is not sacrosanct*. That is, homonymous interpretation based on the assumption that proponents of two different theories are speaking in a common language sometimes leads to misinterpretation, even when the two theorists are in fact speaking in a common language.

A different reason for thinking that homonymous interpretation is not sacrosanct has to do with the observation that theories come with their own *terms of art*. For instance, an interpretation of the term 'fitness', as used in biological theory, as meaning the same as 'fitness' as the word is used in the vernacular would lead to misinterpretation.

The third data point can be captured by the slogan that *theories are sometimes incommensurable*. Thus, sometimes a theory lacks the conceptual resources to fully interpret a different theory. This point is made by both Lewis [35], a Quinean, and Priest [49], a noneist, with respect to the relationship between Quineanism and noneism.

Since homonymous interpretation leads to imputing a commitment to an absurdity, Lewis holds that Quineans should interpret noneists non-homonymously, by taking the noneists' assertion of 'Santa Claus, the seventh son of Kripke and the number π are all something' to mean the same as what Quineans mean by the sentence 'Santa Claus, the seventh son of Kripke and the number π all *exist*'. More generally, Lewis holds that Quineans should understand 'is something', as used by noneists, as having the same meaning as 'exists' as used by Quineans. Thus, according to him, Quineans should understand noneists as advocating *allism*:

Allism: Fictional entities, possibilia, mathematical entities and the like all exist.

Importantly, Lewis holds that interpreting noneists as allists *does not suffice* to make noneism *fully understandable* to Quineans. He argues that (several) Quinean theories lack the conceptual resources required for understanding the noneist's use of 'exists'. For instance, Quineans should not understand 'exists' as meaning the same as 'is present', nor as 'is actual'. Even when the noneists say 'it is exactly the present or actual things that exist', they still take this to be a substantive claim, rather than one true simply by definition.

On Lewis's view, Quineans should understand noneists as being committed to there being a specific distinction between things that they purport to capture through their use of 'exists'. Still, understanding noneists in this way does not make their position fully intelligible to Quineans, since (several) Quinean theories lack the linguistic resources for fully understanding the noneists' use of 'exists'. That is, they cannot themselves talk about the purported distinction between things that noneists' take to be captured by their use of 'exists'.[15]

By contrast with Lewis, Priest explicitly rejects the view that Quineans should interpret the noneists' 'is something' as meaning the same as what they, Quineans, mean with 'exists'. Instead, he holds that Quineans should interpret 'is something' homonymously. Notwithstanding, Lewis and Priest *agree* on the view that Quinean theories *lack* resources required for enabling Quineans to *fully understand* noneist theories. Thus, according to Priest [49, p. 251],

> 'There is absolutely no reason why, in a dispute between noneists and Quineans, everything said by one side must be translated into terms intelligible to the other. No one ever suggested that the notions of Special Theory of Relativity need to be translated into categories that make sense in Newtonian Dynamics (or vice versa) (. . .). Though there may be partial overlap, each side may just have to learn a new language game.'

So, Priest holds that Quinean theories' lack the linguistic resources enabling Quineans to fully understand noneist theories. Furthermore, he takes this feature of relationship between Quinean and noneist theories to also be a feature of the relationship between other pairs of theories purporting to describe the same target phenomena (e.g., the Special Theory of Relativity and Newtonian Dynamics).

Thus, Lewis and Priest both hold that Quineans lack the expressive resources to fully understand noneists. They differ in that Priest thinks the noneists' use of 'exists' marks a genuine distinction between things, whereas Lewis thinks that it doesn't. That is, by contrast with Priest, Lewis's view is that the sentences of the noneists' language that Quineans cannot interpret are uninterpretable *tout court*: they simply fail to express a proposition. Still, the view that such sentences are uninterpretable is rather different

[15] One quick remark. It is not meant by this that there are expressive resources such that, if Quineans had them, then they would be able to fully understand noneists. There might not be such expressive resources. Furthermore, the purported distinction that noneists think is captured by their use of 'exists' may turn out not to be there at all.

from the view, rejected by Lewis, that once they are correctly interpreted noneists are committed to absurdities.

The fourth and fifth data points resulting from the discussion of how Quineans' should understand and engage with noneism are present in Woodward's argument for the claim that *allism* and *noneism* are one and the same view. In a nutshell, the argument is the following:

> 'Now imagine that we *rewrite* our noneism theory: whereas previously we said that an object exists, we now say that an object is actually concrete, and where we previously said that an object is self-identical, we now say that an object exists. No one seriously thinks that this relabelling exercise has changed anything: all we've done is rewritten the theory in a different way. But our rewritten noneist theory *just is* allism and our new quantifiers are defined in exactly the same way as Quine's!' [64, p. 191]

In this passage Woodward is alluding to a specific translation from the noneist vocabulary to the allist vocabulary that translates 'exists' as 'actually concrete' and 'something' as 'exists'. Woodward claims that this translation 'is guaranteed to always take us from truths to truths and from falsehoods to falsehoods' [64], and takes this to be evidence for the claim that noneism just is allism.

The present interest is not in Woodward's claim that noneism is allism. Even though, as shall be seen, there arguably is a sense in which noneism just is allism, in our view this claim must be qualified in important respects. Instead, our interest is in two observations that fall out of Woodward's discussion. The first of these, which corresponds to our fourth data point, is that *theories that would appear to be contradictory if interpreted homonymously are sometimes equivalent*. Woodward's argument, if successful, shows that noneism and allism are one such pair of theories. Furthermore, even if his argument for the equivalence of noneism and allism turns out to be unsuccessful, once it is appreciated that homonymous interpretation sometimes leads to misinterpretation it can also be realised that there can be pairs of equivalent theories that would appear to be contradictory if homonymously interpreted.

The other observation which we take from Woodward's discussion, and that corresponds to our fifth and final data point, is his appeal to *nontrivial* translations between theories' languages as means of showing that 'there is *total* overlap between the conceptual resources of the two theories' [64, p. 191]. In light of the observation that *homonymous translations are not* *sacrosanct*, an appeal to nontrivial translations as a way of showing the equivalence between some pairs of theories is to be expected.

Summing up, the discussion involving the Quineans' reception of noneism reveals that a good account of theory individuation …

1. … should predict the conditions under which it is likely for a theory to be received as absurd by the proponents of another theory;
2. … should not have homonymous interpretation as a mandatory feature of the interpretation of one theory by the proponents of another theory, even when the proponents of the two theories are members of the same linguistic community;
3. … should allow for cases in which a theory is intelligible to the proponents of another theory even though the first theory cannot be fully understood in terms of the resources afforded by the second theory;
4. … should explain how theories that would appear to be contradictory if interpreted homonymously are sometimes equivalent, and offer the means of predicting when this will happen;
5. … should yield conditions under which translations, such as the one proposed by Woodward, count in favour of the claim that 'there is total overlap between the conceptual resources of the two theories'.

3. The Synonymy Account

The aim of this section is to present the *Synonymy Account* of metaphysical equivalence. Section 3.1 is devoted to the characterisation of the technical notion of a *theory formulation*. In Section 3.2 we offer our proposed explication of theory equivalence as *theory synonymy*. In Section 3.3 we propose criteria for *determining* whether two theories are metaphysically equivalent (given the proposed explication of metaphysical equivalence as *synonymy account*).

3.1. Formulations of Theories

The *synonymy relation* is specified in terms of what we will call a *formulation* F_T of a theory. We presuppose that any formulation of a theory is:

- Given in a language L_{F_T};
- Proposed by some agents, the xx_{F_T}s;

- Committed to the truth of all and only the sentences of some subset Com_{F_T} of L_{F_T};
- Presupposes an entailment relation $\Vdash_{F_T}$ between the sentences of L_{F_T}.

The xx_{F_T}s are the *proponents of* T, *as formulated via* F_T, and L_{F_T} is *the language of formulation* F_T *of theory* T. By 'language' is meant, in the present context, nothing but a set of *interpreted* sentences. Since sentences are understood as *meaningful strings*, we will assume that they are decomposable into (i) syntactic strings, and (ii) their meanings.

The language L_{F_T} is furthermore assumed to be the language of the xx_{F_T}s as proponents of F_T. So, the xx_{F_T}'s use of L_{F_T} will be governed by some particular conventions. These conventions will determine the *meanings* of the sentences of L_{F_T} *as used by* xx_{F_T} *in formulation* F_T *of* T.

For ease of exposition, sometimes 'sentence' will be used to refer to a "merely" syntactic string, with its meaning abstracted away. Similarly sometimes 'language' will be used to speak of sets of such "syntactically individuated" sentences. We will rely on context to disambiguate between these senses of 'sentence' and 'language'.

Importantly, and as mentioned in Section 2, even if L_T is an interpreted language, the sentences in Com_T, as used by the xx_Ts, might not have the meanings that they in fact have in the xx_Ts' broader common language (e.g., owing to the xx_Ts erroneous views on the semantics of English). For instance, even if the xx_{F_T}s are speakers of English, they may be using 'the tree in the quad' to stand for an idea, rather than for the *tree in the quad*, or they may be using 'Hesperus' to refer to Sirius A instead of Venus.

The entailment relation $\Vdash_{F_T}$ of formulation F_T of theory T, is a relation between sets of sentences of L_{F_T} and sentences of L_{F_T}. φ *is entailed by* Γ *from the standpoint of* F_T, $\langle \Gamma, \varphi \rangle \in \Vdash_{F_T}$ (alternatively, $\Gamma \Vdash_{F_T} \varphi$), just in case, from the standpoint of F_T, part of what the members of Γ, as used by the xx_{F_T}s in formulation F_T, require of the world in order to be jointly true is what φ, as used by xx_{F_T}s in the formulation F_T, requires of the world in order to be true.

Finally, Com_{F_T}, the set of *commitments* of formulation F_T of theory T, is the set of sentences of L_{F_T} to whose truth F_T is committed, as these sentences are used by the xx_{F_T} in F_T.[16]

[16] As we will make clear, Com_{F_T} is closed under consequence. Also, we are assuming that Com_{F_T} captures all the commitments that its proponents incur in virtue of proposing T, including whatever commitments might be implicit. For more on this issue, see, e.g., [26, §2.3].

For ease of exposition, and whenever confusion is unlikely to arise, we will talk of formulations as if they were theories themselves. So, we will often use 'theory X' when what is officially meant is 'formulation X of some theory'. Two assumptions will be in place with respect to the *commitments* and *entailment* relation of a theory T (i.e., formulation T of any theory):

1. Com_T is the same set as the set of sentences φ such that $Com_T \Vdash_T \varphi$.
2. The entailment relation of F_T is *Tarskian*; that is, it is *reflexive*, *transitive* and *monotonic*.[17]

The first assumption is relatively common. Furthermore, it is a consequence of the view that, if, from the standpoint of proponents of T, part of what it is for the elements of Γ to jointly be true is for φ to be true, then φ is a commitment of T if all the elements of Γ are commitments of T. As for, the assumption that $\Vdash_T$ is a Tarskian relation, we will return to it later in this section.[18]

For illustration, we may conceive a formulation A of a theory which determines a quadruple $\langle xx_{\mathrm{A}}, L_{\mathrm{A}}, \Vdash_{\mathrm{A}}, Com_{\mathrm{A}}\rangle$. The proponents xx_{A} of A are Sue and Bob, L_{A} is a first-order language without identity and containing as its only non-logical expressions the constant a and the unary predicate P, $\Vdash_{\mathrm{A}}$ is the set of all multiple premise/single conclusion sequents in language L_{A} which are classically valid, and $Com_{\mathrm{A}} = \{\varphi : Pa \Vdash_{\mathrm{A}} \varphi\}$.

Before proceeding, it will be useful to clarify some issues concerning how formulations are being understood. For the present purposes, to count as a sentence it suffices to be used by theorists as a representation of a way things could (or couldn't) have been. In particular, we allow for theories formulated in languages which are not *compositional*.

Furthermore, and in connection with the discussion, in Section 1, on the semantic view of theories, even formulations of theories in terms of *models* can be seen as having an underlying *language*, *entailment relation* and "set" of *commitments*. Suppose that a theory is presented as a certain subclass X of the class M of models for first-order languages. In such case, the theory's language may be seen as consisting in the subclasses of M. The theory's entailment relation is that relation that obtains between a class of sentences

[17] A relation R on $\wp(X) \times X$ is: i) *reflexive* if and only if, if $\gamma \in \Gamma$, then $\langle \Gamma, \varphi\rangle \in R$; ii) *transitive* if and only if, if $\langle \Gamma, \varphi\rangle \in R$ for all $\varphi \in \Gamma'$ and $\langle \Gamma', \psi\rangle \in R$, then $\langle \Gamma, \psi\rangle \in R$; iii) *monotonic* if and only if, if $\langle \Gamma, \varphi\rangle \in R$ and $\Gamma \subseteq \Gamma'$, then $\langle \Gamma', \varphi\rangle \in R$.

[18] Later on we will point out that '$\Vdash_T$' can be understood in a model-theoretic manner. Still, we are not committed to this being the right way to understand it. It can also be understood proof-theoretically, provided that the defined relation is Tarskian.

Γ and a sentence φ just in case the intersection of Γ is a subclass of φ. The commitment "set" of T consists in the class (of classes) containing all classes of models that are superclasses of X.[19]

3.2. Theory Synonymy

As a first gloss, according to the *Synonymy Account* two theories are equivalent just in case they have formulations such that:

1. They have the same *theoretical structure*; and
2. Each theory's occupiers of each *place* in the theoretical structure require the same of the world for their truth as the other theory's occupiers of that same place.

Our first aim will be to make precise the relation of having the *same theoretical structure*. The following is a preliminary gloss on this notion:

Sameness of theoretical structure (Preliminary Gloss). Theories T_1 and T_2 have the same theoretical structure just in case:

1. T_1 and T_2 possess the same *entailment structure*, and
2. The sentences to whose truth T_1 is committed and the sentences to whose truth T_2 is committed occupy indiscernible places in their shared entailment structure.

---

[19] Arguably, this proposal can also accommodate van Fraassen's [60] view, according to which what is asserted by a theory is that reality can be embedded in some model of a certain class Y of models. Just let Com_T consists in the classes of models that are superclasses of the union of the class Z of classes of models that is such that z belongs to Z if and only if there is some model m in Y such that every model in z can be embedded in m.

Also, a more refined account of entailment can be given provided that a relation $\cong$ between models telling us when two models are representationally the same—e.g., isomorphism—is available. For each sentence φ, let $\varphi_{\cong}$ be that class which, for each model m in φ, contains the class of all models which bear relation $\cong$ to m. Then, $\Gamma \Vdash_T \varphi$ just in case $\bigcap \Gamma_{\cong}$ is a subclass of $\varphi_{\cong}$, where $\bigcap \Gamma$ is the intersection of Γ. Yet a different account is possible, provided the availability of a relation $\equiv$ between classes of models U and V telling us when U and V are representationally the same. In such case entailment may be understood as follows: $\Gamma \Vdash \varphi$ just in case $\bigcap \Gamma \equiv \varphi \cap \bigcap \Gamma$.

There are several other options available. Which one to take will depend on the theorists' conventions with respect to how they are using the language in which their theory is formulated. For the present purposes, the important point is that the notion of a *formulation of a theory* allows for theories to be formulated in radically different ways.

We now turn to precisifying the notion of *sameness of entailment structure*. After doing so we will then make precise the notion of having the same theoretical structure.

3.2.1. Entailment Structure

Spurious Differences

We will begin by offering an *incorrect* precisification of *having the same entailment structure*. By starting this way we will be able to offer insight on our preferred way of fleshing out the notion. A prima facie natural gloss on the conditions under which the entailment structure of T_1 is the same as the entailment structure of T_2 is the following:

Sameness of Entailment Structure (Incorrect). T_1 and T_2 have the same entailment structure if and only if there is a bijection f from L_{T_1} to L_{T_2} such that, $f(\Vdash_{T_1}) = \Vdash_{T_2}$.

Here,

$$f(\Vdash_{T_1}) =_{\text{df}} \{f(\langle\Gamma, \varphi\rangle) : \langle\Gamma, \varphi\rangle \in \Vdash_{T_1}\},$$

where

$$f(\langle\Gamma, \varphi\rangle) =_{\text{df}} \langle f(\Gamma), f(\varphi)\rangle,$$

and

$$f(\Gamma) =_{\text{df}} \{f(\gamma) : \gamma \in \Gamma\}.$$

For each pair in $\Vdash_{T_1}$, there is a "mirror pair" in $f(\Vdash_{T_1})$, and vice versa. Thus, according to this first (incorrect) gloss theories T_1 and T_2 have the same entailment structure just in case $\Vdash_{T_1}$ and $\Vdash_{T_2}$ "mirror each other".

To see why this gloss on sameness of entailment structure is incorrect, consider theories T_{I}, T_{II} and T_{III}. These theories are formulated, respectively, in (rudimentary) languages $L_{T_{\text{I}}}$, $L_{T_{\text{II}}}$, and $L_{T_{\text{III}}}$, where the only sentences of $L_{T_{\text{I}}}$ are $\bot$, A, B and $\top$, the only sentences of $L_{T_{\text{II}}}$ are $\bot$, C, D and $\top$, and the only sentences of $L_{T_{\text{III}}}$ are $\bot$, E, F, G and $\top$.

We will specify the entailment relations of, respectively, T_{I}, T_{II}, and T_{III} by appealing to the following diagrams (Figures 1, 2 and 3):

For every $i \in \{\text{I}, \text{II}, \text{III}\}$, let $\varphi \leq_{T_i} \psi$ if and only if

1. $\varphi = \psi$, or
2. ψ is reachable from φ by going upwards or horizontally along the diagram's edges.

Also, say that $\psi \leq_{T_i} \Gamma$ if and only if $\psi \leq_{T_i} \gamma$, for every $\gamma \in \Gamma$. Then, we define $\Vdash_{T_i}$ as follows:

$$\Vdash_{T_i} =_{\mathrm{df}} \{\langle \Gamma, \varphi \rangle : \text{ for every } \chi \leq_{T_i} \Gamma,\, \chi \leq_{T_i} \varphi\}.$$

Thus, for instance:

1. $A, B \Vdash_{T_{\mathrm{I}}} \bot$;
2. $\top, C \nVdash_{T_{\mathrm{II}}} D$; and
3. $F \dashv\Vdash_{T_{\mathrm{III}}} G$ (i.e., $G \Vdash_{T_{\mathrm{III}}} F$ and $F \Vdash_{T_{\mathrm{III}}} G$).

It should be immediate that there is a bijection f from $L_{T_{\mathrm{I}}}$ to $L_{T_{\mathrm{II}}}$ such that $f(\Vdash_{T_{\mathrm{I}}}) = \Vdash_{T_{\mathrm{II}}}$. Thus, T_{I} and T_{II} count as having the same entailment structure by the above criterion. However, there is no bijection from $L_{T_{\mathrm{I}}}$ or

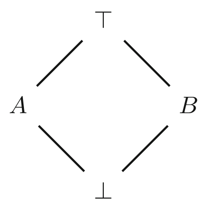

Figure 1. $\leq_{T_{\mathrm{I}}}$

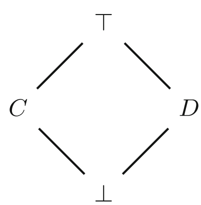

Figure 2. $\leq_{T_{\mathrm{II}}}$

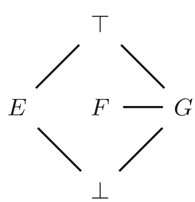

Figure 3. $\leq_{T_{\mathrm{III}}}$

$L_{T_{\mathrm{II}}}$ to $L_{T_{\mathrm{III}}}$. The reason is that these languages have different cardinalities to begin with. So, on the gloss on entailment structure presently being discussed, none of T_{I} and T_{II} has the same entailment structure as T_{III}.

But there is some reason to think that this is the wrong prediction. For it is reasonable to think that if, from the standpoint of a theory, φ and ψ are mutually entailing, then, from the theory's standpoint, φ and ψ require the *same* of the world in order to be true, and so ought to count as *one* vis-à-vis a notion of sameness of entailment structure relevant for *metaphysical equivalence*. Since the preliminary gloss on sameness of entailment structure just considered does not treat F and G as one, that gloss is inappropriate for the purposes of giving an account of metaphysical equivalence.

Same requirements on the world

Let us further develop the above considerations. Say that a set of sentential meanings C '*entails*' a meaning p just in case part of what the members of C jointly require of the world in order to be true is that p be true.[20] Then, it is natural to think that the following hypothesis is true:

Meaning Equivalence Hypothesis. Meanings p and q require the same of the world in order to be true just in case:

(i) for every set C of meanings, C entails p if and only if C entails q; and

(ii) for every set C of meanings, and every meaning s, $C \cup \{p\}$ entails s if and only if $C \cup \{q\}$ entails s.

Whereas before *entailment* was being understood as a relation between *sentences*, it is now being understood as a relation between sentential *meanings*. Notwithstanding, entailment between sentences can be defined from entailment between sentential meanings in the following natural way: a set of sentences of Γ entails a sentence φ just in case the meanings expressed by the members of Γ jointly entail the meaning expressed by φ.

Given the *meaning equivalence hypothesis* and the assumption that entailment, qua relation between *meanings*, is Tarskian, it follows that meanings require the same of the world in order to be true just in case each entails the other. Thus, the *meaning equivalence hypothesis* and the assumption that entailment is a Tarskian relation jointly justify an appeal to the following presupposition:

[20] In what follows we restrict our attention to sentential meanings which are "truth-apt", and which are neither vague nor context-sensitive. We hope to generalise the account to deal with vagueness and context-sensitivity in future work.

Meaning Equivalence Presupposition. For each theory T, $\varphi \dashv\Vdash_T \psi$ if and only if, according to the xx_Ts, φ and ψ, as used by the xx_Ts in the formulation of T, require the same of the world in order to be true.

By the *meaning equivalence hypothesis*, two meanings require the same of the world just in case they are mutually entailing. By the *meaning equivalence presupposition*, if according to a theory sentences φ and ψ are mutually entailing, then, according to theory's proponents, φ and ψ require the same of the world in order to be true.

Similarity

In the remainder of the paper we will adopt the *meaning equivalence presupposition.* As we have seen, under this presupposition the gloss on *sameness entailment structure* adopted so far is *too* sensitive to differences between *sentences.* Furthermore, given the *meaning equivalence presupposition*, it is possible to filter out spurious differences between the (sentential) entailments of two theories by considering *equivalence classes* of sentences under *mutual* (sentential) entailment, thus treating mutually entailing sentences as one. In such case "entailment", qua relation between meanings, is appropriately represented as a relation between equivalence classes of sentences under mutual sentential entailment.

More precisely, let $[\varphi]_T^{\dashv\Vdash}$ be the *equivalence class of* φ *under mutual entailment.* That is,

$$[\varphi]_T^{\dashv\Vdash} =_{\text{df}} \{\psi \in L_T : \varphi \dashv\Vdash_T \psi\}.^{21}$$

Also, for any subset X of L_T, let

$$[X] =_{\text{df}} \{[\varphi] : \varphi \in X\}.$$

Finally, let

$$\Vdash_T^{\dashv\Vdash} =_{\text{df}} \{\langle[\Gamma], [\varphi]\rangle : \Gamma \Vdash_T \varphi\}.$$

Then, in order for two theories to have the same *entailment structure* what should be required is their *similarity* in the following, technical sense:

DEFINITION. (*Similarity*) T_1 and T_2 are *similar*, $T_1 \sim T_2$, if and only if there is a bijection f from $[L]_{T_1}^{\dashv\Vdash}$ to $[L]_{T_2}^{\dashv\Vdash}$ such that, $f(\Vdash_{T_1}^{\dashv\Vdash}) = \Vdash_{T_2}^{\dashv\Vdash}$.

---

[21] We will omit T in $\dashv\Vdash_T$ whenever confusion is unlikely to arise. Similarly, we will omit $\dashv\Vdash$ and T in $[\cdot]_T^{\dashv\Vdash_T}$ whenever confusion is unlikely to arise.

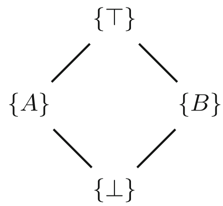

Figure 4. $\leq^{\dashv\Vdash}_{T_{\mathrm{I}}}$

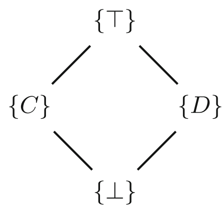

Figure 5. $\leq^{\dashv\Vdash}_{T_{\mathrm{II}}}$

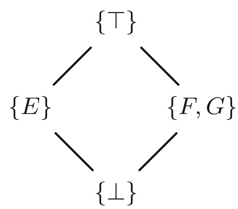

Figure 6. $\leq^{\dashv\Vdash}_{T_{\mathrm{III}}}$

The notion of *similarity* is already defined in Kuhn [31]. Our proposal, not unrelated to that of Kuhn's, is to precisify sameness of entailment structure in the following way:

Sameness of Entailment Structure (First Version). T_1 and T_2 have the same entailment structure if and only if $T_1 {\sim} T_2$.

Now, consider once more the relations $\Vdash_{T_{\mathrm{I}}}$, $\Vdash_{T_{\mathrm{II}}}$ and $\Vdash_{T_{\mathrm{III}}}$. Despite the fact that there is no bijection from $L_{T_{\mathrm{I}}}$ or $L_{T_{\mathrm{II}}}$ to $L_{T_{\mathrm{III}}}$, there are such bijections, *modulo mutual entailment*. To see this, let

$$[\varphi] \leq^{\dashv\Vdash}_{T_i} [\psi] \text{if and only if} \varphi \Vdash_{T_i} \psi,$$

for each $i \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$. Then, the following figures provide a representation of $\leq^{\dashv\Vdash}_{T_{\mathrm{I}}}$, $\leq^{\dashv\Vdash}_{T_{\mathrm{II}}}$ and $\leq^{\dashv\Vdash}_{T_{\mathrm{III}}}$ (Figures 4, 5 and 6):

The diagrams also allow for the representation of the relations $\Vdash^{\dashv\Vdash}_{T_{\mathrm{I}}}$, $\Vdash^{\dashv\Vdash}_{T_{\mathrm{II}}}$ and $\Vdash^{\dashv\Vdash}_{T_{\mathrm{III}}}$, since

$$[\Gamma] \Vdash^{\dashv\Vdash}_{T_i} [\varphi] \quad \text{if and only if,} \quad \text{for every} \quad \psi \leq \Gamma, [\psi] \leq^{\dashv\Vdash}_{T_i} [\varphi].$$

Indeed, the diagrams reveal that there are bijections $f : [L]^{\dashv\Vdash}_{T_{\mathrm{I}}} \to [L]^{\dashv\Vdash}_{T_{\mathrm{III}}}$ and $g : [L]^{\dashv\Vdash}_{T_{\mathrm{II}}} \to [L]^{\dashv\Vdash}_{T_{\mathrm{III}}}$ such that:

$$f(\Vdash^{\dashv\Vdash}_{T_{\mathrm{I}}}) = \Vdash^{\dashv\Vdash}_{T_{\mathrm{III}}} = g(\Vdash^{\dashv\Vdash}_{T_{\mathrm{II}}}).$$

As a consequence, theories T_{I}, T_{II} and T_{III} are all similar to each other. Therefore, they have the same entailment structure.

Again following Kuhn, we will introduce a notion related to that of similarity, except that this novel notion directly appeals to mappings between *sentences*, rather than between *equivalence classes* of sentences:

Definition. (*Similarity via f and g*). Let $f : L_{T_1} \to L_{T_2}$ and $g : L_{T_2} \to L_{T_1}$. T_1 and T_2 are *similar via* f and g, $T_1 \overset{f,g}{\sim} T_2$ if and only if:

1. For every $\Gamma \subseteq L_{T_1}$ and every $\varphi \in L_{T_1}$: $\Gamma \Vdash_{T_1} \varphi$ only if $f(\Gamma) \Vdash_{T_2} f(\varphi)$
2. For every $\Gamma \subseteq L_{T_2}$ and every $\varphi \in L_{T_2}$: $\Gamma \Vdash_{T_2} \varphi$ only if $g(\Gamma) \Vdash_{T_1} g(\varphi)$;
3. For every $\varphi \in L_{T_1}$: $\varphi \dashv\Vdash_{T_1} g(f(\varphi))$;
4. For every $\varphi \in L_{T_2}$: $\varphi \dashv\Vdash_{T_2} f(g(\varphi))$.

By a (small) generalisation of the result reported in Kuhn [31, p. 69], it can be shown that $T_1 \sim T_2$ if and only if there are functions f and g such that $T_1 \overset{f,g}{\sim} T_2$, assuming that both $\Vdash_{T_1}$ and $\Vdash_{T_2}$ are Tarskian.[22] This allows us to provide a second, equivalent explication of sameness of entailment structure, namely:

Sameness of Entailment Structure (Second Version). T_1 and T_2 have the same entailment structure if and only if there are functions $f : L_{T_1} \to L_{T_2}$ and $g : L_{T_2} \to L_{T_1}$ such that T_1 and T_2 are similar via f and g.

As examples of *dissimilar* theories, let Cl and Int be theories such that:

1. $L_{\mathrm{Cl}} = L_{\mathrm{Int}}$ is a propositional language with logical constants $\neg$, $\vee$, $\wedge$, $\to$, $\perp$, and whose only non-logical constant is the propositional letter A;

[22] The notion of similarity via f and g is also defined in Segerberg [54, p. 43] where it is called *syntactic equivalence via* f and g. Pelletier and Urquhart [46, p. 263] define the notion of translational equivalence. Translational equivalence is quite close to similarity via f and g, except that Pelletier and Urquhart impose the restriction that f and g must be compositional. They obtain a notion also defined in Kuhn [31, p. 80], which is called there simply *equivalence via* f and g.

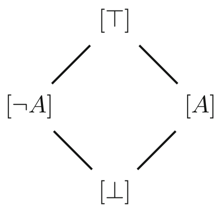

Figure 7. $\leq^{\dashv\Vdash}_{\mathrm{Cl}}$

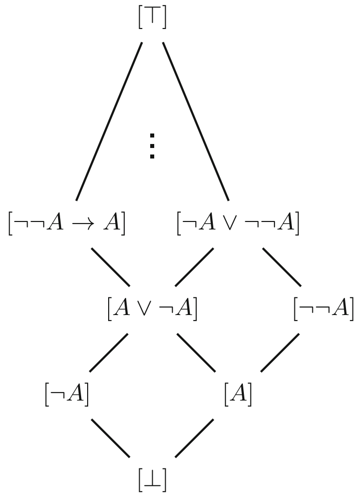

Figure 8. $\leq^{\dashv\Vdash}_{\mathrm{Int}}$

2. $\Vdash_{\mathrm{Cl}}$ is the set of valid sequents of classical propositional logic in language L_{Cl};
3. $\Vdash_{\mathrm{Int}}$ is the set of valid sequents of intuitionist propositional logic in language L_{Int}.

By following the previous diagrammatic conventions, these theories' entailment structures may be represented via the following diagrams (the intuitionistic theory's diagram is perforce incomplete, as there are infinitely many equivalence classes of sentences of L_{Int} under $\dashv\Vdash_{\mathrm{Int}}$) (Figures 7 and 8):

Even though there are only *four* elements in $L^{\dashv\Vdash}_{\mathrm{Cl}}$, namely, the elements of the Lindenbaum algebra on one generator for classical logic, there are *infinitely* many elements in $L^{\dashv\Vdash}_{\mathrm{Int}}$, the elements of the Lindenbaum algebra on one generator for intuitionistic logic (i.e., the elements of the Rieger–Nishimura lattice). Hence, Cl and Int are *dissimilar*, and thus do not count as having the same entailment structure. This is the intuitively right result.

Even when spurious differences between mutually equivalent sentences are filtered out, Cl and Int still possess a very different structure of entailments.

Inclusion of Entailment Structure

A related notion also characterised by Kuhn [31, pp. 73–79] is the relation of *being a fragment*:

DEFINITION. (*Fragment*) Let $f : L_{T_1} \to L_{T_2}$. Then, T_1 is a *fragment* of T_2 *via* f, $T_1 \overset{f}{<} T_2$ if and only if:

For every $\Gamma \subseteq L_{T_1}$ and every $\varphi \in L_{T_1}$: $\Gamma \Vdash_{T_1} \varphi$ if and only if $f(\Gamma) \Vdash_{T_2} f(\varphi)$.

T_1 is a *fragment* of T_2, $T_1 < T_2$, if and only if there is some f such that $T_1 \overset{f}{<} T_2$.

We are interested in a related, albeit more *stringent* notion. Let

$$\Vdash_T^+ =_{\text{df}} \{\varphi \in L_T : \forall \psi \in L_T(\varphi \Vdash_T \psi \Rightarrow \psi \Vdash_T \varphi)\}$$

and

$$\Vdash_T^- =_{\text{df}} \{\varphi \in L_T : \forall \psi \in L_T(\psi \Vdash_T \varphi \Rightarrow \varphi \Vdash_T \psi)\}.$$

Thus, $\Vdash_T^+$ and $\Vdash_T^-$ consist of, respectively, the set of *maximal* sentences and the set of *minimal* sentences relative to T's entailment ordering. Then, the notion of a *stringent fragment* is defined as follows:

DEFINITION. (*Stringent Fragment*) Let $f : L_{T_1} \to L_{T_2}$. T_1 is a *stringent fragment* of T_2 via f, $T_1 \overset{f}{\sqsubset} T_2$, if and only if:

1. $T_1 \overset{f}{<} T_2$,
2. $f(\Vdash_{T_1}^+) \subseteq \Vdash_{T_2}^+$, and
3. $f(\Vdash_{T_1}^-) \subseteq \Vdash_{T_2}^-$.

T_1 is a *stringent fragment* of T_2, $T_1 \sqsubset T_2$ if and only if there is a function f such that $T_1 \overset{f}{\sqsubset} T_2$.

In what follows we will use 'sfragment' instead of 'stringent fragment'. In order for a theory to count as a sfragment of another theory it is not enough for the first theory to be a fragment of the second theory. It is also required that all the minimal and all the maximal elements in the entailment structure of the first theory be mapped to, respectively, minimal and maximal elements of the second theory.

The notion of a sfragment affords the resources to explicate a different relationship between the entailment structures of two theories, specifically:

Inclusion of Entailment Structure. The entailment structure of T_2 includes the entailment structure of T_1 if and only if $T_1 \sqsubset T_2$.

The requirement that minimal elements be mapped to minimal elements and maximal elements be mapped to maximal elements concerns the fact that minimal and maximal elements may be understood as having a special status in a theory. Whereas minimal elements are used by theorists to express what they take to be *absurdities*, as they make maximal demands on the world in order to be true, maximum elements are used by theorists to express meanings which are, according to the theorists, *trivial*, as they make minimal demands on the world in order to be true. Accordingly, mapping a minimal element to something other than a minimal element would misrepresent the entailment structure of a theory, as would mapping a maximal element to something other than a maximal element.

Consider again theories Cl and Int. As noted, it is not the case that these theories are similar. However, Cl $\sqsubset$ Int. That is, Cl is a sfragment of Int. One of the functions witnessing this fact is given by the well-known Gödel–Gentzen translation.[23] A more straightforward (for the present purposes) mapping witnessing that Cl is a sfragment of Int is given by the following function $f : L_{\mathrm{Cl}} \to L_{\mathrm{Int}}$:

1. $f(\varphi) = A$, for all φ such that $\varphi \dashv\Vdash\Vdash_{\mathrm{Cl}} A$;
2. $f(\varphi) = \neg A$, for all φ such that $\varphi \dashv\Vdash\Vdash_{\mathrm{Cl}} \neg A$;
3. $f(\varphi) = \bot$, for all φ such that $\varphi \dashv\Vdash\Vdash_{\mathrm{Cl}} \bot$;
4. $f(\varphi) = \top$, for all φ such that $\varphi \dashv\Vdash\Vdash_{\mathrm{Cl}} \top$.

By contrast, Int is not a fragment of Cl, i.e., Int $\not\sqsubset$Cl. This is the intuitively correct result. Structure-wise, it would appear that proponents of Cl just lack the resources to distinguish between some of the ways of characterizing the world that there are according to Int's proponents.

Conceptions of Logical Space

Rayo's [52, Ch. 2] views on theorizing afford further insight on the notion of *sameness of entailment structure* and its relevance for determining how theories relate. According to Rayo, inquiry can be divided into three stages. The first of these consists in the choice of a language suitable for certain theoretical purposes. The second stage consists in the formulation of a theoretical hypothesis concerning *logical space*—i.e., of a *conception* of *logical*

[23] The Gödel–Gentzen translation is given by the following mapping: $f : L_{\mathrm{Cl}} \to L_{\mathrm{Int}}$: (i) $f(\bot) = \bot$; (ii) $f(\varphi) = \neg\neg\varphi$ where φ is atomic and distinct from $\bot$; (iii) $f(\varphi \wedge \psi) = f(\varphi) \wedge f(\psi)$; (iv) $f(\varphi \vee \psi) = \neg(\neg f(\varphi) \wedge \neg f(\psi))$; (v) $f(\varphi \to \psi) = f(\varphi) \to f(\psi)$.

space. By *conception* of logical space what Rayo has in mind is a conception of the distinctions there are between ways things might or might not have been,.[24] and of how these are ordered according to whether part of what it is for the world to fall under some distinctions is for it to fall under another distinction. The third and final stage of theorizing consists in identifying which of these distinctions are the ones taken to truly characterise how the world is.

On this picture, conceptions of logical space are prerequisites for fruitful inquiry. It is against the background of a conception of logical space that theorists formulate their commitments and what exactly is being advocated by a theorist depends on its background conception of logical space.[25]

For instance, to use one of Rayo's examples, even if two theorists advocate the truth of both p and $\neg\neg p$, they may still be committed to radically different and opposing theories. One of the theorists—e.g., a "classically minded" proponent of Cl—, may think that p and $\neg\neg p$ require the same of the world for their truth, whereas another theorist—e.g., an "intuitionistically minded" proponent of Int—, may think that p requires more of the world to be true than $\neg\neg p$.[26] While the proponent of Int will find the discovery that p was true, over and above $\neg\neg p$, to be a significant discovery, the proponent of Cl will find that nothing new has been discovered. Given

[24]Rayo is only concerned with ways things *might have been* Here, we remain neutral on whether there are not only ways things might have been but also ways things might not have been. According to Miller [43], a notion of sentential meaning appropriate for metaphysical equivalence must be capable of discriminating between the ways things might have been in which a sentence is true as well as the ways things might not have been in which the sentence is true. Miller's [43] hyperintensional account of metaphysical equivalence will be discussed in Section 5.2.

[25]For a similar idea, see Pérez Carballo [47]. As Pérez Carballo points out, one need not think of the positing of a conception of logical space as a way of undergoing a factual commitment, but simply as the adoption of "putative" conceptual resources.

[26]One reason why the logician is a proponent of Int may have to do with the Brouwer–Heyting–Kolmogorov (BHK) interpretation of intuitionistic logic. According to BHK, a proposition follows from other propositions only if there is a method to construct a proof for the former given proofs for the latter. For instance, the inference from $\neg\neg\varphi$ to φ would be valid only if, given a proof that there is no proof of $\neg\varphi$ one could always construct a proof of φ. Since a reductio argument does not show how to construct such a proof for φ from the inconsistency arrived at by assuming $\neg\varphi$, such an inference is intuitionistically invalid. However, this is not the only way to interpret what the intuitionist logician has in mind. One could instead think of sentences as standing for pieces of information that have been proven or verified at certain stages of inquiry or points in time, in the manner of Kripkean semantics for intuitionistic logic. For our purposes, what is important is the possibility of thinking that there is a "wedge" between p and $\neg\neg p$.

that $\neg\neg p$ was true, p was *ipso facto* also true, since, from this theorist's standpoint, $\neg\neg p$ and p require the *same* of the world for their truth. Thus, in general, theorists may agree on the truth of the same sentences while disagree on what it takes for each of those sentences to be true.

If inquiry requires a conception of logical space, it is crucial to have the means to say when two theories have "isomorphic" conceptions of logical space, to use Rayo's terminology, even in those cases in which the theories are formulated in different languages. Rayo's idea is that two theories' conceptions of logical space are isomorphic when each theory can find a distinction that "matches" the other theory's distinction, in that they require the same of the world for their truth. Besides discussion of some examples, no precise account of when conceptions of logical space are isomorphic is offered by Rayo. But one way of making precise (or improving) Rayo's notion of *isomorphism* between conceptions of logical space is in terms of the relation of *sameness of entailment structure*.

The hypothesized distinctions between ways things might or might not have been that compose a conception of logical space may be *represented* by equivalence classes of L_T's sentences, under mutual entailment, insofar as, mutually equivalent sentences require, according to the theory, the same of the world for their truth. Furthermore, entailment between a set X of equivalence classes of sentences and a sentence y may be seen as encapsulating the fact that part of what it is for the sentences in y to be true is for the sentences in X to be jointly true.

Then, a *necessary condition* for two theories to have isomorphic conceptions of logical space is that each theory's entailment relation, modulo mutual entailment, be the "mirror image" of the other theory's entailment relation, modulo mutual entailment. That is, on our proposed precisification a necessary condition for two theories to have isomorphic conceptions of logical space is that they be similar, and so have the *same entailment structure*. For in such case the theories' *conceptions of logical space* have the *same structure*.[27]

So understood, what the fact that T_{Cl} and T_{Int} are *dissimilar* shows is that these theories' underlying conceptions of logical space have different structures. Even if their proponents end up committed to the truth of the same sentences, this agreement masks a robust disagreement with respect to what distinctions there are between ways things might or might not have been.

[27]Our full precisification is in terms of *congruence* a notion characterised in Section 3.2.3.

3.2.2. Theoretical Structure With the characterisation of sameness of entailment structure in place, the explication of sameness of *theoretical structure* may now be offered. The relevant notion is that of *solid similarity.* Let

$$Com_T^{\dashv\Vdash} =_{\mathrm{df}} \{[\varphi] : \varphi \in Com_T\}.$$

Then:

Definition. (*Solid Similarity*). T_1 and T_2 are *solidly similar*, $T_1 {\approx} T_2$, if and only if there is a bijection f from $[L]_{T_1}^{\dashv\Vdash}$ to $[L]_{T_2}^{\dashv\Vdash}$ such that, $f(\Vdash_{T_1}^{\dashv\Vdash}) = \Vdash_{T_2}^{\dashv\Vdash}$ *and* $f(Com_{T_1}^{\dashv\Vdash}) = Com_{T_2}^{\dashv\Vdash}$.

Our proposal is to explicate sameness of theoretical structure in the following way:
Sameness of Theoretical Structure (First Version). T_1 and T_2 have the same *theoretical structure* if and only if $T_1 {\approx} T_2$.

If a mapping witnessing the *solid similarity* between theories T_1 and T_2 exists, then not only do T_1 and T_2 share their *entailment structure*, they are furthermore committed to the truth of sentences whose equivalence classes under mutual entailment are indistinguishable vis-à-vis that entailment structure. Or, to speak in terms of conceptions of logical space, if two theories are solidly similar, then not only are their conceptions of logical space structured in the same way, their commitments amount to distinctions that occupy indiscernible places in that structure.

For some examples, consider once more the theories T_{I}, T_{II} and T_{III}. Let

$$Com_{T_{\mathrm{I}}} =_{\mathrm{df}} \{A, \top\}, \quad Com_{T_{\mathrm{II}}} =_{\mathrm{df}} \{\top\}, \quad \text{and} \quad Com_{T_{\mathrm{III}}} =_{\mathrm{df}} \{F, G, \top\}.$$

Then, by appealing to the previous representations of $\leq_{T_{\mathrm{I}}}^{\dashv\Vdash}$, $\leq_{T_{\mathrm{II}}}^{\dashv\Vdash}$ and $\leq_{T_{\mathrm{III}}}^{\dashv\Vdash}$ we can represent the theoretical structures of T_{I}, T_{II} and T_{III}, doing so by representing the sets $Com_{T_{\mathrm{I}}}^{\dashv\Vdash}$, $Com_{T_{\mathrm{II}}}^{\dashv\Vdash}$ and $Com_{T_{\mathrm{III}}}^{\dashv\Vdash}$ with the points in the respective structures that are inside the dotted lines (Figures 9, 10 and 11):

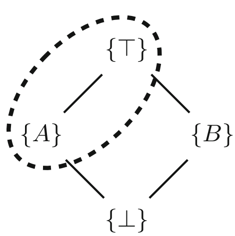

Figure 9. Theoretical structure of T_{I}

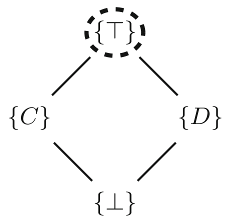

Figure 10. Theoretical structure of T_{II}

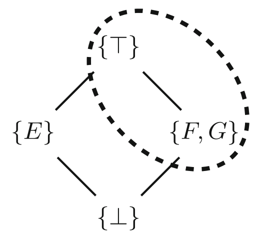

Figure 11. Theoretical structure of T_{III}

There are two bijections from $[L]^{\dashv\Vert\vdash}_{T_{\mathrm{I}}}$ to $[L]^{\dashv\Vert\vdash}_{T_{\mathrm{II}}}$ witnessing the similarity between T_{I} and T_{II}, namely, the bijection that maps $\{A\}$ to $\{C\}$ and the bijection that maps $\{A\}$ to $\{D\}$. In both of these cases, $\{A\}$ is mapped to a set that does not belong to $Com^{\dashv\Vert\vdash}_{T_{\mathrm{II}}}$, even though $\{A\}$ belongs to $Com^{\dashv\Vert\vdash}_{T_{\mathrm{I}}}$. This shows that no equivalence class of sentences, under mutual entailment, to whose truth T_{II} is committed occupies a role in T_{II}'s entailment structure that is indiscernible from the role occupied by A's equivalence in T_{I}'s entailment structure. Thus, $T_{\mathrm{I}} \not\approx T_{\mathrm{II}}$. Hence, T_{I} and T_{II} do not count as having the same theoretical structure according to the present criterion.

On the other hand, according to the present proposal, T_{I} and T_{III} do share the same theoretical structure. For the bijection f from $[L]^{\dashv\Vert\vdash}_{T_{\mathrm{I}}}$ to $[L]^{\dashv\Vert\vdash}_{T_{\mathrm{III}}}$ that witnesses the similarity between T_{I} and T_{III} and which maps $\{A\}$ to $\{F, G\}$ is such that $f(Com^{\dashv\Vert\vdash}_{T_{\mathrm{I}}}) = Com^{\dashv\Vert\vdash}_{T_{\mathrm{III}}}$. Intuitively, that T_{I} and T_{III} share the same theoretical structure is the correct result. If anything breaks the metaphysical equivalence between theories T_{I} and T_{III}, it must be something having to do with what exactly the proponents of these theories mean by the sentences of their respective languages, not with the common structure of their conceptions of logical space, nor with how their commitments fit in that structure.

One can also define a notion close to that of solid similarity, except that it appeals directly to mappings between languages L_{T_1} and L_{T_2}. Where f is any function from L_{T_1} to L_{T_2}, let:

$$T_1 \overset{f}{\lhd} T_2 \quad \text{if and only if,} \quad \text{for all} \quad \varphi \in Com_{T_1}, \quad f(\varphi) \in Com_{T_2};$$

Then, solid similarity *via* functions f and g is defined as follows[28]

Definition. (*Solid Similarity via f and g*). Let $f : L_{T_1} \to L_{T_2}$ and $g : L_{T_2} \to L_{T_1}$. T_1 and T_2 are *solidly similar via* f and g, $T_1 \overset{f,g}{\approx} T_2$, if and only if:

1. $T_1 \overset{f,g}{\sim} T_2$;
2. $T_1 \overset{f\ g}{\lhd\rhd} T_2$.

By appealing to the notion of solid similarity via functions f and g it is possible to explicate sameness of theoretical structure in an alternative, albeit equivalent way:

Sameness of Theoretical Structure (Second Version). T_1 and T_2 have the same *theoretical structure* if and only if there are functions f and g such that: $T_1 \overset{f,g}{\approx} T_2$

[28]The following proofs establish that T_1 and T_2 are solidly similar if and only if T_1 and T_2 are solidly similar via functions f and g.

Proof.: [$T_1 \overset{f,g}{\approx} T_2$ implies $T_1 \approx T_2$] Suppose $T_1 \overset{f,g}{\approx} T_2$ and define $h : [L]^{\dashv\Vdash}_{T_1} \cup [L]^{\dashv\Vdash}_{T_2} \to [L]_{T_1} \cup [L]_{T_2}$ in such a way that $h([\varphi]) = [f(\varphi)]$. Then, h is a bijection witnessing $T_1 \sim T_2$, by a small generalisation of the result shown in Kuhn [31, pp. 69–70]. It will now be shown that (i) $h(Com^{\dashv\Vdash}_{T_1}) \subseteq Com^{\dashv\Vdash}_{T_2}$ and ii) $Com^{\dashv\Vdash}_{T_2} \subseteq h(Com^{\dashv\Vdash}_{T_1})$.

(i) Suppose $x \in h(Com^{\dashv\Vdash}_{T_1})$. Then, $x = h([\varphi])$, for some $\varphi \in Com_{T_1}$. So, $x = [f(\varphi)]$. By $T_1 \overset{f}{\lhd} T_2$, there is a $\psi \in L_{T_2}$ such that $f(\varphi) \dashv\Vdash_{T_2} \psi \in Com_{T_2}$. Hence $x = h([\varphi]) = [f(\varphi)] \in Com^{\dashv\Vdash}_{T_2}$. So, $h(Com^{\dashv\Vdash}_{T_1}) \subseteq Com^{\dashv\Vdash}_{T_2}$.

(ii) Suppose $x \in Com^{\dashv\Vdash}_{T_2}$. Then, $[x] = [\varphi]$, for some $\varphi \in Com_{T_2}$. By $T_2 \overset{g}{\lhd} T_1$, there is a $\psi \in Com_{T_1}$ such that $g(\varphi) \dashv\Vdash_{T_1} \psi$, and thus $g(\varphi) \in Com_{T_1}$. Hence, $f(g(\varphi)) \in f(Com_{T_1})$. So, $\varphi \in f(Com_{T_1})$, by $T_1 \overset{f,g}{\sim} T_2$. Therefore, $[\varphi] \in [f(Com_{T_1})] = h(Com^{\dashv\Vdash}_{T_1})$. Hence, $Com^{\dashv\Vdash}_{T_2} \subseteq h(Com^{\dashv\Vdash}_{T_1})$. ■

Proof. [$T_1 \approx T_2$ implies $T_1 \overset{f,g}{\approx} T_2$] Suppose $T_1 \approx T_2$. Let h be any bijection witnessing $T_1 \approx T_2$. Let $ch : [L]^{\dashv\Vdash}_{T_1} \cup [L]^{\dashv\Vdash}_{T_2} \to L_{T_1} \cup L_{T_2}$ be any function such that $ch([\varphi]) \in [\varphi]$. Define $f : L_{T_1} \to L_{T_2}$ and $g : L_{T_2} \to L_{T_1}$ in such way that $f(\varphi) = ch(h([\varphi]))$ and $g(\varphi) = ch(h^{-1}([\varphi]))$. Then, $T_1 \overset{f,g}{\sim} T_2$, by a small generalisation of the result shown in Kuhn [31, pp. 69–70]. It will now be shown that (i) $T_1 \overset{f}{\lhd} T_2$ and (ii) $T_2 \overset{g}{\lhd} T_1$.

(i) Suppose that $\varphi \in Com_{T_1}$. Then, $[\varphi] \in Com^{\dashv\Vdash}_{T_1}$. So, $h([\varphi]) \in h(Com^{\dashv\Vdash}_{T_1}) = Com^{\dashv\Vdash}_{T_2}$, by $T_1 \approx T_2$, by $T_1 \approx T_2$. Thus, $f(\varphi) = ch(h([\varphi])) \in Com_{T_2}$. Hence, $T_1 \overset{f}{\lhd} T_2$.

(ii) Suppose that $\varphi \in Com_{T_2}$. Then, $[\varphi] \in Com^{\dashv\Vdash}_{T_2} = h(Com^{\dashv\Vdash}_{T_1})$, by $T_1 \approx T_2$. So, $h^{-1}([\varphi]) \in Com^{\dashv\Vdash}_{T_1}$. Thus, $g(\varphi) = ch(h^{-1}(\varphi)) \in Com_{T_1}$. Hence, $T_2 \overset{g}{\lhd} T_1$. ■

So, theories have the same theoretical structure just in case there are pairs of functions from the language of each to the language of the other which witness their sameness of theoretical structure and furthermore map the commitments of each theory to the commitments of the other theory. When that is the case, two theories' conceptions of logical space will not only have the same structure but their *commitments* will occupy indiscernible places in that structure.

Inclusion of Theoretical Structure

We can also characterise what it is for the theoretical structure of a theory to be *included* in the theoretical structure of another theory. To do so, the key notion is that of a *solid sfragment*:

DEFINITION. (*Solid sfragment*) T_1 is a *solid sfragment* of T_2 via f if and only if:

1. T_1 is a sfragment of T_2 via f; and
2. $f(Com_{T_1}) \subseteq Com_{T_2}$.

Then, inclusion of theoretical structure is characterised as follows:

Inclusion of Theoretical Structure. T_2's theoretical structure *includes* T_1 if and only if there is some function f such that T_1 is a solid sfragment of T_2 via f.

For an example, consider once more theories Cl and Int. Let

$$Com_{\text{Cl}} =_{\text{df}} \{\varphi : A \Vdash_{\text{Cl}} \varphi\}, \quad \text{and} \quad Com_{\text{Int}} =_{\text{df}} \{\varphi : A \Vdash_{\text{Int}} \varphi\}.$$

That is, each theory is committed to what, according to them, is a consequence of A. Then, we get the following representations of their respective theoretical structures (Figures 12 and 13).

In such case, Cl's theoretical structure is *included* in Int's theoretical structure. Indeed, as it turns out, the theories are committed to the truth of the same sentences, i.e., $Com_{\text{Cl}} = Com_{\text{Int}}$. Yet, as the representations of

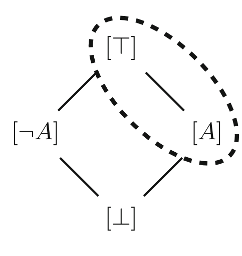

Figure 12. Theoretical structure of Cl

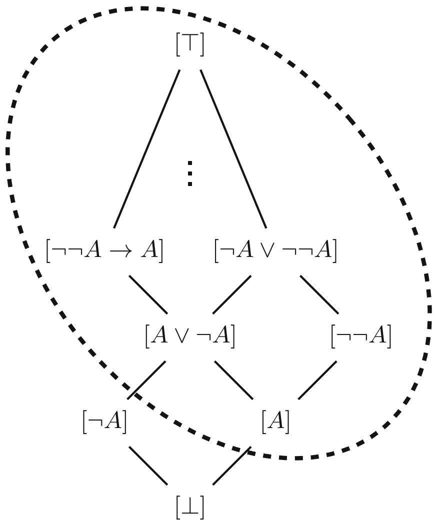

Figure 13. Theoretical structure of Int

their respective theoretical structures reveal, the two theories are anything but metaphysically equivalent, owing to the previous observation that they have radically different entailment structures.

3.2.3. Theory Synonymy We are almost in a position to offer our proposed precisification of metaphysical equivalence. To do so, we will distinguish between *correct translations* and *deeply correct translations*:

Definition. (*Correct Translation*). A function f from L_{T_1} to L_{T_2} is a *correct translation* if and only if, for all $\varphi \in L_{T_1}$, φ and $f(\varphi)$ literally require the same of the world to be true.

Definition. (*Deeply Correct Translation*). A function f from L_{T_1} to L_{T_2} is a *deeply correct translation* if and only if, for all $\varphi \in L_{T_1}$, φ and $f(\varphi)$, as used by, respectively, the xx_{T_1} and the yy_{T_2}s in the formulations of their respective theories, require the same of the world for their truth.

Thus, whereas correct translations between the languages of two theories are sensitive to the sentences' literal meanings, deeply correct translations between those languages are sensitive to the meanings of sentences as they are used by the proponents of the respective theories in their corresponding formulations. *Synonymy* is defined as follows[29]:

Definition. (*Theory Synonymy*). T_1 and T_2 are *synonymous via* functions f and g, $T_1 \overset{f,g}{\equiv} T_2$, if and only if $T_1 \overset{f,g}{\approx} T_2$ and both f and g are deeply correct translation schemes.

T_1 and T_2 are *synonymous* if and only if there are functions f and g such that $T_1 \overset{f,g}{\equiv} T_2$.

Our proposed explication of metaphysical equivalence is as follows[30]: **Metaphysical Equivalence is Theory Synonymy**. Theories T_1 and T_2 are metaphysically equivalent if and only if there are formulations F_{T_1} of T_1 and F_{T_2} of T_2 such that $F_{T_1} \equiv F_{T_2}$.

The reason why theory synonymy is characterised in terms of *deeply* correct translations, rather than in terms of correct translations, has to do with Lewis's observations mentioned in Section 2. As was shown there, the interpretation of a theory needs to be sensitive to what proponents of a theory intend to express with the sentences used in their formulation of the theory, independently of whether that matches the literal meaning of those sentences.

Consider again theories T_{I} and T_{III}. In order to determine whether they are synonymous it is not sufficient to determine whether they are solidly similar (and so, whether they have the same theoretical structure). For proponents of T_{I} might mean with A something quite different from what proponents of T_{III} mean with F and G. In effect, it might be that what proponents of T_{I} mean with A is that dinosaurs are extinct, whereas what proponents of T_{III} mean with F and G is that dinosaurs are *not* extinct. In such case, even though the two theories have the same *theoretical structure*, they are *not synonymous*, and thus *not equivalent*. Notwithstanding, if two theories have a different entailment structure this already shows that they are not equivalent. There is no need to consider what their proponents mean with the sentences of their languages.

One of the aims of appealing to solid similarity was that of having a minimally satisfactory necessary condition for theory equivalence which did not require interpretation of the theory's language. Even though there is

[29] There is an equivalent formulation of theory synonymy that appeals to bijections witnessing the solid similarity between T_1 and T_2. However, the present formulation will suffice for our purposes.

[30] This is a place where it is relevant to distinguish between theories and their formulations.

a sense in which this is indeed the case, note that interpretation is still required at some level. Interpretation plays a crucial role when determining the entailment relation of each theory. For it is a tacit assumption of our proposal that when, according to a theory T_1, Γ entails φ and according to a theory T_2, Δ entails ψ, the same is meant with the two occurrences of 'entails'.[31]

The structural relation of being a stringent fragment via f, $\overset{f}{\sqsubset}$, in conjunction with the notion of a deeply correct translation, gives rise to the notion of *embeddability*:

Definition. (*Embeddability*). A theory T_1 is *embeddable* in theory T_2 just in case there is a deeply correct translation f such that $T_1 \overset{f}{\sqsubset} T_2$.

In Rayo's terms, when T_1 is *embeddable* in T_2, all distinctions in T_1's conception of logical space are also present in T_2's conception of logical space, while it is perhaps the case that according to T_2 there are further distinctions between ways things might have been. Or, to put it differently, T_1 is embeddable in T_2 when T_1's conception of logical space is a *part* of T_2's conception of logical space. Depending on the further specifics of the theories Cl and Int, it might be that the former theory is embeddable in the latter. In such case the intuitionistically aligned theory Int will have available all the distinctions available to the classically aligned theory Cl, and perhaps some more. Embeddability will play an important role in the discussion to take place in Section 4.

A notion stronger than *embeddability* is that of *theoretical embeddability*. We will call an embedding of T_1 in T_2 *theoretical* if and only if T_1 is a *solid* sfragment of T_2 via f. Thus a theoretical embedding also preserves theoretical commitments.

Another relevant relation between theories is that of *congruence*:

Definition. (*Congruent*). Theories T_1 and T_2 are *congruent* just in case there are deeply correct translations $f : L_{T_1} \to L_{T_2}$ and $g : L_{T_2} \to L_{T_1}$ such that T_1 and T_2 are similar via f and g.

As previously mentioned, the *similarity* between two theories constitutes a necessary condition for the "isomorphism", in Rayo's sense, between their conceptions of logical space. The theories' conceptions of logical space will indeed be isomorphic provided that there are *deeply correct* translations

[31] Furthermore, it is a tacit assumption of our proposal that both occurrences mean that part of what it is for the premises to be jointly true is for the conclusion to be true.

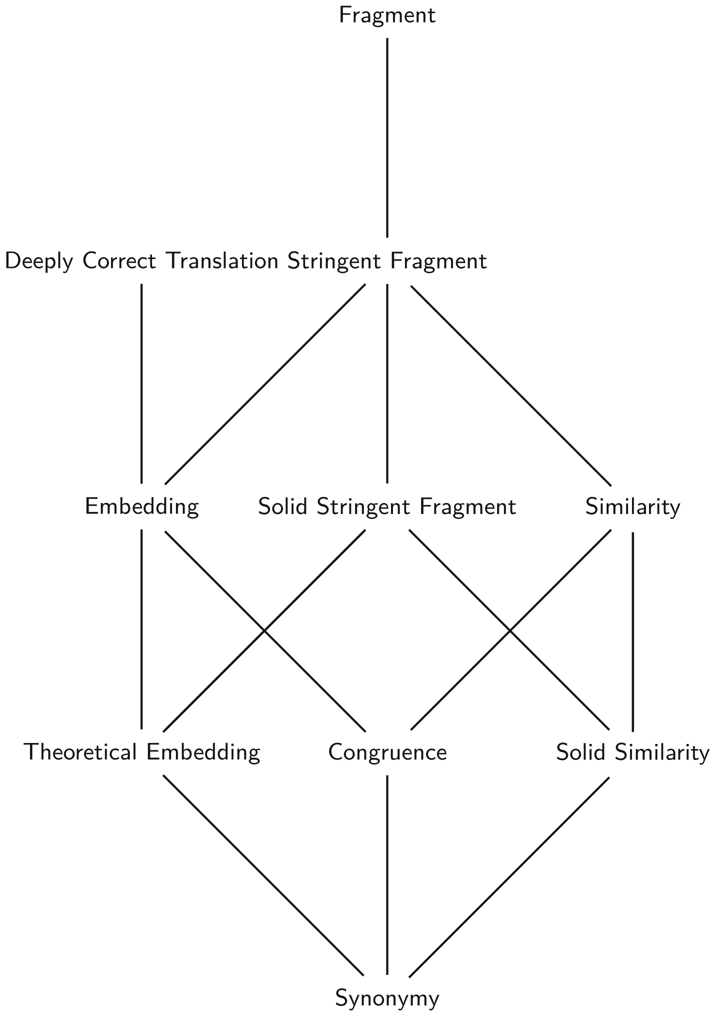

Figure 14. Synonymy Account

witnessing the theories' similarity. That is, *congruence* constitutes a *precisification* of *isomorphism* between conceptions of logical space.

This concludes the presentation of the first component of the *Synonymy account* of metaphysical equivalence, namely the explication of metaphysical equivalence as theory synonymy. A summary of the relations that have been discussed can be found in Figure 14. We now turn to the second component of the *Synonymy account*: the formulation of criteria for determining when mappings between languages constitute deeply correct translations.

3.3. Deeply Correct Translations

One difficulty with determining whether (formulations) T_1 and T_2 are synonymous has to do with the fact that the *syntactic* information contained in T_1 and T_2 does not, on its own, suffice to determine whether functions $f : L_{T_1} \to L_{T_2}$ and $g : L_{T_2} \to L_{T_1}$ are deeply correct translations. To add to this difficulty, proponents of T_i may be using the sentences of their theory's language to mean something other than what φ literally means (in the language of the theory's proponents' broader community).

Consider a language L'_{T_i} syntactically just like L_{T_i} and such that what is literally meant by each sentence φ of L'_{T_i} is what the proponents of T_i mean by φ in formulation T_i. Then, the question whether f and g are *deeply correct* translations can be substituted by the question whether $f' : L'_{T_1} \to L'_{T_2}$ and $g' : L'_{T_2} \to L'_{T_1}$ are *correct* translations. One way to determine whether this is the case consists in determining whether they are *convention-friendly translations*, where the notion of a *convention-friendly translation* is defined as follows:

Definition. (*Convention-Friendly Translation*). Let L_1 be a language of a linguistic community C_1 and L_2 be a language of a linguistic community C_2. Also, let $L \leq L'$ if and only if L' is a superlanguage of L—i.e., a language which includes all the sentences of L, with the same literal meanings as the ones those sentences have in L—which is also a language of the community of speakers of L.

A translation f mapping L_1 to L_2 is a *convention-friendly translation* if and only if there could be a language L such that:

1. $L_2 \leq L$;
2. There is a correct description of the beliefs, desires and intentions of the members of C_1 in L;
3. This description, in conjunction with the translation of L_1 given by f, yields a description, in L, of the linguistic practices of C_1 as a community of speakers conforming to a *convention of truthfulness and trust*, in Lewis's [34] sense, in the used fragment of L_1.[32]

Lewis [34, p. 167] offers the following characterisation of what it is for a community to be truthful and trusting in a language L:

[32] See also Lewis [32]. Also, we want to stress that here we are only concerned with sentential meaning, not with word meaning. On the face of it, even if a convention of truthfulness and trust settles sentential meaning, it does not settle word meaning—see Janssen-Lauret and MacBride [27] for more on Lewis's views on this issue.

'To be truthful in L is to act in a certain way: to try never to utter any sentences of L that are not true in L. Thus, it is to avoid uttering any sentence of L unless one believes it to be true in L. To be trusting in L is to form beliefs in a certain way: to impute truthfulness in L to others, and thus to tend to respond to another's utterances of any sentence of L by coming to believe that the uttered sentence is true in L.'

Let us illustrate what it is for a translation to be convention-friendly with a simple example. Suppose that we have a true description, in English, of the beliefs, intentions and desires of the community of speakers of French. Suppose also that we have a translation of French into English according to which the sentence 'le chat est sur le paillasson' is translated as 'Paris is located in England'. Furthermore, suppose that in the large majority of the occasions in which a speaker of French utters 'le chat est sur le paillasson', he intends to communicate that the cat is on the mat. In such case the translation is *not* convention-friendly. The reason is that the description that we obtain in English is not one in which the sentence is commonly uttered by speakers of French when they believe that Paris is located in England. Furthermore, as the example shows, the translation is in fact *incorrect*.

The following principle offers some guidance in determining the correctness of a translation in terms of *convention-friendliness*:

Convention-Friendliness Principle. If a plausible candidate for being a correct translation f from L_1 to L_2 is a convention-friendly translation, and all the other translations from L_1 to L_2 that are plausible candidates for being correct translations from L_1 to L_2 are not convention-friendly, then this fact is an excellent reason to believe that f is a correct translation.

Determining whether a translation is convention-friendly is in part a matter of determining the beliefs, intentions and desires of the members of the community of speakers of the source language. Two principles that help in this task are Lewis's [33] *rationalisation principle* and *principle of charity*.

In a nutshell, according to the *rationalisation principle* each agent should be represented as rational, in such a way that the agent's physical description, as well as the system of beliefs and desires assigned to him, jointly offer explanations of the agent's behaviour that conform to the canons of decision theory. According to the *principle of charity*, roughly, we should assign to each agent those beliefs that we would have had if we had been exposed to the same evidence and training of the agent, and the same desires that we would have had if we had the agent's beliefs, training and history.[33]

Briefly, the reason why *convention-friendliness* requires that the community of speakers conforms to a convention of truthfulness and trust *only* in the used part of L_1 is the following. Suppose that the requirement was extended to *all* of L_1. Then, a convention-friendly translation would most probably fail to be a correct translation. For a convention-friendly translation may assign the wrong meanings to some of the sentences of the *unused* part of L_1. In particular, it will assign the wrong propositions to at least some of those unused, very long and complicated sentences of the language.

The problem is, as Lewis notes, that if a speaker were to use such strings, then he would not be *trusted*. Rather, he would be understood as 'trying to win a bet or set a record, or feigning madness or raving for real, or doing it to annoy, or filibustering, or making an experiment to test the limits of what is humanly possible to say and mean' [36, p. 108]. For this reason, there will be no convention of truthfulness and trust with respect to the unused, very long and complicated sentences of the language. So, in general, a convention-friendly translation can be expected to be *incorrect* when defined for the *unused* and *cumbersome* sentences of L_1. Members of the community of speakers of L_1 would think that those sentences would not be used truthfully in L_1, and so they would not be trusting.[34]

Also, the Convention-Friendliness principle appeals to a distinction between the *plausible* and the *implausible* convention-friendly translations because, in general, there will be many different convention-friendly translations from L_1 to L_2. Where f is a convention-friendly translation from L_1 to L_2, any mapping g from L_1 to L_2 agreeing with f on the used part of L_1 will be a convention-friendly translation. But not all of those will be plausible.

[33] Lewis [33] puts these principles at work in a strategy for determining an agent's beliefs, desires and meanings on the basis of our complete knowledge of the agent, *qua* a physical system. No such limited knowledge needs to be assumed for the present purposes. The principles are here given simply as extra resources available to the task of determining whether a certain translation is convention-friendly.

[34] The reason why the *Synonymy account* is not committed to the stronger principle according to which a translation is convention-friendly if and only if it is correct has to do with the different problems that have been identified in the literature concerning Lewis's account of what it is for a community to speak a language in terms of the members of the community conforming to a convention of truthfulness and trust in the language. These problems have led us to propose a weaker connection between convention-friendliness and correctness. See Burge [5], Hawthorne [19], O'Leary-Hawthorne [45] and Kölbel [30] for some criticisms of Lewis's account.

One way to make precise the notion of a plausible convention-friendly translation would appeal to naturalness, with some account of what makes a translation more natural than another one. This move would be in agreement with what Lewis [36] says about preferring the straight rather than the bent grammars generating assignments of semantic values for L_1 compatible with there being a convention of truthfulness and trust in the used part of L_1. But there may be other ways. For the present purposes, it is perhaps best to leave the notion of a plausible translation as a primitive. Inquirers aiming to establish the equivalence between theories will often have already selected the mappings which they take to be plausible candidates for being correct translations.

Why is it that the fact that a plausible candidate for being a correct translation scheme f from L_1 to L_2 is a convention-friendly translation, and all the plausible alternative translations from L_1 to L_2 are not convention-friendly, gives only *excellent reason* for believing that f is correct, instead of *implying* that f is correct? The worry here is that there might be no *correct* translation from L_1 to L_2 whatsoever. The existence of one and only one plausible convention-friendly translation f does not rule out this scenario. Despite this, it is difficult to see what sort of evidence would supports the view that there is no correct translation from L_1 to L_2, rather than supporting the view that f is a correct translation from L_1 to L_2.

One important aspect of convention-friendly translations is their *insensitivity* to the sort of *fine-grained* distinctions between meanings that are *irrelevant* for metaphysical equivalence. For instance, even if 'Cambridge is north of NY' and 'NY is south of Cambridge' have different meanings, a mapping sending each sentence of English to itself except that it maps 'Cambridge is north of NY' to 'NY is south of Cambridge' and 'NY is south of Cambridge' to 'Cambridge is north of NY' will count as a convention-friendly translation—assuming that, as used by speakers of English, 'Cambridge is north of NY' and 'NY is south of Cambridge' do require the same of the world for their truth. Another example concerns the sentences 'John runs' and 'John runs or John runs'. Perhaps these have different meanings, at least according to some fine-grained conceptions of meaning (e.g., if meanings are conceived of as structured propositions). Regardless, what matters for determining whether a translation mapping one of these sentences to the other is convention-friendly is simply whether they require the same of the world for their truth—not more fine-grained distinctions between meanings.

Indeed, it is not unreasonable to think that any convention-friendly translation f will be sensitive to differences with respect to whether φ and $f(\varphi)$

require the same or different things of the world for their truth, while being insensitive to more fine-grained differences between φ and $f(\varphi)$. Thus, convention-friendly translations presumably afford a sweet-spot for determining metaphysical equivalences.

Notwithstanding, one can expect that it will be difficult to determine whether a translation is convention-friendly. For this reason it is desirable to have a simple procedure, even if defeasible, for determining whether translations are deeply correct. For this purpose, and drawing inspiration from Hirsch's [22–25] work on verbal disputes, we propose what we call *Hirsch's rule of thumb*.[35] It consists in appealing to judgements concerning the truth of a particular counterfactual hypothesis. For each pair of theories T_1 and T_2, the antecedent of the counterfactual consists in the description of the following (counterfactual) scenario:

> **Disjoint Communities Scenario.** There are two different communities, C_{T_1} and C_{T_2}. In C_{T_1} theory T_1 is acknowledged to be the best theory available, and a vast majority of the members of C_{T_1} know all the intricacies of T_1. In effect, T_1 has become a part of the folk theory of C_{T_1} (what is meant with T being a part of the 'folk theory' of C_T is simply that T is a theory that is implicit in the everyday thought and action of the members of C_{T_1}, just as it is implicit in everyday thought and action that people have beliefs). Furthermore, the sentences of L_{T_1} have as their literal meanings, in the language of C_{T_1}, those meanings that they are used to express by T_1's proponents in its formulation. Similarly for T_2 with respect to C_{T_2}. Also, each of these societies was initially unaware of the existence of the other. Later on, some members mm_{T_2} of C_{T_2} become aware of C_{T_1}, and are given sufficient time to get to know it in detail.

Now, let f be a mapping from L_{T_1} to L_{T_2}. The counterfactual hypothesis is as follows.
Hirschean Counterfactual. If the disjoint communities scenario had obtained, then f would have been a correct translation of a part of the language of C_{T_1} (specifically, of L_{T_1}) to the language of C_{T_2} by mm_{T_2}.
Hirsch's rule of thumb consists in the following claim:

[35] This does not imply that the disputes which Hirsch takes to be verbal turn out to be disputes between equivalent theories. It also does not imply that we agree with Hirsch that what he calls 'common sense ontology' is the correct ontology. We remain neutral on these questions.

Hirsch's Rule of Thumb If the Hirschean counterfactual is true, then f is a deeply correct translation scheme from L_{T_1} to L_{T_2}.

As previously mentioned, the question whether a translation scheme is *deeply correct* may be substituted by the question whether a related translation scheme is *correct*. The focus on communities C_{T_1} and C_{T_2} and their languages makes it possible to shift attention from the non-literal use of L_{T_1} and L_{T_2} to the literal use of the languages of the communities C_{T_1} and C_{T_2}. Furthermore, the fact that, initially, each one of the communities is unaware of the existence of the other makes it possible for the history of disputes between the theories' proponents not to play a role on how the language of each linguistic community is best translated.

To mention the obvious, judgements concerning the truth of the Hirschean counterfactual require some hold on what would constitute a correct translation. This is a place where the convention-friendliness principle and Lewis's principles of rationalisation and charity come into play, as these principles offer some guidance on how to judge the truth of the Hirschean counterfactual. Still, it may turn out to be easier to judge the truth of the Hirschean counterfactual than to use other means to determine whether a given translation is convention-friendly.

4. Applying the Synonymy Account

In this section we show that the *Synonymy account* satisfies the desiderata listed in Section 2, and that it affords a nuanced understanding of the debate between Quineans and noneists.

4.1. Satisfaction of the Desiderata

According to the first of the desiderata laid out in Section 2, an account of metaphysical equivalence should predict some of the conditions under which it is likely for a theory to be received as *absurd* by the proponents of another theory. The *Synonymy account* yields some predictions concerning when this is likely to happen. Furthermore, these predictions very much agree with the diagnosis as to why some Quineans have understood noneists as being committed to an absurdity.

It is reasonable to suppose that any theory whose entailment structure is such that there is a sentence φ which entails every sentence of the language attributes to φ the status of being maximally informative, i.e., of expressing an *absurdity*. Suppose that theories T_1 and T_2 appear to be formulated in the same language (broadly construed), and that T_1 is committed to the truth of a sentence whose homonymous interpretation by the proponents of T_2 is a sentence that, as used by T_2's proponents, expresses an absurdity. In such case the proponents of T_2 will take T_1 to be absurd.

This prediction of the *Synonymy account* can be generalised. The account predicts that a sufficient condition for a theory T_1 to be, on a preliminary interpretation, understood as absurd by the proponents of T_2, when T_1's and T_2's languages are a part of a broader language common to the xx_{T_1}s and the xx_{T_2}s, is that the homonymous interpretation of some of the sentences to whose truth T_1 is committed be sentences that entail some element in $\Vdash^{-}_{T_2}$. For each sentence in $\Vdash^{-}_{T_2}$ is understood by the proponents of T_2 as expressing an absurdity, insofar as this is the set of minimal elements with respect to $\Vdash_{T_2}$'s entailment relation. As we saw in Section 2, Quineans appear to have understood noneists as speaking gibberish for precisely this reason. Noneists are committed to the truth of 'some things do not exist', a sentence which, as used by Quineans, expresses an absurdity. Similar situations may be expected to happen in other debates.

According to the second desideratum, an account of metaphysical equivalence should not have homonymous interpretation as a mandatory facet of the interpretation of a theory by another theory's proponents, even when the two theories' languages are both part of a broader language spoken by a linguistic community that has the proponents of both theories amongst its members. As we have seen in Section 3.3, satisfaction of this desideratum has been written into the *Synonymy account*, via the definition of the relation of *theory synonymy* in terms of *deeply correct* translations, rather than of correct translations.

The third desideratum on accounts of metaphysical equivalence singled out in Section 2 is that an appropriate account should allow for cases in which a theory is intelligible to the proponents of another theory even though the first theory cannot be fully understood in terms of the resources afforded by the second theory. The distinctions introduced in Section 3 afford straightforward ways of explicating *full understanding* and *intelligibility*.

Full understanding may be explicated via *embeddability*:

Full Understanding Theory T_1 is fully understandable in terms of the resources of theory T_2 just in case T_1 is embeddable in T_2.

Thus, to speak in terms of conceptions of logical space, T_1 is *fully understandable* in terms of T_2's resources just in case T_1's conception of logical space is a *part* of T_2's conception of logical space.

According to the notion of *intelligibility* here at play, intelligibility is easy to get. For T_1 to be intelligible by the lights of T_2 it is enough that T_1 not be interpreted as an absurd theory by the lights of T_2 via a deeply correct translation.

It should be clear that it is possible for a theory T_1 not to be embeddable in T_2 while no deeply correct translation maps a commitment of T_1 to a sentence in $\Vdash^-_{T_2}$. In such case T_1 is intelligible, and yet not fully understandable, by the lights of T_2. Thus, the third desideratum on an adequate account of theory equivalence is satisfied.

In fact, in the following sections we will encounter an example that shows that intelligibility does not entail full understanding. Although the noneist theory Non_1 is intelligible in terms of the Quinean theory Qui_1, it is not fully understandable in terms of it, for the Quinean theory lacks the conceptual resources to make sense of some of the commitments of this noneist theory.

According to the fourth desideratum, an account of metaphysical equivalence should explain how theories that would appear to be contradictory if homonymously interpreted are sometimes equivalent, and offer some means of predicting when this will happen. As already remarked, there are cases in which homonymous interpretation is not the correct interpretation of the theories in question. Furthermore, it may happen that two theories that would turn out to be contradictory if homonymously interpreted are such that there are functions f and g establishing their solid similarity. For instance, in the following sections we will argue that the Noneist theory Non_1 and a theory Qui_2 that results from adding more conceptual resources to the original Quinean theory, are in fact solidly similar but would contradict each other if they were homonymously interpreted. In such case the question arises as to whether f and g are both *deeply correct* translation schemes. We can expect this to be the case at least for some pairs of theories, in which case those theories are in fact *synonymous*, and so equivalent.

Furthermore, by coupling the explication of metaphysical equivalence as theory synonymy with an account of what is required for a translation to be deeply correct, and a procedure for determining when this is so—*Hirsch's rule of thumb*—, the *Synonymy account* has the resources for not only explicating theory equivalence and explaining how theories that would appear contradictory if interpreted homonymously are sometimes equivalent, but also for generating predictions concerning when two theories in fact turn to be equivalent.

According to the last of the desiderata on accounts of metaphysical equivalence identified in Section 2, any adequate account of metaphysical equivalence should yield conditions under which translations, such as the one proposed by Woodward, count in favour of the claim that 'there is total overlap between the conceptual resources of the two theories'. Translations between theories' languages are of primary importance for revealing whether they share their entailment structure. Moreover, if these translations are deeply correct, then they establish the theories' *congruence*, and so that there is indeed a 'total overlap' between their conceptual resources. In such case the theories will have "isomorphic" conceptions of logical space. Thus, the *Synonymy account* takes seriously Woodward's considerations involving translations, revealing their importance for determining when two theories totally overlap with respect to their conceptual resources.

4.2. The Noneism versus Quineanism Debate

We have just shown that the *Synonymy account* satisfies the desiderata laid out in Section 2. We now turn to applying it to the debate between noneists and Quineans. We begin by spelling out in some detail simple versions of Noneism and Quineanism—respectively, the theories Non_1, and Qui_1.

The languages of Non_1 and of Qui_1, respectively, L_{Non_1} and Qui_1, are syntactically indistinguishable. Each consists of a first-order modal language with boolean connectives $\rightarrow$ and $\neg$, necessity operator $\Box$, the quantifier $\forall$ (respectively, the noneist's neutral general quantifier and the Quinean's universal quantifier), the (binary) identity predicate $=$, and the monadic predicates E (the noneist's existence predicate), F ('is fictional'), P ('is a "mere possibile"'; i.e., P is satisfied by a thing just in case that thing could have existed but *actually* does not),[36] and M ('is a mathematical entity'). The remaining boolean connectives are defined in the usual way, the same applying to $\Diamond$. The quantifier $\exists$ (respectively, the noneist's neutral particular quantifier and the Quinean's existential quantifier) is defined as the dual of $\forall$ (i.e., $\exists x\varphi =_{\text{df}} \neg\forall x\neg\varphi$). The sets L_{Non_1} and L_{Qui_1}, consist, respectively, in the set of well-formed formulae of Non_1 and the set of well-formed formulae Qui_1.

The characterisations of theories Non_1 and Qui_1 to be given make use of the following axioms and inference rules[37]:

Axioms and Rules of Non_1:

(PL) All propositional tautologies.

(K) $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$.

(T) $\Box\varphi \rightarrow \varphi$.

(5) $\Diamond\varphi \rightarrow \Box\Diamond\varphi$.

[36] Note that 'actually' is being used with its rigidified reading.

[37] We could have appealed to a set of models instead. Nothing hangs on this.

($\mathfrak{A}$1)	$\forall v(\varphi) \to \varphi[v'/v]$.[38]
(=1)	$v = v$.
(=2)	$v = v' \to (\varphi \to \psi)$.[39]
(E–F)	$\forall v(Fv \to \neg Ev)$.
(E–P)	$\forall v(Pv \to \neg Ev)$.
(E–M)	$\forall v(Mv \to \neg Ev)$.
(MP)	$\Vdash_{Non_1} \varphi \to \psi$, $\Vdash_{Non_1} \varphi \Rightarrow \Vdash_{Non_1} \psi$.
(Nec)	$\Vdash_{Non_1} \varphi \Rightarrow \Vdash_{Non_1} \Box\varphi$.
($\mathfrak{A}$2)	$\Vdash_{Non_1} \varphi \to \psi \Rightarrow \Vdash_{Non_1} \varphi \to \forall v(\psi)$.[40]

Axioms and Rules of Qui_1:

(PL)	All propositional tautologies.
(K)	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.
(T)	$\Box\varphi \to \varphi$.
(5)	$\Diamond\varphi \to \Box\Diamond\varphi$.
($\forall$1)	$\forall v(\varphi) \to \varphi[v'/v]$.[41]
(=1)	$v = v$.
(=2)	$v = v' \to (\varphi \to \psi)$.[42]
(EDef)	$\exists v'(v = v') \leftrightarrow Ev$[43]
(MP)	$\Vdash_{Qui_1} \varphi \to \psi$, $\Vdash_{Qui_1} \varphi \Rightarrow \Vdash_{Qui_1} \psi$.
(Nec)	$\Vdash_{Qui_1} \varphi \Rightarrow \Vdash_{Qui_1} \Box\varphi$.
($\forall$2)	$\Vdash_{Qui_1} \varphi \to \psi \Rightarrow \Vdash_{Qui_1} \varphi \to \forall v(\psi)$.[44]

The intended reading of these axioms by, respectively, noneists and Quineans should be clear. Now, where $\bigwedge\Gamma'$ is any conjunction of all the elements in Γ', let

$$\Gamma \Vdash \varphi \quad \text{if and only if there is a finite set } \Gamma' \text{ such that } \Gamma' \subseteq \Gamma, \text{ and } \wedge\Gamma' \Vdash \varphi.$$

Furthermore, let

$$\Vdash_{Non_1} =_{\text{df}} \{\langle\Gamma, \varphi\rangle : \Gamma \Vdash_{Non_1} \varphi\}, \quad \text{and} \quad \Vdash_{Qui_1} =_{\text{df}} \{\langle\Gamma, \varphi\rangle : \Gamma \Vdash_{Qui_1} \varphi\}.$$

---

[38] Provided that v is free for v', where $\varphi[v'/v]$ results from replacing each free occurrence of v in φ by v'.

[39] Where ψ differs from φ at most in having v' free at some places where φ has v free.

[40] Provided that v is not free in φ.

[41] Provided that v is free for v', where $\varphi[v'/v]$ results from replacing each free occurrence of v in φ by v'.

[42] Where ψ differs from φ at most in having v' free at some places where φ has v free.

[43] Where v' is the first variable of the alphabet if v is not, and v' is the second variable of the alphabet otherwise.

[44] Provided that v is not free in φ.

Then, the commitments of Non_1 and Qui_1 are the following[45]

$$Com_{Non_1} =_{\mathrm{df}} \{\varphi : \exists vFv, \exists vPv, \exists vMv \Vdash_{Non_1} \varphi\};$$

and

$$Com_{Qui_1} =_{\mathrm{df}} \{\varphi : \exists vFv, \exists vPv, \exists vMv \Vdash_{Qui_1} \varphi\}.$$

It is worth pointing out that $\exists x \neg Ex$, the statement of noneism (in the mouths of noneists) is one of the commitments of Non_1. Also, note that Qui_1 is an *allist* theory. It is committed to the *existence* of fictional entities, mere possibilia and mathematical objects. That is,

$$\exists x(Ex \wedge Fx), \exists x(Ex \wedge Px), \quad \text{and} \quad \exists x(Ex \wedge Mx)$$

are all commitments of Qui_1. Similarly, the claims that every fictional entity exists, every mere possibilia exists and every mathematical object exists, that is,

$$\forall x(Fx \rightarrow Ex), \forall x(Px \rightarrow Ex), \quad \text{and} \quad \forall x(Mx \rightarrow Ex)$$

are all commitments of Qui_1.

4.2.1. Noneism, Allism and Expressive Resources The homonymous translation of Non_1 to Qui_1 maps each sentence of L_{Non_1} to the syntactically identical sentence of L_{Qui_1}. Lewis notes that homonymous translation has the effect that Quineans take noneism to be absurd. Whereas the sentence $\exists x(\neg Ex)$ (of L_{Non_1}) is one of the commitments of Non_1, the homonymous sentence $\exists x(\neg Ex)$ (of L_{Qui_1}) is such that:

$$\text{For every formula} \quad \varphi \text{ of } L_{Qui_1}, \exists x(\neg Ex) \Vdash_{Qui_1} \varphi.$$

The problem is that, insofar this is the case, proponents of Qui_1 take $\exists x(\neg Ex)$ (of L_{Qui_1}) to express an absurdity. Hence, the *homonymous* translation constitutes an *uncharitable* interpretation of the proponents of Non_1 by the proponents of Qui_1 insofar as inputs to proponents of Non_1 a commitment to an absurdity.

Lewis's [35, p. 29] suggestion with respect to how Non_1 should be interpreted by the proponents of Qui_1 can be understood as the suggestion that the following function f (a restriction of the homonymous function to formulae without occurrences of 'E') affords an appropriate *partial* translation from L_{Non_1} to L_{Qui_1}:

---

[45] Note that we are leaving it open whether the sentences of L_{Non_1} and L_{Qui_1}, and the principles in As_{Non_1} and As_{Qui_1} are *semantically* the same.

1. $f(\neg\varphi)$ is $\neg f(\varphi)$.
2. $f(\varphi \to \psi)$ is $f(\varphi) \to f(\psi)$.
3. $f(\Box\varphi)$ is $\Box f(\varphi)$.
4. $f(\forall v(\varphi))$ is $\forall v(f(\varphi))$.
5. $f(Fv)$ is Fv.
6. $f(Pv)$ is Pv.
7. $f(Mv)$ is Mv.
8. $f(v = v')$ is $v = v'$.

Note that, on this interpretation, the sentences $\exists x(Fx)$, $\exists x(Px)$ and $\exists x(Mx)$ are translated homonymously. Furthermore, we have that

$$\begin{aligned}
&\exists x(Fx) \dashv\Vdash_{Qui_1} \exists x(Fx \wedge Ex), \\
&\exists x(Px) \dashv\Vdash_{Qui_1} \exists x(Px \wedge Ex), \quad \text{and} \\
&\exists x(Mx) \dashv\Vdash_{Qui_1} \exists x(Mx \wedge Ex).
\end{aligned}$$

Thus, proponents of Qui_1 interpret 'there are fictional characters, possibilia as well as mathematical objects' as 'there *exist* things which are fictional characters, possibilia as well as mathematical objects'.

Moreover, the sentences $\forall x(Fx \to \exists y(y = x))$, $\forall x(Px \to \exists y(y = x))$, and $\forall x(Mx \to \exists y(y = x))$ are also translated homonymously. Furthermore, we have that

$$\begin{aligned}
&\forall x(Fx \to \exists y(y = x)) \dashv\Vdash_{Qui_1} \forall x(Fx \to Ex), \\
&\forall x(Px \to \exists y(y = x)) \dashv\Vdash_{Qui_1} \forall x(Px \to Ex), \text{and} \\
&\forall x(Mx \to \exists y(y = x)) \dashv\Vdash_{Qui_1} \forall x(Mx \to Ex).
\end{aligned}$$

Thus, we get the result that, under translation f, the proponents of Qui_1 would describe the proponents of Non_1 as being committed to the claims that fictional, mathematical and merely possible objects *all exist*. That is, the function f affords an interpretation of noneism to Quineans whereby noneists are committed to *allism*.

There are different interpretive hypotheses available. Since the present purpose is to illustrate the workings of the *Synonymy account*, we'll assume that the function f indeed affords a correct, albeit partial, interpretation of L_{Non_1} to the proponents of Qui_1.

Still, a challenge remains: how should proponents of Qui_1 interpret the sentences of L_{Non_1} in which the noneist's existence predicate occurs? In particular, how should proponents of Qui_1 interpret *noneism* (i.e., the sentence '$\exists x \neg Ex$'), as it is expressed in the mouths of noneists?

Arguably, Lewis's (and Priest's) remarks that Quineans lack the expressive resources allowing them to fully understand noneists, applied to theories Non_1 and Qui_1, are correct. That is, arguably, theory Qui_1 *does not* possess the resources required to provide a correct interpretation of $\exists x(\neg Ex)$ and other sentences of L_{Non_1} in which the noneist's existence predicate occurs. Thus, proponents of Qui_1 do not have available the expressive resources to *fully understand* Non_1. In particular, proponents of Qui_1 lack the expressive resources to interpret *noneism*, as this thesis is formulated in L_{Non_1} by noneists. The tools of the *Synonymy account* enable us to state the fact that proponents of Qui_1 do not have available the expressive resources to *fully understand* Non_1 in terms of *embeddability*, specifically, as the fact that Non_1 *is not embeddable* in Qui_1.

As for how proponents of Non_1 should interpret Qui_1, Lewis's remarks in [35, p. 29] suggest that they should do so in agreement with the following function g:

1. $g(\neg\varphi)$ is $\neg g(\varphi)$.
2. $g(\varphi\rightarrow\psi)$ is $g(\varphi)\rightarrow g(\psi)$.
3. $g(\Box\varphi)$ is $\Box g(\varphi)$.
4. $g(\forall v(\varphi))$ is $\forall v(g(\varphi))$.
5. $g(Fv)$ is Fv.
6. $g(Pv)$ is Pv.
7. $g(Mv)$ is Mv.
8. $g(Ev)$ is $v = v$.
9. $g(v = v')$ is $v = v'$.

By contrast with the homonymous translation from L_{Qui_1} to L_{Non_1}, g maps predications of existence to self-identity statements, as revealed by clause 8 of g's definition. It is easy to see that, Qui_1 is a stringent fragment of Non_1 via function g. So, if Lewis's suggestion is right, then g is a deeply correct translation and therefore Qui_1 *is embeddable* in Non_1, and so all the distinctions in Qui_1's conception of logical space are present in Non_1's conception of logical space.

At least part of the nature of the disagreement between the proponents of, respectively, Non_1 and Qui_1 becomes clearer after the previous observations. Proponents of Non_1 endorse the view that there are certain expressive resources, certain distinctions in logical space—e.g., the distinction expressed by 'something does not exist' ($\exists x\neg Ex$), as used by proponents of Non_1—, which are not available in Qui_1.

Proponents of Qui_1 will disagree insofar as they don't think that there are such extra distinctions in logical space. According to them, the purported extra distinctions between ways things might or might not have been posited by noneists are just not there. If, instead, proponents of Qui_1 *accept* the existence of such distinctions, then they must acknowledge that their theory is *deficient* in ways that Non_1 is not, since their own theory is *embeddable* in Non_1.

The diagnosis of the disagreement between noneists and Quineans as a disagreement concerning the truth of 'some things do not exist' is thus shallow. On the one hand, this diagnosis fails to take into consideration the (real) possibility that noneists and Quineans mean different things by sentences such as 'something does not exist'. On the other hand, the diagnosis neglects the fact that one of the main points of disagreement between noneists and Quineans concerns the distinctions required to appropriately describe the world. These theorists are fighting about what distinctions there are in logical space. The *Synonymy account* thus provides tools which enable a more nuanced and better understanding of the debate between noneists and Quineans.

4.2.2. A Different Quinean Theory Consider now a different Quinean theory, Qui_2. Qui_2's language is just like Qui_1's language except for the extra predicate, C, which a thing satisfies just in case it is *concrete*.

Following Linsky and Zalta [37] and Williamson [62], the interest is on a notion of concreteness that is *modally flexible*, in the sense that concrete things, such as trees and tables, could have been non-concrete. Thus, 'non-concrete' is not intended to be synonymous with 'abstract', even though part of what it is to be abstract is to be non-concrete. Paradigmatic examples of concrete things are trees, tables, Kripke and the planet Venus. Paradigmatic instances of non-concrete things are Sherlock Holmes, the number two and the merely possible seventh son of Kripke,[46]

The theory Qui_2 is obtained from Qui_1 by adding the following axioms to those of Qui_1:

$$(\text{C}-\text{F})\ \forall v(Fv\rightarrow\neg Cv). \quad (\text{C}-\text{P})\ \forall v(Pv\rightarrow\neg Cv). \quad (\text{C}-\text{M})\ \forall v(Mv\rightarrow\neg Cv).$$

The inference rules of Qui_2 are the same as those of Qui_1, and Qui_2's entailment relation $\Vdash_{Qui_2}$ is defined in a manner similar to $\Vdash_{Qui_1}$, now having in mind the extra axioms (C−F), (C−P) and (C−M).

The commitments of Qui_2 are the following:

$$Com_{Qui_2}=_{\text{df}}\{\varphi : \exists vFv, \exists vPv, \exists vMv \Vdash_{Qui_2} \varphi\}.$$

[46] A different suggestion, given in Woodward [64], is to augment the language of the Quinean with predicates intended to stand for concreteness and *being actual* with the intended reading of *actual* being one according to which the seventh son of Kripke is not actual but could have been, and translate the noneist's 'exists' by 'is concrete and actual'.

We also want to make it clear that we are not committed to Linsky and Zalta's or Williamson's theories being theories that Quine would endorse (nor are we committed to them being theories that Quine would not endorse). We have labelled theory Qui_2 "Quinean" only insofar as Qui_2 has the formula '$E(x) \leftrightarrow \exists y(x = y)$' among its commitments—a formula that would appear to capture the thesis of Quineanism (as it has been formulated in §1 and in §2) in the formal language.

Now, let f' be a mapping from L_{Non_1} to L_{Qui_2} obtained from f by adding

$$f'(Ev) \,\text{is}\, Cv.$$

Also, let g' be a mapping from L_{Qui_2} to L_{Non_1} obtained from g by adding the following clause:

$$g'(Cv) \,\text{is}\, Ev.$$

It should be clear that Non_1 and Qui_2 are solidly similar via f' and g'. This does not suffice to establish the *synonymy* between Non_1 and Qui_2, since whether these theories are synonymous depends on the existence of a pair of *deeply correct translations* which witness the theories' solid similarity. But let us suppose, for the present purposes, that the functions f' and g' are indeed deeply correct translations. In such case Non_1 and Qui_2 are synonymous via f' and g'. So, according to the *Synonymy account*, Non_1 and Qui_2 are *metaphysically equivalent.*

Even if f' and g' are inded deeply correct translations, and so Non_1 and Qui_2 turn out to be metaphysically equivalent, from this it should *not* be concluded that *noneism just is allism.* As previously shown, Non_1 and the *original* quinean theory Qui_1 are arguably *not* equivalent, since Qui_1 does not even appear to have the expressive resources required for a deeply correct translation of the claim of *noneism* (in the mouths of proponents of Non_1), i.e., of '$\exists x \neg Ex$'. Yet, Qui_1 and Qui_2 would typically both be counted as *allist* theories. Hence, the claim that noneism just is allism requires qualification, since some theories that typically count as *allist* do not even possess the expressive resources to express *noneism.*

The Synonymy Account reveals that it is often more useful to focus on *theories* rather than their labelling according to one or another *slogan* (such as labelling them as "Noneism", "Quineanism" and "Allism"). Suppose that S_1 (e.g., Quineanism) is thought to have the drawback of possessing insufficient expressive resources in comparison to those of S_2 (say, Noneism). Suppose that S_1ers then show how, by appealing to certain extra primitives, they may avoid the objection that S_1 has insufficient expressive resources, thus allegedly revealing that S_1 is a relevant alternative to S_2.

The *Synonymy account* affords the resources to see the ways in which this dialectic is misguided. To begin with, when noneists argue that allism is not satisfactory on the basis of insufficient expressive resources, this is best understood as an argument not against allism itself, but instead against a certain theory, or family of theories (at least apparently committed to allism). In addition, by resorting to extra primitives allists in effect express their adherence to theories that are *different* from the ones they started with. Those theories may indeed be better than the ones they started with, and allists may be right in changing their minds. But they are different (and nonequivalent) theories nonetheless.

Finally, if the starting theory under consideration is Qui_1, the rival noneist theory is Non_1, and the improved theory is Qui_2, then the allist will be wrong in claiming that Allism is still a relevant alternative to Noneism on the basis that Qui_2 does not lack expressive resources when compared to Non_1. Given the assumptions presently in play, Qui_2 and Non_1 are synonymous theories, and so *metaphysically equivalent*. Thus, to characterise the theories Qui_2 and Non_1 as alternatives insofar as one of them is an allist theory whereas the other is a noneist theory is to *mischaracterise* the situation. What proponents of Qui_2 mean with 'some things do not exist' is different from what proponents of Non_1 mean with 'some things do not exist', and so the theories are not in conflict.

Relatedly, observe that the translations f' and g' are not homonymous. The sentence 'some things do not exist' ($\exists x \neg Ex$), as used by Quineans$_2$, gets translated by g' into 'there is something that is self-distinct' ($\exists x \neg (x = x)$). This sentence is absurd by the lights of both theorists. However, 'some things do not exist' ($\exists x \neg Ex$), as used by noneists$_1$, gets translated by f' into the Quineans$_2$' non-homonymous sentence 'some things are not concrete' ($\exists x \neg Cx$); and just as noneists$_1$ don't take $\exists x \neg Ex$ to be absurd, Quineans$_2$ don't take 'some things are not concrete' to be absurd. The more general observation is that by augmenting the expressive resources of a theory, the enriched theory may turn out to be metaphysically equivalent to a theory which was previously a rival.

How would the *Synonymy account* recommend that the question whether the functions f' and g' indeed constitute deeply correct translations, and so whether Non_1 and Qui_2 are indeed metaphysically equivalent, be addressed? The account recommends the use of *Hirsch's rule of thumb*. So, consider two societies Soc_{Non_1} and Soc_{Qui_2}. To make the case rather extreme, imagine that Soc_{Non_1} and Soc_{Qui_2} descend from two different populations of English speakers, which, due to some cataclysmic event, were forced to move to distinct and far away planets. Society Soc_{Non_1} is constituted by the descendants of one of those populations, whereas Soc_{Qui_2} is constituted by the descendants of the other. Now, suppose that:

1. Soc_{Non_1} and Soc_{Qui_2} developed for ages without having any contact with each other;
2. In each of these planets some event took place that led to the destruction of most of the knowledge concerning the origins of the society inhabiting

it, in such a way that the society's current members are all unaware of the fact that they travelled from the Earth to their current planet, and that other inhabitants of planet Earth had to move to a different planet;

3. Theory Non_1 becomes part of the folk theory of Soc_{Non_1}, and theory Qui_2 becomes part of the folk theory of Soc_{Qui_2};
4. At some point in their histories Soc_{Non_1} and Soc_{Qui_2} both developed the technological means to send tripulated missions to space, in search of alien life;
5. Some members mm_{Qui_2} of Soc_{Qui_2} manage to travel to the planet where Soc_{Non_1} is based, and to interact with the inhabitants of Soc_{Non_1}.

The scenario just described is one corresponding to the antecedent of a Hirschean counterfactual. According to *Hirsch's rule of thumb*, f' is a deeply correct translation from L_{Non_1} to L_{Qui_2} just in case, if the scenario described had obtained, then f' would have been a correct translation of a portion of the language of Soc_{Non_1} (that corresponding to L_{Non_1}) by mm_{Qui_2}.

To determine whether this is so, the question to be considered is whether f' affords an interpretation of the language of Soc_{Non_1} whereby the members of Soc_{Non_1} turn out to conform to a convention of truthfulness and trust in their language by the lights of mm_{Qui_2}. In the vast majority of cases in which the inhabitants of Qui_2 would assent to sentences such as the sentence 'some fictional character, α, does not exist and ...', mm_{Qui_2} would describe them as believing the truth of 'there exists a fictional character, α, that is not concrete and ...', and as intending to communicate its truth to others. Moreover, this generalises to the different sentences of the language of Soc_{Non_1} for which f' is defined. If this is correct, and there are no other plausible alternative translations, then it should indeed be concluded that the Hirschean counterfactual is true.

To make things more dramatic, we can even conceive mm_{Qui_2} returning to their planet, publishing a translation manual based on f', and this translation manual being used by other members of Soc_{Qui_2} in their visits to Soc_{Non_1}. We can also conceive the possibility of some of these other members of Soc_{Qui_2} at some point also becoming members of Soc_{Non_1}, quickly becoming speakers of the language of Soc_{Non_1}. Arguably, all this may be conceived as being the case without the members of Soc_{Qui_2} and Soc_{Non_1} ever questioning the adequacy of the translation manual based in f'. To be sure, they may question whether there are other translations that better preserve certain hyperintentional aspects of meaning. But it seems reasonable to think that they would maintain that a sentence and its translation afforded by f' require the same of the world in order to be true.

If all this is correct, then the Hirschean counterfactual is indeed true about f'. That is, it is true that if the scenario described had obtained, then f' would have been a correct translation of the language of Soc_{Non_1} by mm_{Qui_2}. Furthermore, a symmetric case may also be considered, with members mm_{Non_1} of Soc_{Non_1} visiting Soc_{Qui_2}. Symmetric considerations would lead to judge as true the claim that if this counterfactual scenario had obtained, then g' would have been a correct translation of the language of Soc_{Qui_2} by mm_{Non_1}. Thus, by *Hirsch's rule of thumb*, f' and g' are deeply correct translations. Given that Non_1 and Qui_2 are strongly similar via f' and g', Non_1 and Qui_2 are synonymous theories. Given the explication of metaphysical equivalence as theory synonymy, Non_1 and Qui_2 are metaphysically equivalent.

Now, it is important to bear in mind that the theories that have been proposed in connection with the Noneism–Allism debate are more nuanced than Non_1, Qui_1 and Qui_2. For this reason, we do not want to give much importance to the fact that Non_1 and Qui_2 are, arguably, synonymous, and so metaphysically equivalent. The aim of the present discussion has been solely that of offering an example of an application of the *Synonymy account*.

To conclude, in this section it was shown that the *Synonymy account* satisfies the desiderata laid out in Section 2. The account was also shown to offer tools enabling a deeper understanding of metaphysical debates. On the one hand, the *Synonymy account* makes salient the fact that sometimes what is at issue between rival theories is whether to accept certain distinctions in logical space. On the other hand, the account rightly changes the focus of debates from labelling according to slogans to theories.

5. Objections and Replies

In this section we compare the *Synonymy account* to two other recent approaches to metaphysical equivalence, specifically, McSweeney's *epistemic account* and Miller's *hyperintensional account*, and show why the *Synonymy account* is preferable to both. Contra the *epistemic account*, we argue that the existence of a common definitional extension of two theories is not a necessary condition for their equivalence. The counterexamples we introduce also cast doubt on the Unified Perspective Condition, the main thesis of the *epistemic account*. Against the *hyperintensional account*, we argue that it yields wrong predictions vis-à-vis the metaphysical equivalence of some theories, owing to its insensitivity to theories' conceptions of logical space.

5.1. McSweeney's Epistemic Account of Metaphysical Equivalence

McSweeney [41] defends the following *epistemic account* of metaphysical equivalence:

Unified Perspective Condition (UPC): 'in order for us to be justified in claiming that two theories, T and T', are equivalent, there must be an occupiable *perspective* from which T and T' can be conceived of as a single unified theory, $T+$, which (in some to-be-determined sense) *says nothing over and above* either T of T', and which *says everything* that T and T' do' (p. 273).

Furthermore, McSweeney argues that a good technical explication of the UPC can be gotten by appealing to the notion of a *common definitional extension** (CDE*). Roughly, $T+$ is a CDE* of theories T and T' just in case: (i) T, T' and $T+$ are all first-order languages; (ii) the language of $T+$ extends the language of both T and T'; (iii) $T+$ is a *conservative extension* of both T and T'[47] and (iv) the primitive expressions of $T+$ are all explicitly definable in terms of both the expressions of T and the expressions of T'.[48]

---

[47]A theory T_1 is a *conservative extension*; of a theory T_2 iff $Com_{T_1} \Vdash_{T_1} \varphi \Leftrightarrow Com_{T_2} \Vdash_{T_2} \varphi$ for all formulae φ of L_{T_2} (fleshed out in terms of our distinction between *entailments* and *commitments*).

[48]Here are the definability requirements (fleshed out in terms of *entailments* and *commitments*):

1. For every n-ary relation symbol R of L_{T+} there is a formula φ of L_T ($L_{T'}$) such that
$$Com_{T+} \Vdash_{T+} \forall x^1 \ldots \forall x^n (Rx^1 \ldots x^n \leftrightarrow \varphi(x^1, \ldots, x^n));$$
2. For every n-ary function symbol f of L_{T+} there is a formula φ of L_T ($L_{T'}$) such that
$$Com_{T+} \Vdash_{T+} \forall x^1 \ldots \forall x^n \forall y (f(x^1 \ldots x^n) = y \leftrightarrow \varphi(x^1, \ldots, x^n, y));$$
3. For every constant symbol c of L_{T+} there is a formula φ of L_T ($L_{T'}$) such that
$$Com_{T+} \Vdash_{T+} \forall x (x = c \leftrightarrow \varphi(x));$$
4. For every n-ary operator O of L_{T+} and all formulae φ^1_+, ..., φ^n_+ of L_{T+} there is a formula φ of L_T ($L_{T'}$) such that
$$Com_{T+} \Vdash_{T+} \forall x^1 \ldots \forall x^n (O(\varphi^1_+(x^1), \ldots, \varphi^n_+(x^n)) \leftrightarrow \varphi(x^1, \ldots x^n));$$

McSweeney's proposed technical explication of the UPC in terms of the notion of a CDE* is as follows:

CDE* as a Unified Perspective: in order for us to be justified in claiming that two theories, T and T', are equivalent, T and T' must have a CDE*.

Arguably, McSweeney's main reason for thinking that the existence of a CDE* falls short of being a *sufficient* condition for the metaphysical equivalence of theories T and T', when these are first-order theories, is that T and T''s CDE* may be formulated in terms which fail to be *fundamental* or "*joint-carving*".[49] But the existence of a CDE* *fails* to be a sufficient condition for the metaphysical equivalence of two first-order theories *even* when their CDE* is formulated in fundamental or joint-carving terms.

Sider [56, p. 179] offers a nice counterexample. Consider two theories, both formulated in classical first-order logic, one of which has as its sole axiom 'Some molecule is part of some electron' ($\exists x(Mx \wedge \exists y(Ey \wedge P(x, y)))$, while the other has as its sole axiom 'some electron is part of some molecule' ($\exists x(Mx \wedge \exists y(Ey \wedge P(y, x)))$. These theories have a CDE*. Yet, surely, they are *not* equivalent. Their truth clearly makes different demands on the world. Furthermore, even if it is rejected that the theories are formulated in joint-carving terms, it should be clear how Sider's example constitutes a *template* from which to generate, in a fundamental language, counterexamples to the view that the existence of a CDE* is a sufficient condition for metaphysical equivalence. The lesson is that whether two theories are metaphysically equivalent is a function of more than just the structural commonalities of the devices through which they are formulated.

Does McSweeney's official view—that the existence of a CDE* is a necessary condition for believing that two theories are metaphysically equivalent—fare better? One obvious limitation of McSweeney's view is that it applies only when the theories in question are all formulated in a first-order language. But surely there can be metaphysically equivalent theories formulated in terms of plural languages, higher-order modal languages and many-sorted languages.

---

5. For every quantifier Q of L_{T+} and all formulae φ_+ of L_{T+} there is a formula φ of L_T ($L_{T'}$) such that

$$Com_{T+} \Vdash_{T+} \forall x(Q\varphi_+(x) \leftrightarrow \varphi(x)).$$

[49] For instance, McSweeney [41, p. 227] points out in a parenthetical remark that the UPC 'is only, and only should be for the hardcore realist, a necessary condition for a justified belief in equivalence—perhaps some will want to treat it as also sufficient, but hardcore realists should not'.

For instance, one promising application of metaphysical equivalence is to theories formulated in, respectively, plural logic and monadic second-order logic. Imposing the existence of a CDE* as a necessary condition for being justified in believing that two theories are metaphysically equivalent does not allow for the comparison of theories formulated in plural and higher-order languages. Yet, there are pairs of theories formulated in such languages which are *clearly* metaphysically equivalent. For a theorist may just *decide* to use the language of monadic second-order logic to speak plurally (as is done in, e.g., [62, Ch. 5.8]). The restriction to first-order languages is thus no negligible cost, in particular given the advent of plural and higher-order languages.

A related problem is that McSweeney's proposed precisification of the UPC presupposes that the quantifiers of metaphysically equivalent first-order theories that have a CDE* range over the *same* domain of things. But theories may be metaphysically equivalent even if their quantifiers range over distinct domains (in this section we offer an example of two metaphysically equivalent theories whose respective quantifiers range over distinct domains—in the form of theories T_I and T_S, with T_I quantifying over urelements and T_S quantifying over singletons of urelements). Furthermore, and importantly for the present purposes, in some cases metaphysically equivalent theories whose quantifiers range over different domains only have CDEs* that say things *over and above* what is said by those theories. Consequently, the view that the existence of a CDE* is a necessary condition for two theories to be metaphysically equivalent fails to afford a good precisification of the UPC, since the unification afforded by a CDE* says *more than* what is said by the theories it was supposed to "unify".

For instance, consider two theories, T_I and T_S, both officially formulated in first-order languages with identity (for clarity, the predicates, operators and quantifiers of these languages will be subscripted). T_I's quantifiers range over the individuals (urelements) whereas T_S's quantifiers range over the unit sets of the actually existing individuals. The only predicates of each theory are their corresponding identity predicates. $\ulcorner t =_I u \urcorner$ states that t and u are identical *individuals*, whereas $\ulcorner t =_S u \urcorner$ states that t and y are *unit sets* with identical members. Furthermore, each language possesses a single individual constant, respectively, 'a_I' and 'a_S'. 'a_I' refers to an individual whereas 'a_S' refers to *the unit set* of the individual that 'a_I' refers to. T_I's and T_S's entailment relations are those of first-order logic (in their respective languages), and they have no distinctive commitments. That is, they are committed to the first-order consequences, in their respective languages, of the empty set.

It is straightforward to set up mappings $f : L_{T_S} \to L_{T_I}$ and $g : L_{T_I} \to L_{T_S}$ witnessing the *sameness of theoretical structure* of T_I and T_S. It suffices to change each one of the language's subscripts. Then, f arguably maps each sentence of L_{T_S} to a sentence of L_{T_I} that requires the same of the world for its truth. *Mutatis mutandis* for g. For instance, what $\exists x_I(x \neq_I a_I)$ ('there is some individual other than a_I') requires of the world for its truth is the same as what $\exists x_S(x \neq_S a_S)$ ('there is some unit set such that the individual that belongs to it is distinct from the individual (a_I) that belongs to a_S') requires for its truth. Thus, the theories are arguably *synonymous*, and so, on our view, *metaphysically equivalent*.

Yet, finding a CDE* for T_I and T_S is arguably not the best way of unifying these theories. Prima facie, both theories' commitments are all true. But suppose that T_I and T_S have a CDE* T_{I+S}. Then,

$$\exists y_S(x =_S y)$$

is a commitment of T_{I+S}, since

$$\exists y_S(x =_S y)$$

is a commitment of T_S. In such case, by universal generalization,

$$\forall x \exists y_S(x =_S y)$$

is a commitment of T_{I+S}. Moreover,

$$\forall x \exists y_S(x =_S y) \to \exists y_S(a_I =_S y)$$

is a commitment of T_{I+S}. So,

$$\exists y_S(a_I =_S y)$$

is a commitment of T_{I+S}. But

$$\exists y_S(a_I =_S y)$$

is false, for a_I is *not* a *set*. Thus, no CDE* of T_I and T_S amounts to a single, unified theory which says nothing over and above either T_I or T_S. For all CDEs* of T_I and T_S must perforce say false things whereas T_I and T_S say only true things.

The example may be resisted. Perhaps McSweeney would find the inference in T_I and T_S's CDE* from $\exists y_S(x =_S y)$ to $\forall x \exists y_S(x =_S y)$ to be illegitimate. A possible reply is that $\exists y_S(x =_S y)$ is a commitment of T_{I+S} only when 'x' is a variable of T_S. That is, perhaps the quantifiers of L_I and L_S's CDE* must come with their dedicated variables, distinct from the variables of both T_{L_I} and T_{L_S}.

We are unsure whether the view that the quantifiers of L_I and L_S come with their own quantifiers calls into question McSweeney's conception of two theories' CDE* as constituting a unification of those theories. Be that as it may, there is a related worry.

Any CDE* T_{I+S} of T_I and T_S must have $\ulcorner\forall x(x = a_I \leftrightarrow \varphi)\urcorner$ amongst its commitments, for some formula φ of L_S. But what could φ be? The only relevant alternative is '$x =_S a_S$'.[50] Now, we have that

$$\forall x(x = a_I \leftrightarrow x =_S a_S) \rightarrow (a_I = a_I \leftrightarrow a_I =_S a_S)$$

is a commitment of T_{I+S}. So,

$$\forall x(x = a_I \leftrightarrow x =_S a_S)$$

is a commitment of T_{I+S} only if

$$a_I = a_I \leftrightarrow a_I =_S a_S$$

is a commitment of T_{I+S}. Moreover,

$$a_I = a_I$$

is a commitment of T_{I+S}. So,

$$a_I =_S a_S$$

is a commitment of T_{I+S}. But $a_I =_S a_S$ is false. After all, a_I is *not* a *set*. So, any CDE* of T_I and T_S says some things over above those said by either T_I or T_S.

The main feature of the counterexamples is their use of the fact that one can say things about entities in a domain via expressions whose semantic values are entities pertaining to a different domain. While this idea finds one major exponent on abstractionist philosophy of mathematics, in

[50]To be sure, there is a different alternative, specifically, $x \neq_S a_S$. But that cannot be. If that were the case, then $\neg\exists y_I \exists z_I(y \neq_I z \wedge y \neq_I a_I \wedge z \neq_I a_I)$ would be a commitment of T_{I+S}, and so $\neg\exists y_I \exists z_I(y \neq_I z \wedge y \neq_I a_I \wedge z \neq_I a_I)$ would be a commitment of T_I, since T_{I+S} is a conservative extension of T_I. Yet, $\neg\exists y_I \exists z_I(y \neq_I z \wedge y \neq_I a_I \wedge z \neq_I a_I)$ is not a commitment of T_I. Contradiction.

The argument from claim that $\forall x(x = a_I \leftrightarrow x =_S a_S)$ is a commitment of L_{I+S} to the claim that $\neg\exists y_I \exists z_I(y \neq_I z \wedge y \neq_I a_I \wedge z \neq_I a_I)$ is a commitment of T_{I+S} relies on universally instantiating L_{I+S}'s dedicated universal quantifier with variables pertaining to L_I (assuming that only these can be bound by the quantifiers of L_I). But presumably *this* move is allowed. Otherwise, the logic of L_{I+S}'s dedicated universal quantifier can't be classical first-order logic. Furthermore, what would the unification afforded by a CDE* of two theories be if such inferences were not allowed?

particular on the view that the left-hand side of Hume's principle's biconditional constitutes a *recarving* of its right-hand side,[51] there are myriad more innocuous examples. Not only can we talk about urelements through expressions whose semantic values are unit sets, we can also talk about points through expressions whose semantic values are lines, and about individuals through expressions whose semantic values are haecceities, to mention just two other examples.

Thus, as the counterexamples show, unifying in the manner required by the existence of a CDE* wrongly presupposes that T's predicates, truly applicable to things in T's domain of quantification, are also truly applicable to things in T''s domain of quantification. But this may fail to be so, even when the sentences of each theory can be translated to sentences of the other theory in a way witnessing their metaphysical equivalence.

Now, McSweeney claims that the existence of a CDE* of two theories rules out the possibility of those theories being formulated in terms of equally joint-carving "unrestricted" quantifiers, and so witnessing "quantifier-variance". So, are the quantifiers of T_I and T_S equally joint-carving unrestricted quantifiers? Prima facie, they are not. For it is plausible to think that the quantifiers of T_I and T_S are restrictions of a more encompassing quantifier, and so are not unrestricted in the relevant sense. Quantifier variance is not the issue.[52]

So far, we have shown that the existence of a CDE* is inapplicable in interesting cases of metaphysical equivalence—specifically, it is inapplicable when the theories are formulated in languages other than first-order languages. We have also shown that the view that the existence of a CDE* is a necessary condition for being justified in believing that two theories are metaphysically equivalent sometimes fares poorly as a technical precisification of the UPC. What about the UPC itself? May it be true nonetheless?

It is rather unclear what the UPC amounts to, in particular the idea of a "unified" theory. Still, the technical precisification in terms of the notion of a *common definitional extension** makes it reasonable to think that part of what the idea of "unification" involves is the existence of a theory which unifies the *vocabularies* of T and T' in a *single language*, says nothing over above either T or T', and says everything said by T and T'.

---

[51] See, e.g., Wright and Hale [66].

[52] We have mentioned joint-carvingness here only by way of addressing a possible objection to our views. We ourselves take no stand on the issue whether some but not all quantifiers are joint-carving.

When two theories T and T' are *synonymous*, it is rather straightforward to find a theory $T+$ whose language is a *superlanguage* of both T and T', says nothing over and above either T or T', and says everything said by T and T'. Just union T and T''s languages, union their commitments and close their entailments under translation. The resulting theory will say nothing over and above either T or T', will say everything that is said by T and T', and will be formulated in a "common language". In *this* sense, T and T''s *synonymy* trivially guarantees the existence of a single theory that "unifies" T and T'.

If that is not enough for *unification*, what is? Arguably, another idea behind the appeal to the notion of a *common definitional extension** in offering a technical precisification of the UPC is that the language of any theory that unifies T and T' must be *compositionally determined* solely in terms of logical expressions and the primitive (nonlogical) expressions of the languages of T and T'.

McSweeney could reply that the Synonymy account is equivalent to her account in virtue of a recent result by Barrett and Halvorson [3]. According to them, two theories (with disjoint signatures) are intertranslatable iff they are definitionally equivalent (i.e., iff they have a CDE). At first sight, the notion of intertranslatability they make use of seems like the notion of Theory Similarity we gave in Section 3.2.1. If the two notions are equivalent, then, by this result, any objection to McSweeney's account could perhaps be reformulated as an objection to the Synonymy account.

However, there is an important difference between the two approaches. As stated in the definition of Theory Similarity, the Synonymy account does not assume a *compositionality* constraint, in contrast to the previously mentioned result. Barrett and Halvorson define intertranslatability in terms of the Quinean notion of a *reconstrual.* A reconstrual is a mapping η from the signature of some first-order language to the formulas of the other language which has to satisfy additional conditions. The reconstrual induces by recursion another map that 'translates' the formulas of the first language to the formulas of the second language. The recursive definition requires that the translations in the second language of the formulas of the first language are obtained compositionally by preserving the effect of the logical connectives and quantifiers.

Keeping things simple, consider the following example (Figures 15 and 16):

In the cases of both Cl and DevCl (*deviant* classical logic) $[\varphi]$ consists of the equivalence class of φ under classical mutual entailment. The two natural candidate reconstruals η_1 and η_2 from the signature of DevCl to the signature of Cl are such that $\eta_1(B) = A$ and $\eta_2(B) = \neg A$. These induce mappings F_1 and F_2 from the formulae of DevCl to the formulae Cl such that $F_1(B) = A$ and $F_1(\neg B) = \neg A$, and $F_2(B) = \neg A$ and $F_2(\neg B) = \neg\neg A$. Mapping F_1 will not witness the intertranslatability of Cl and DevCl, owing to the fact that $\neg B \Vdash_{\mathrm{DevCl}} B$, even though $F_1(\neg B) = \neg A \nVdash_{\mathrm{Cl}} A = F_1(B)$. Mapping F_2 will not witness the intertranslatability of Cl and DevCl, owing to the fact that $\neg B \Vdash_{\mathrm{DevCl}} B$, even though $F(\neg B) = \neg\neg A \nVdash_{\mathrm{Cl}} \neg A = F(B)$. So, while Cl and DevCl are similar, they are not intertranslatable.

Since our notion of Theory Similarity does not impose a compositionality constraint, it is more generous. It allows for theories to be similar while not being common definitional extensions. But that is not a problem for our account. As the example shows, it is similarity, not intertranslatability, which is on the right track.

Our counterexample to McSweeney's technical precisification of the UPC in terms of *common definitional extension** raises serious doubts concerning the prospects of the UPC when 'unification' is glossed in this manner. Furthermore, we are typically able to determine the equivalence between two theories without *establishing* the existence of any theory that consists of their unification, in the sense of 'unification' now being examined. For instance, among the paradigmatic cases of theories whose equivalence one may be interested in ascertaining are theories formulated in different languages.

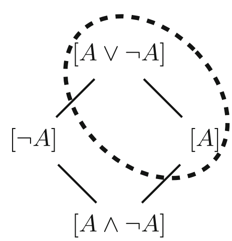

Figure 15. Theoretical structure of Cl

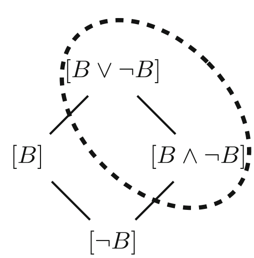

Figure 16. Theoretical structure of DevCl

In such cases the problem of determining whether T and T' are equivalent typically amounts to a problem of *translation*. While being able to speak T's and T''s languages is no doubt helpful, demonstrating the existence of a *third* theory in a language that is a superlanguage of T's and T''s languages just seems besides the point.

Furthermore, it is doubtful whether, in general, such "unificatory" languages exist or even possibly exist. For instance, if T is formulated in English while T' is formulated in German, is a necessary condition for being justified in believing that T and T' are equivalent that there be a language whose complex expressions are compositionally determined from the primitive expressions of English and German? Not only does this view seem absurd, it is also doubtful that there is or even could be such a language. After all, English and German have substantially different sentential structure.

Moreover, even if no such language exists, and so no theory unifies the English and German theories, on the sense of 'unification' presently being examined, the lack of such a "unificatory" theory in no way reveals that one cannot be justified in believing that a theory formulated in English is metaphysically equivalent to a theory formulated in German. Rather, if this is the sense of 'unification' in question, then it is the UPC that must be rejected, not the possibility of being justified in believing that some theory formulated in English is equivalent to some theory formulated in German.

A final remark on the UPC. The condition talks about there being a perspective from which two theories can be *conceived* of as a single unified theory. But the existence of a CDE* is not enough for us to be able to *conceive* two theories as unified, even bracketing all the problems previously mentioned for the precisification of the UPC in terms of the existence of a CDE*. Arguably, what does afford a perspective from which T and T' can be *conceived of* as a *single* unified theory is, at best, the *metatheory* in which T and T' are shown to have a CDE*.

Furthermore this metatheory need not, and typically will not, be formulated in a language that is a super-language of the languages of T and T'. The metatheory also need not say all the things said by either T or T', and it will most probably say things over and above T and T' (e.g., it will say that T and T' stand in some interesting relations). Overall, the metatheory need not be a *unification* of T and T', in the sense of 'unification' presently being examined.

McSweeney worries that if there is no theory that is the unification of two metaphysically equivalent theories, then nothing justifies us in believing in the theories' metaphysical equivalence. But why think so? Simple cases of translation between languages reveal equivalences between theories without any need of positing further theories unifying the previous ones. For instance, theories having as their commitments, respectively, the sentences 'Biden is the US president' and 'Biden es el presidente de los EEUU' are, all things being equal, metaphysically equivalent. We are capable of so ascertaining even if there is no language that "unifies" English and Spanish.

So, what justifies us in believing that theories are equivalent? In part, what justifies us are facts about the languages in which the theories are formulated, how these languages are used by the respective theories' proponents and whether that use makes it so that the same is meant by particular sentences of those languages. It is information bearing on these issues that is directly relevant to determine metaphysical equivalences—not information about whether languages can be "unified".

Moreover, we typically have available plenty of information of the relevant kind, even in those hard cases in which we remain unsure whether the theories in question are metaphysically equivalent. By contrast with the UPC, this is precisely the sort of information that the *Synonymy account* is sensitive too. Thus, the *Synonymy account* is arguably much better placed than the UPC for guiding our inquiry on whether particular theories are metaphysically equivalent.

5.2. Miller's Hyperintensional Account of Metaphysical Equivalence

Miller [43] has proposed a *hyperintensional* approach to metaphysical equivalence. Her starting point is the following claim:

Sameness of Meaning: Two theories are metaphysically equivalent if and only if they have the same meaning.

Miller's idea is that, on some conception of meaning, *sameness of meaning* is true. This reduces the problem of offering an account of metaphysical equivalence to that of ascertaining which conception of meaning renders true *sameness of meaning*.

Many metaphysical theories are thought to be necessary. If true, they are necessarily true. If false, they are necessarily false. So, if we take the meaning of a metaphysical theory to be given by its *intension*, the set of possible worlds where the theory is true, then *sameness of meaning* implies that any two necessarily true metaphysical theories are not only *intensionally* but also *metaphysically* equivalent. This is unsatisfying. Suppose that *Permanentism*, understood as the thesis that always, everything always exists, and

Constitution is Identity, understood as the thesis that for one thing to constitute another is for the two things to be identical, are both true,[53] Plausibly, if *permanentism* and *constitution is identity* are true, they are necessarily so. But, arguably, they do not mean the same. More importantly, *permanentism* and *constitution is identity* ought not to count as metaphysically equivalent.

Thus, Miller argues, in order to distinguish the meanings of any two such theories we should appeal to *hyperintensions*. These constrast with intensions in that hyperintensionally equivalent expressions, but not intensionally equivalent expressions, can be substituted by one another *salva veritate* even in contexts induced by propositional attitudes such as belief.

Miller's appeal to hyperintensions prompts the following worry. A philosopher may believe that a theory T is true without believing that another theory T^* is true, despite the fact that both theories are equivalent. This is reasonable. After all, we do disagree on which theories are equivalent. So, aren't hyperintensions *too* fine-grained for metaphysical equivalence?

Miller avoids this worry by distinguishing between *strong hyperintensionality* and *weak hyperintensionality*. Whereas strongly hyperintensional distinctions are those that allow us to distinguish between the *genuine* ways in which things could or could not be, merely weakly hyperintensional distinctions are due entirely to features of our representational system. Therefore, a philosopher can believe that T is true but T^* isn't, despite the fact that T and T^* are strongly hyperintensionally equivalent, because T and T^* are not weakly hyperintensionally equivalent.

Miller proposes to model *strong hyperintensionality* via impossible worlds. A theory T's *strong hyperintension* consists of the set of possible and impossible worlds at which all of T's commitments are true. Two theories are strongly hyperintensionally equivalent if and only if their commitments are jointly true at exactly the same possible and impossible worlds. Equipped with these notions, Miller offers the following account of metaphysical equivalence:

ME: Metaphysical theories T and T^* are *metaphysically equivalent* iff T and T^* are strongly hyperintensionally equivalent.

[53] For more on *permanentism* see Sullivan [57], Williamson [62], Cameron [6] and Deasy [13]. For more on *constitution is identity*, see Johnston [28], Noonan [44] and Baker [2].

ME is not Miller's *full* account of metaphysical equivalence. In her view, before considering ME we should already have a criterion for inter-translatability:

> 'Indeed, Miller [42] suggests that if two theories are correctly inter-translatable—if there is a function that maps the sentences of one theory onto the sentences of the other theory in a way that is a correct translation—then said theories are metaphysically equivalent. The idea is that if there really is a correct translation, then the two theories are saying the very same thing, and hence are metaphysically equivalent. *Holding fixed that we know what a correct translation is*, the question then becomes on what grounds we could reasonably come to believe that a purported translation is correct.' (p.2, our emphasis)

Furthermore, in Miller [42] she offers the following account of correct inter-translatability:

> 'So let us say that we have a correct translation between theories just if there is an assertibility mapping that is truth preserving and where it preserves truth in virtue of the same truth-makers.' (p. 2)

By being formulated in terms of accessibility mappings, Miller's account of correct inter-translatability is akin to our notion of a deeply correct translation. Notwithstanding, in our view ME falls short of providing a correct account of metaphysical equivalence. Recall that, according to Rayo [52], any theoretical inquiry involves a hypothesis regarding the distinctions that compose logical space—i.e., a conception of logical space. A sentence makes a distinction between ways for the world to be, those in which the sentence is true and those in which it isn't. If a sentence φ entails another ψ, then all the ways for the world to be in which φ is true are also ways for the world to be in which ψ is true. The fact that some ways for the world to be are among others thus reflects the entailment relations between their corresponding sentences. In particular, sentences that entail each other make the same distinctions and therefore describe the same worlds. A theory's conception of logical space consists in a conception of the distinctions there are between ways for the world to be, and of how these relate. Note that we make room for the view that the ways for the world to be include *possible* as well as *impossible* ways. Accordingly, our view is compatible with a hyperintensional account of the meaning of sentences, vis-à-vis metaphysical equivalence, such as the one defended by Miller.

In a nutshell, our reason for rejecting ME is that ME is sensitive only to theories' commitments, thus neglecting the fact that theories may bring with them rather different conceptions of *logical space*. But two researchers may agree on the language in which they formulate their theories, and even on which sentences of that language are true, while still disagreeing with respect to their respective conceptions of logical space.

For instance, consider once more theories Cl and Int, introduced in Section 3.2. These theories are formulated in the same language, and are committed to the truth of the same sentences. Furthermore, under at least some ways of further specifying Cl and Int, the two theories are strongly hyperintensionally equivalent since, for this to be the case, it suffices that the strong hyperintension of A, as used by the proponents of Cl be the same as the strong hyperintension of A as used by the proponents of Int.

In such case ME implies that Cl and Int are metaphysically equivalent. The problem is that they aren't. For Cl and Int have radically different, non-isomorphic conceptions of logical space (in our terms, they are not *congruent*), as they differ with respect to their entailment structure. Indeed, Int's conception of logical space is much finer. By contrast to Cl's, Int's conception of logical space includes infinitely many distinctions. Relatedly, whereas the fact that p is true, given that $\neg\neg p$ is true, is surprising from the standpoint of proponents of Int, there is nothing surprising in this fact from the standpoint of proponents of Cl.

If the example involving theories Cl and Int looks too simple, owing to the meagre stock of propositional letters of their language, note that theories Cl* and Int* just like, respectively, Cl and Int except that they have countably many propositional letters will be *dissimilar*.[54] Still, whenever A, as used by the proponents of Cl*, has the same strong hyperintension as A, as used by the proponents of Cl*, theories Cl* and Int* will be hyperintensionally equivalent. Since Cl* and Int* are metaphysically inequivalent, owing to their "non-isomorphic" conceptions of logical space, ME delivers the wrong result.

So, there is good reason to prefer the *Synonymy account* to Miller's ME. In addition, and while we will here remain neutral on the right grain of content vis-à-vis metaphysical equivalence, we want to point out that the *Synonymy account* affords proponents of an intensional conception of meaning some resources for resisting Miller's reasons for advocating ME.

For the purposes of exposition, let us concede that 'always, everything always exists' and 'for one thing to constitute another is for the two things to

[54]For a proof of this result, see Kocurek [29].

be identical' are both metaphysically necessary. Still, proponents of *constitution as identity* will furthermore take the sentences 'the statue is constituted by the piece of clay' and 'the statue is identical to the piece of clay' to require the same of the world for their truth—to amount to the same distinction in logical space. By contrast, and to the extent that an *eternalist* theory even considers the distinctions in logical space that the sentences 'the statue is constituted by the piece of clay' and 'the statue is identical to the piece of clay' correspond to, it will not take them to amount to the same distinction in logical space.

Similarly, a proponent of *permanentism* will take the sentences 'in the past, there were dinosaurs' and 'there are things such that, in the past, they were dinosaurs', to require the same of the world for their truth—to amount to the same distinction in logical space. By contrast, and to the extent that a *constitution as identity*-theory even considers the distinctions in logical space that the sentences 'in the past, there were dinosaurs' and 'there are things such that, in the past, they were dinosaurs' correspond to, it will not take them to amount to the same distinction in logical space.

So, even if 'always, everything always exists' and 'for one thing to constitute another is for the two things to be identical' are both metaphysically necessary, it is reasonable to expect *constitution as identity* and *permanentism* to be theories with different conceptions of logical space. So, they are not metaphysically equivalent, regardless of whether 'always, everything always exists' and 'for one thing to constitute another is for the two things to be identical' have the same intension.

The general strategy suggested by the example to proponents of intensions as having the right grain for metaphysical equivalence is the following: to explain differences between theories that would seem to arise from the distinct requirements imposed on the world by their respective commitments as in fact having their source in the theories' distinct conceptions of logical space. No doubt, it is hard to resist the claim that the sentences 'always, everything always exists' and 'to be constituted by something is to be identical to it' have different meanings, even if they happen to be metaphysically true. But "intensionalists" need not disagree. What they will nonetheless claim is that, for the purposes of metaphysical equivalence, hyperintensions are not really required.

To sum up, Miller proposes a hyperintensional approach to metaphysical equivalence according to which two theories are metaphysically equivalent if and only if their respective commitments are jointly true at exactly the same possible and impossible worlds. Our main objection to Miller's hyperintensional account is that it wrongly implies that metaphysical equivalence is insensitive to theories' conceptions of logical space, and so predicts the metaphysical equivalence of clearly inequivalent theories. Moreover, and while remaining open on the issue, we noted that the *Synonymy account* affords some resources for rejecting Miller's contention that intensions are too coarse-grained for the purposes of metaphysical equivalence.

6. Conclusion

A better understanding of what it takes for metaphysical theories to be metaphysically equivalent, and of how to ascertain when this is the case, promises to lead to progress in theorising. For instance, it will make it possible to ascertain which disputes are merely verbal, and so to avoid double-counting.

In this paper we proposed an account of metaphysical equivalence, the *Synonymy account*, and defended its adequacy. We began by isolating, in Section 2, some desiderata that any correct account of metaphysical equivalence must satisfy. In Section 3 we offered an explication of metaphysical equivalence as *theory synonymy*. In a nutshell, according to this explication, theories are equivalent just in case they have the same conception of logical space and their respective commitments require the same of the world for their truth. We also proposed, in Section 3, some principles for determining whether a given translation is deeply correct, given how the notion of a deeply correct translation is central to the proposed explication of metaphysical equivalence as theory synonymy.

Then, in Section 4, we argued that the *Synonymy account* satisfies all the desiderata on accounts of metaphysical equivalence previously laid out. We also showed that the account affords a more nuanced and better understanding of the dialectic between noneists and Quineans. Furthermore, the discussion revealed that the account also promises to deliver a better understanding of other debates in metaphysics.

In Section 5 we considered two other accounts of metaphysical equivalence, namely McSweeney's epistemic account and Miller's hyperintensional account. Regarding the former, we argued that the requirement that two theories have a common definitional extension* in order to be justified in believing their metaphysical equivalence is not really a necessary condition. We also argued that the thesis that we should be able to conceive both theories from a unified perspective in order to be justified in believing their metaphysical equivalence is implausible. About the latter, we argued that by being insensitive to theories' underlying conceptions of logical space the hyperintensional account gives the wrong predictions with respect to the metaphysical equivalence of certain theories.

Funding Open access funding provided by FCT|FCCN (b-on). The first author acknowledges the financial support of the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the projects CEEC/IND/0467/2017, and the PHI-ASM Project - Philosophical Investigation of Applications of Science in Mathematics, with reference PeX.2022.06784.PTDC, https://doi.org/10.54499/2022.06784.PTDC. The second author acknowledges the financial support of the project PID2019-106078GB-I00 (2020-2024, MCI/AEI/FEDER,UE, 10.13039/501100011033) of the Spanish Ministry of Science and Innovation, the project IT1612-22 (2022-2025) of the Basque Government and the contract with UPV/EHU funded by the Postdoctoral Training Programme 2021–2022 (POS202110027) of the Department of Education of the Basque Government."

References

[1] Azzouni, J. A new characterisation of scientific theories, *Synthese* 191: 2993–3008, 2014.

[2] Baker, L. R., Why constitution is not identity, *Journal of Philosophy* 94(12): 599–621, 1997.

[3] Barrett, T. W., and H. Halvorson, Glymour and quine on theoretical equivalence, *Journal of Philosophical Logic* 45(5): 467–483, 2016.

[4] Benovsky, J., Tropes or universals: How (not) to make one's choice, *Metaphilosophy* 45(1): 69–86, 2014.

[5] Burge, T., On knowledge and convention, *Philosophical Review* 84(2): 249–255, 1975.

[6] Cameron, R. P., On characterizing the presentism/eternalism and actualism/possibilism debates, *Analytic Philosophy* 57(2): 110–140, 2016.

[7] Carnap, R., *The Logical Syntax of Language*, K. Paul, Trench, Trubner & Co, London, 1937.

[8] Carnap, R., Empiricism, semantics and ontology, *Revue Internationale de Philosophie* 4(11): 20–40, 1950.

[9] Carnap, R., The methodological character of theoretical concepts, in H. Feigl, and M. Scriven, (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis, Minnesota Studies in the Philosophy of Science* 1, Minnesota Studies in the Philosophy of Science, Minneapolis, 1956, pp. 38–76.

[10] Chalmers, D., Verbal disputes, *Philosophical Review* 120(4): 515–544, 2011.

[11] Chalmers, D., D. Manley, and R. Wasserman, *Metametaphysics*, Oxford University Press, Oxford, 2009.

[12] Davey, B. A., and H. A. Priestley, *Introduction to Lattices and Order*, 2nd edn., Cambridge University Press, Cambridge, 2002.

[13] Deasy, D., Characterising theories of time and modality, *Analytic Philosophy 60*(3): 283–305, 2019.

[14] Feigl, H., The 'orthodox' view of theories: remarks in defense as well as critique, in M. Radner, and S. Winokur, (eds.), *Analyses of Theories and Methods in Physics and Psychology, Minnesota Studies in the Philosophy of Science* vol. 4, University of Minnesota Press, Minneapolis, 1970, pp. 3–16.

[15] Giere, R., *Explaining Science: A Cognitive Approach*, University of Chicago Press, Chicago, 1988.

[16] Glymour, C., Theoretical equivalence and the semantic view of theories, *Philosophy of Science* 80(2): 286–297, 2013.

[17] Halvorson, H., What scientific theories could not be, *Philosophy of Science* 79: 183–206, 2012.

[18] Halvorson, H., The semantic view, if plausible, is syntactic, *Philosophy of Science* 80(3): 475–478, 2013.

[19] Hawtorne, J., A note on 'Languages and language', *Australasian Journal of Philosophy* 68(1): 116–118, 1990.

[20] Hempel, C. G., The theoretician's dilemma, in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Free Press, New York, 1965, pp. 173–226.

[21] Hirsch, E., Quantifier variance and realism, *Philosophical Issues* 12(1): 51–73, 2002.

[22] Hirsch, E., Physical-object ontology, verbal disputes, and common sense, *Philosophy and Phenomenological Research* 70(1): 67–97, 2005.

[23] Hirsch, E., Ontological arguments: interpretive charity and quantifier variance, in T. Sider, J. Hawthorne, and D. W. Zimmerman, (eds.), *Contemporary Debates in Metaphysics*, Basil Blackwell, Maiden, 2007, pp. 367–381.

[24] Hirsch, E., Language, ontology and structure, *Nous* 42(3): 509–528, 2008.

[25] Hirsch, E., Ontology and alternative languages, in D. Chalmers, D. Manley, and R. Wasserman, (eds.), *Metametaphysics*, Oxford University Press, Oxford, 2009, pp. 231–258.

[26] Janssen-Lauret, F., Committing to an individual: Ontological commitment, reference and epistemology, *Synthese* 193(2): 583–604, 2016.

[27] Janssen-Lauret, F., and F. MacBride, Lewis's global descriptivism and reference magnetism, *Australasian Journal of Philosophy* 98(1): 192–198, 2020.

[28] Johnston, M., Constitution is not identity, *Mind* 101(401): 89–106, 1992.

[29] Kocurek, A. W., On the concept of a notational variant, in A. Baltag, J. Seligman, and T. Yamada, (eds.), *Logic, Rationality, and Interaction (LORI 2017, Sapporo, Japan)*, 2017, pp. 284–298.

[30] Kölbel, M., Lewis, language, lust and lies, *Inquiry* 41(3): 301–315, 1998.

[31] Kuhn, S. T., *Many-Sorted Modal Logics*, Ph.D. Thesis, Uppsala University Filosofiska Studier Number 29, 1977.

[32] Lewis, D., *Convention: A Philosophical Study*, Harvard University Press, Cambridge, 1969.

[33] Lewis, D., Radical interpretation, *Synthese* 23: 331–344, 1974.

[34] Lewis, D., Languages and language, in *Philosophical Papers*, volume I, Oxford University Press, Oxford, 1983, pp. 163–188.

[35] Lewis, D., Noneism or Allism, *Mind* 99(393): 23–31, 1990.

[36] Lewis, D., Meaning without use: Reply to Hawthorne, *Australasian Journal of Philosophy* 70(1): 106–110, 1992.

[37] Linsky, B., and E. N. Zalta, In defense of the contingently nonconcrete, *Philosophical Studies* 84: 283–294, 1996.

[38] Lutz, S., On a straw man in the philosophy of science: A defence of the received view. *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 2(1): 77–120, 2012.

[39] Lycan, W., The trouble with possible worlds, in M. J. Loux, (ed.), *The Possible and the Actual*, Cornell University Press, Ithaca, 1979, pp. 274–316.

[40] McCall, S., and E. J. Lowe, 3d/4d equivalence, the twins paradox and absolute time, *Analysis* 63(2): 114–123, 2002.

[41] McSweeney, M. M., An epistemic account of metaphysical equivalence, *Philosophical Perspectives* 30(1): 270–293, 2016.

[42] Miller, K., The metaphysical equivalence of three and four dimensionalism, *Erkenntnis* 62(1): 91–117, 2005.

[43] Miller, K., A hyperintensional account of metaphysical equivalence, *Philosophical Quarterly* 67(269): 772–793, 2017.

[44] Noonan, H., Constitution is identity, *Mind* 102(405): 133–146, 1993.

[45] O'Leary-Hawthorne, J., Meaning and evidence: A reply to Lewis, *Australasian Journal of Philosophy* 71(2): 206–211, 1993.

[46] Pelletier, F. J., and A. Urquhart, Synonymous logics, *Journal of Philosophical Logic* 32: 259–285, 2003.

[47] Pérez Carballo, A., Structuring logical space, *Philosophy and Phenomenological Research* 92(2): 460–491, 2016.

[48] Priest, G., *Towards Non-Being: The Logic and Metaphysics of Intentionality*, Oxford University Press, Oxford, 2005.

[49] Priest, G., Against against non-being, *Review of Symbolic Logic* 4(1): 237–253, 2011.

[50] Putnam, H., *The Many Faces of Realism*, Open Court, 1987.

[51] Quine, W.V.O., On what there is, *Review of Metaphysics* 2(5): 21–38, 1948.

[52] Rayo, A., *The Construction of Logical Space*, Oxford: Oxford University Press, 2013.
[53] Routley, R., *Exploring Meinong's Jungle and Beyond: An Investigation of Noneism and the Theory of Items*, Australian National University, Canberra, 1980.
[54] Segerberg, K., *Classical Propositional Operators*, Clarendon Press, Oxford, 1982.
[55] Sidelle, A., The method of verbal dispute, *Philosophical Topics 35*(1&2): 83–113, 2007.
[56] Sider, T., *The Tools of Metaphysics and the Metaphysics of Science*, Oxford University Press, 2020.
[57] Sullivan, M., The minimal a-theory, *Philosophical Studies* 158(2): 149–174, 2012.
[58] Suppe, F., *The Semantic View of Theories and Scientific Realism*, University of Illinois Press, Urbana and Chicago, 1989.
[59] Suppes, P., *Representation and Invariance of Scientific Structures*, CSLI Publications, Stanford, CA, 2002.
[60] van Fraassen, B., *The Scientific Image*, Oxford University Press, Oxford 1980.
[61] van Inwagen, P., Meta-ontology, *Erkenntnis* 48: 233–250, 1998.
[62] Williamson, T., *Modal Logic as Metaphysics*, Oxford: Oxford University Press, 2013.
[63] Wittgenstein, L., *Tractatus Logico-Philosophicus*, Routledge, New York, 1990.
[64] Woodward, R., Towards being, *Philosophy and Phenomenological Research* LXXXVI(1): 183–193, 2013.
[65] Worrall, J., Underdetermination, realism and empirical equivalence, *Synthese* 180: 157–172, 2011.
[66] Wright, C., and B. Hale, *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*, Clarendon Press, Oxford, 2001.

B. Jacinto
Center for the Philosophy of Sciences and Faculty of Sciences
University of Lisbon
Lisbon
Portugal
bmjacinto@fc.ul.pt

J. Belastegui
ILCLI, University of the Basque Country (UPV/EHU); MCMP, Ludwig-Maximilians-Universität München (LMU)
Donostia
Spain
javier.belastegui@ehu.eus