Equivalence and Convention

Neil Dewar

November 9, 2022

The goal of this paper is to analyse the role of convention in interpreting physical theories—and, in particular, how the distinction between the conventional and the non-conventional interacts with judgments of equivalence. We will begin with a discussion of what, if anything, distinguishes those statements of a theory that might be dubbed "conventions". This will lead us to consider the conventions that are not themselves part of a theory's content, but are rather *applied* to the theory in interpreting it. Finally, we will consider the idea that what conventions to adopt might, itself, be regarded as a matter of convention.

1 Conventions within a theory

As is well-known, a major plank of the logical-empiricist program—associated especially with the work of Carnap—was concerned with analysing a theory into its "factual" and "conventional" (or synthetic and analytic) components. Perhaps the best-developed element of this was Carnap's proposal that the factual component of a theory T could be identified with its *Ramsey sentence* T^R, where the vocabulary to be Ramseyfied was the theoretical vocabulary; the conventional component could then be identified with what was subsequently called the *Carnap sentence* $(T^R \rightarrow T)$.

However, both this specific proposal and the general program of analysing a theory into factual and conventional parts have fallen from favour. In particular, the idea of a hard-and-fast distinction between the factual and the conventional is widely held to have been dealt a decisive blow by Quine (1951). Quine's argument may be summarised as based on two compelling observations. The first is that any attempt to explicate the analytic-synthetic distinction only leads us in a circle of tightly interconnected concepts (of meaning, synonymy, etc.). The second is that we cannot distinguish in any robust fashion between those parts of a theory that are immune to revision and those which are subject to empirical input. If an observation runs contrary to a theory's predictions, then the theory must be modified *somewhere*; but there is no matter of fact about where the modification must land. No part of a theory, says Quine, is not in principle available for modification in light of recalcitrant evidence.

And yet, and yet: there do, in fact, seem to be clear cases of conventions in physical theories. Consider, for example, the statement of a gauge condition (e.g. the Lorenz gauge condition $\partial_a A^a = 0$), or a commitment to a particular system of units (e.g. declaring that one will use units in which $c = 1$).[1] So we have a puzzle. On the one hand, Quine's analysis appears to provide very convincing *general* reasons for being sceptical that the conventions in a theory can be singled out. But on the other, we seem to have at least some clear examples where it is, in fact, possible to do this.

The purpose of this section, then, is to revisit the project of identifying what it is that makes something a convention. The aim is to see if we can find some properties that are characteristic of conventions. Later, we will consider how to reconcile this with Quine's critique.

As a starting-point, we will in fact follow Quine's lead: by taking *definitions* to be paradigmatic examples of conventions.[2] As in that paper, I will take a definition to be a statement that fixes the meaning of some newly-introduced term by assigning it the meaning already associated to some complex of existing terms. Examples might include "let the kinetic energy of a body be half its mass multiplied by the square of its speed", "the centre of mass of a system of bodies is obtained by dividing the mass-weighted sum of those bodies' positions by their combined mass", or "$\tan x = \sin x / \cos x$".

In order to discuss the role of definitions in a precise formal manner, we will start by thinking about the theory of definitions in logic. In that context, a definition of a new symbol R in terms of an existing language $\mathcal{L}$ is a sentence of the form

$$\forall x_1 \ldots \forall x_n (R x_1 \ldots x_n \leftrightarrow \phi(x_1, \ldots, x_n)) \tag{1}$$

where ϕ is an $\mathcal{L}$-formula. If we take a theory T, all of whose sentences are $\mathcal{L}$-sentences, and augment it with definitions of new symbols in terms of $\mathcal{L}$, then theory T^+ so obtained is said to be a *definitional extension* of T.

Definitional extensions have two characteristic features. A definitional extension T^+

[1] What of conventions in the theories of other sciences? I hope that what I say here might apply outside of physics, too; but for the sake of not exposing my ignorance, I will confine myself to explicit discussion only of examples from physics.

[2] Quine (1936).

of some $\mathcal{L}$-theory T will:

- *explicitly define*, in terms of $\mathcal{L}$, any symbol in the language of T^+ that is not in $\mathcal{L}$; and
- be a *conservative* extension of T.

The first condition means that for any new symbol, T^+ will entail a definition of that symbol (i.e. a sentence of the form (1)). The second condition means that for any $\mathcal{L}$-sentence ϕ, if $T \not\models \phi$ then $T^+ \not\models \phi$. Thus, the first condition is what guarantees that T^+ provides definitions of the new vocabulary; and the second condition is what guarantees that this is *all* T^+ does.

So, when a theory is augmented by definitions—which we take to be paradigmatic examples of conventions—the resulting theory explicitly defines the vocabulary introduced by those definitions, and is conservative with respect to the old theory. Now we ask the question: do these features extend to any convention? Unfortunately, at this point we butt up against the limitations of the use of a notion from logic. What we would like to do is take some uncontroversial examples of conventions—such as those listed above—and see if they abide by these two conditions. But those examples were drawn from physics

First, it does not seem plausible to suppose that the condition of explicit definability might be a hallmark of conventions. For one thing, not all conventions even introduce new vocabulary. Take as given Maxwell's theory stated in terms of electromagnetic potentials; then the statement of a gauge condition introduces no new vocabulary, but is certainly a convention. Even where a convention is introducing new vocabulary, however, we need not have definability. Suppose we start with Maxwell's theory formulated in terms of fields, then introduce the electromagnetic potential A_a by stipulating that it is to obey the condition

$$F_{ab} = \partial_{[a}A_{b]} \tag{2}$$

This condition is (I claim) most plausibly interpreted as a convention governing the use of this new symbol. But it does not provide a *definition* of A_a: it does not fix the meaning of A_a, it merely constrains it.[3] One quick way to see this is to observe that one can have two solutions of

The requirement of conservativeness seems more plausible, however. We would indeed expect that a mere convention should not, by itself, add content. Moreover, one finds

[3] Note that the statement $F_{ab} = \partial_{[a}A_{b]}$ does meet the formal conditions for being a definition—but only if it is taken as a definition of F_{ab} in terms of A_a, not the other way around.

that the condition of conservativeness is sometimes presupposed in philosophical analysis of conventions. For example, Gödel's critique of Carnap's conventionalism about mathematics uses as a starting-point the claim that for this to be so, the mathematics must be conservative over the base theory.[4] So perhaps this is a more plausible thing to require of conventions: that supplementing a theory by a convention ought to represent a conservative extension of that theory.

However, this cannot be right as it stands. Once again, consider the theory of electromagnetism formulated in terms of potentials. Imposing a gauge condition is surely a convention. Yet doing so will, in general, have non-trivial consequences. For example, if the Lorenz gauge condition is imposed, then Maxwell's equations may be re-expressed as

$$\partial_a \partial^a A^b = J^b \tag{3}$$

This is not a condition which can be derived in the ungauged version of the theory; hence, the gauged theory is not a conservative extension of the ungauged theory.

We are here butting up against the same issue that arose in discussing definability: in general, a convention need not introduce new vocabulary. Now the good news is that (unlike definability) the condition of conservativeness makes sense even if no new vocabulary is introduced. The bad news is that the condition trivialises. If a theory T and its extension T^+ are both in the same language $\mathcal{L}$, then T^+ is a conservative extension of T just in case the two theories are logically equivalent: in other words, if whatever new sentences are added to T to obtain T^+ are logically trivial.[5] So it would seem that if no new vocabulary is being introduced, then conservativeness is too demanding a condition to impose on putative conventions.

However, this objection misses an important fact about the ungauged theory: that it contains *surplus structure*. A consequence of this is that assessing what counts as "new content" is harder than it might seem. To demonstrate this, it is helpful to move from thinking about conservativity in syntactic terms to instead thinking about it in semantic terms. Specifically, we say that a theory T^+ is a *semantically conservative* extension of T if for every model M of T, there is some model M^+ of T^+ such that the reduct of M^+ to $\mathcal{L}$ is M.[6] In the context of first-order logic, semantic conservativity entails syntactic conservativity but not vice versa; in second-order logic, the two are equivalent.

Electromagnetism in the Lorenz gauge is not a semantically conservative extension

[4] Gödel (1995). See Warren (2020) and Marschall (2021) for more discussion of Gödel's argument.

[5] Unless T is inconsistent, of course.

[6] Recall that, given an $\mathcal{L}^+$-model M^+, the reduct of M^+ to $\mathcal{L} \subset \mathcal{L}$ is the $\mathcal{L}$-model with the same domain as $\mathcal{M}^+$, and which agrees with M^+ on the extension of all terms in $\mathcal{L}$.

of the ungauged theory. For just like syntactic conservativity, semantic conservativity trivialises in the case that $\mathcal{L} = \mathcal{L}^+$: in that case, if T^+ is a semantically conservative extension of T then it is logically equivalent to T. In other words, when we impose the Lorenz gauge condition, then we rule out certain models—those which do not comply with the condition. However, the characteristic feature of a gauge condition is that it must 'admit' at least one model from each gauge-equivalence class: that model (or models, in the case of an incomplete gauge condition) becomes the "representative" of that equivalence class. So although it does indeed rule out some models, any model that it rules out is gauge-equivalent to some model not ruled out. Let us say, then, that a theory T^+ is a *semantically conservative extension of T up to equivalence* if for every model M of T, there is some model M^+ of T such that the reduct of M^+ to $\mathcal{L}$ is equivalent to M. Electromagnetism in the Lorenz gauge is not a semantically conservative extension of ungauged electromagnetism "on the nose"; but it is a semantically conservative extension up to gauge-equivalence.

Generalising, I propose the following condition on something's being a convention: a statement $\phi \in T$ may be regarded as a convention, relative to some theory T' such that $T' \cup \{\phi\}$ is logically equivalent to T, just in case T is a semantically conservative extension of T' up to equivalence. A consequence of this is that whether ϕ is a convention is relative to (i) the theory T in which ϕ is embedded; (ii) the theory T' which we take to be obtained by removing ϕ; and (iii) the standard of equivalence to be adopted. Different standards of equivalence will yield different verdicts on what constitutes a convention. For example, if the standard of equivalence adopted is that of empirical equivalence, then one obtains the classic logical-empiricist position: any claim added to a theory which does not impact its predictive outputs is a mere convention. In the next section, we will devote more attention to the issue of what the criterion of equivalence should be.

The relativity of convention to the choice of comparison theories (i) and (ii) should also be noted. In particular, it strikes me as significant that whether $\phi \in T$ is a convention depends not only on T, but also on what one takes to be the result of "removing" ϕ from T. Of course, if one has a particular method for determining T' from T, one need not specify the two theories separately. One natural choice, for example, would be to simply let $T' = T \setminus \{\phi\}$. But this means that ϕ might be a convention when regarded as a sentence of T, but not when regarded as a sentence of some theory logically equivalent to T. Another way to do this would be to follow Carnap and identify the theory T' with the Ramsey sentence of T (where the vocabulary to be Ramseyfied is the theoretical vocabulary). Since a theory is always a conservative extension of its Ramsey sentence,[7]

[7](Button and Walsh, 2018, Proposition 3.5)

it will follow from this that any consequence of T that is not a consequence of T' will be a convention. In particular, the Carnap sentence $(T' \to T)$ may be identified as the "conventional content" of the theory, in contrast to the Ramsey sentence T' being the "conventional content"—as Carnap wished to argue.

However, I think we are better off not committing to one particular way of extracting a theory T' from the theory T. For, recognising that conventionality is relative to a choice of T' gives us a way to make sense of the observation with which we began: namely, that we seem to have clear examples of conventions in science, despite Quine's critique. The resolution, I claim, is that it does make sense to take a given theory (and a given standard of equivalence), and argue that *adding* a statement to that theory amounts to adding a mere convention. This is, indeed, precisely what we did in the case of gauge conditions in electromagnetism. What does not make sense, in general, is giving a general prescription for how to identify the conventional components of a given theory (as Carnap sought to do)—not unless one gives a general prescription for identifying T' from T. So, in a sense, identifying a convention is a one-way process. One can identify that adding such-and-such a claim to a theory would be merely to add a convention; one cannot say that such-and-such a claim that has already been added to a theory is a convention. This ties in nicely with a remark of Putnam's: "Quine has suggested that the distinction between truths by stipulation and truths by experiment is one which can be drawn only at the moving frontier of science. Conventionality is not "a lingering trait" of the statements introduced as truths by stipulation."[8]

2 Conventions about theories

Assessing whether a statement is a convention, then, depends (in part) on determining which models of the theory T' are equivalent to one another. The relevant sense of equivalence here is that of physical or theoretical equivalence: whether, that is, the models in question depict the same state of the world or not. Gauge-equivalence provides one example. More generally, symmetries are a good example of the kind of phenomenon at play here. For example, suppose one formulates a theory of N Newtonian particles using coordinates. One could then impose the condition that the centre of mass of the system is to be at rest. This is plausibly regarded as a convention, but only if we regard models related by a boost to be physically equivalent to one another. If they are not so equivalent, then "the centre of mass is at rest" is a hypothesis, not a convention. So it seems that to settle the question "is this theoretical statement a convention?" we need

[8](Putnam, 1962, p. 371)

to address the question of whether symmetry-related models are physically equivalent or not.

This is a debate with a sizeable (and still growing) literature. My own previous work on this has treated this debate as one with a determinate answer: namely, that symmetry-related models are indeed physically equivalent. Now, however, I am inclined to take a somewhat different attitude. It seems to me to be better to say that this, too, is an issue of what conventions to adopt. Unlike the conventions we have considered so far, however, these conventions will not be conventions *within* a theory; rather, they are conventions *about* the theory. One cannot have a convention within the theory that stipulates that a pair of models are equivalent to one another. For example, how might I indicate, within the theory of electromagnetism, that gauge-equivalent potentials are equivalent? A statement of the form $A_a = A_a + \partial_a \lambda$ accompanied by the assertion that λ can be any scalar field, for example, will make the theory inconsistent.

From this perspective, the question to ask is not "are symmetry-related models physically equivalent?", but rather "what are the pragmatic advantages or disadvantages of treating symmetry-related models as physically equivalent?" Treating them as equivalent has various pragmatic advantages. Since symmetry-related models are typically empirically equivalent, treating them as physically equivalent avoids concerns about underdetermination. If all such models are just different representations of the same physical situation, one may free choose whichever model is the most calculationally convenient. And finally, for local symmetries (i.e. gauge symmetries), this stance will at least make it possible to have a well-posed initial value problem.

Let us say, then, that the decision to treat certain models as equivalent to one another is a semantic convention. (We'll return below to the tenability of this position.) However, this is not the only kind of semantic convention that is important. There are also the conventions concerning how the theory relates to the world. Clearly, these kinds of conventions play some kind of important role—not least, in determining relationships of equivalence. Bas van Fraassen famously observed that the equation describing heat diffusion is formally identical to that describing gas diffusion, and hence that the difference between them must be a matter of their physical interpretation.[9] In a similar vein, Sklar notes that the statements "all lions have stripes" and "all tigers have stripes" are formally intertranslatable—but, again, would typically receive different interpretations.[10]

So, it seems, it is not enough to delineate a theory's internal standards of synonymy: one must also describe that theory's relationship to the world. This idea is pervasive

[9]van Fraassen (2014)
[10]Sklar (1982)

in recent philosophy-of-physics literature. For example, Maudlin (2018) argues that any theory must specify a physical ontology, not just the mathematical representation of that ontology. For another, De Haro and Butterfield (2018) make use of "interpretation maps", which "map from our theories and models, to 'meanings' and to 'the world'."[11] Indeed, one might even think that providing this kind of interpretation isn't merely needed *in addition to* the sort of "internal" interpretational work I outlined above; one could argue that once it has been specified what the mathematical structures represent, that will determine when two mathematical structures represent the *same* thing. Coffey (2014) articulates a position along these lines, as does Teitel (2021)—the latter of whom, incidentally, also describes interpretations as "mappings from representational vehicles to contents."[12]

Now, it's surely true in some sense that an interpretation consists of a mapping from representations to contents. And it is a tempting idealisation to suppose that we have a box of representations on the one hand, and an array of contents on the other, and the business of interpretation is a matter of correlating the one to the other—like a child with a sticker-book,[13] or a museum curator appending labels to the exhibits.[14] However, I'm a bit concerned about this picture, for two reasons.

First, it suggests that *any* mapping from vehicles to contents counts—at least in principle—as an interpretation. Indeed, Teitel explicitly argues that we need to take account of "trivial semantic conventionality": the "familiar platitude that any representational vehicle can in principle be used to represent the world as being just about any way whatsoever", i.e., that any association between vehicles and contents is an admissible interpretation. Of course, we're free to give the term "interpretation" a wide scope like this. But I think it is a mistake to abstract away so far from the kinds of interpretations *we* could give. The vast, vast majority of such interpretations are not, in any relevant sense, available to us. Only those interpretations—those mappings from representations to contents—which admit of specification by finite means are the sorts of interpretations which we could, in fact, articulate. One might say that this is why trivial semantic conventionality says merely that we have such interpretational lassitude *in principle*. But this doesn't seem right: it's not for lack of time, or resources, or ingenuity that our capacities to specify interpretations are so circumscribed. Make those as generous as you wish, and we will still only be able to articulate an infinitesimal fraction of the possible associations between words and contents. To consider our capacities "in

[11](De Haro and Butterfield, 2018, p. 322)
[12](Teitel, 2021, p. 4125)
[13]Price (2011)
[14]Quine (1969)

principle" is to suppose those capacities to be arbitrarily large; it is not to suppose them infinite.

Second, it seems to imply the wrong direction of explanation. On this picture, what makes something an interpretation is that it is such a mapping. So to interpret a theory is just to "give" such a mapping: to specify what propositions correspond to what sentences, or more generally, what contents correspond to what representational vehicles. This, I think, is misleading because it suggests that when we interpret a theory, we put it into contact with some realm of semantic objects of which we already have a grasp. Now, in some cases the practice of interpretation may involve something like this. In translating a theory from a foreign language, for example, we might indicate what terms in the other language correspond to what terms in the home language—and hence, assuming the home language is understood, what the semantic content of the foreign terms is. But I submit that the sense of interpretation we are interested in as philosophers of science simply is not this kind of thing. We do not start out with a grasp of those propositions the theory of General Relativity might be trying to say, and interpret that theory by putting its sentences into correspondence with those propositions. Rather, it is through the articulation and application of General Relativity itself that we come to be in a position to articulate the propositions that the sentences of General Relativity express.

What can we replace this picture with, then? Unfortunately, I do not have a good answer to this question. However, I do want to suggest that the essence of interpreting a theory lies in making sense of the use and application of that theory—in other words, in characterising its empirical content. Indeed, I am minded to say that so far as interpreting a theory on its own goes, specifying the empirical content is *all* there is to do. On the face of it, one might worry that this is inconsistent with realism; but I think that this worry is misplaced. We need to distinguish two things. On the one hand, there is the claim that all there is to the *content* of a theory is its empirical content: that a theory "says nothing more" than the set of its empirical consequences. That is indeed a strong (and likely unworkable) form of empiricism. However, one can deny this claim—and so endorse the basic realist commitment to theoretical content beyond empirical content—without thinking that there is anything more to the *activity* of interpretation than the specification of empirical content. If, following such a specification, somebody says "Well that's all well and good, but I don't only want to know how to interpret the empirical part of the theory; I want you to also tell me what the theoretical part is saying", we have no choice but to simply *repeat* the theoretical part itself.

However, there is an important missing piece here. Now suppose that somebody proffers a different theory, to which they have assigned the *same* empirical content as that which has been associated to our theory—so, in other words, the two theories are empirically equivalent. As already noted, we are not identifying the content of the theory with its empirical content, so we are not immediately forced to conclude that these two theories have the same content. However, we have not said anything which gives us the capacity to determine whether these two theories do, in fact, have the same content.

In other words, what is needed are appropriate *criteria of equivalence*. This is what I meant when I said that the specification of empirical content is all there is to interpretation when we are considering a theory on its own. When we consider a theory in relation to other theories, there is further work to be done, and that work consists in the specification of equivalence criteria. Again, this is a question that I have defended a particular answer to—by and large, the answer that we should try to adopt fairly liberal criteria of equivalence, such as inter-translatability or categorical equivalence.

However, this stance gives rise to the following dialectical problem. Many of those tempted by liberal criteria of equivalence are attracted by something like the following thought: there should not be questions which possess a definite answer, but where that answer could not be determined, even in principle, by empirical inquiry. (In a slogan: no disagreement without the possibility of resolution.) Yet the question "are these two theories equivalent?" does not appear to be one that could be settled by empirical inquiry. Certainly, if it were to be maintained that two intertranslatable theories were distinct, then it does not seem that we could point to empirical evidence that would refute their position.

Of course, the above problem is not without precedent. It recalls, indeed, the famous (or notorious) issue of whether the principle of verification is, itself, capable of verification. And just as reflection on that question moved Carnap to adopt the Principle of Tolerance in *Logical Syntax of Language*, so—I suggest—we should take a conventionalist attitude toward inter-theoretic equivalence (just as we earlier took a conventionalist attitude toward intra-theoretic equivalence). This then opens up the scope to view claims about theoretical equivalence as, again, recommendations rather than reports. The idea is that our scientific purposes are better served by regarding inter-translatable or categorically equivalent theories as equivalent, than by regarding such theories as distinct.

3 Conventions about conventions

However, all this raises difficulties. I am now advocating that the decision to adopt liberal standards of equivalence—i.e., to regard the differences between inter-translatable theories as merely notational or conventional—is itself a convention. As already discussed, one reason for doing this is by appeal to a Carnapian Principle of Tolerance. But one might worry that adopting the Principle of Tolerance might itself *already* commit one to a more liberal standard of equivalence. For after all, doesn't the Principle of Tolerance argue that when the choice between two theories may be regarded as a convention, it should be so regarded?

In other words, we seem to have a tension between two "levels" at which tolerance might be applied. At the level of comparing theories, the Principle of Tolerance seemingly instructs us to regard the choice between those theories as conventional—in other words, to regard the two theories as equivalent. But at the level of comparing criteria for theoretical equivalence, the Principle seemingly instructs us to regard the choice between liberal and illiberal criteria as a matter of convention: contradicting the earlier instruction to take the side of the liberal criteria! In other words: if we are tolerant about the choices between theories, that appears to commit us to *intolerance* about the choices between criteria of equivalence. This seems an uncomfortable position, as it seems to imply that tolerance can hold only within certain limits.

The key to dissolving this tension lies in looking more carefully at what happens if we do indeed adopt the Principle of Tolerance at both levels. As has just been discussed, at the level of comparing theoretical criteria, the Principle of Tolerance requires regarding the disagreement as merely conventional. This means that we cannot declare advocates of a stricter criterion of theoretical equivalence to be wrong. However, we are permitted to regard them as *unwise*. That is, it is consistent with the Principle of Tolerance to say that an illiberal criterion of equivalence is pragmatically inferior. From this perspective, the lower-level Principle of Tolerance—the one that is marshalled in support of liberal criteria of equivalence—amounts to a claim that being tolerant will bring pragmatic benefits; the higher-level Principle of Tolerance takes these benefits as reasons to adopt the lower-level Principle (as a convention).

This also gives us the resources to address an objection that might arise on the basis of trivial semantic conventionality. Recall that this is the thesis that any representation can be used to represent any content. A version of trivial semantic conventionality could be invoked to argue that any two representations *may* be regarded as equivalent. And if that's so, then it might seem that the Principle of Tolerance will insist that they *should* be regarded as equivalent. This will then collapse the content of all representations into one another—surely a *reductio* of this kind of view!

However, we can resist the pressure towards collapse if we semantically ascend: i.e., if we consider the pragmatic benefits of a tolerant framework rather than an intolerant one (using tolerance at the higher level to explain why it is pragmatic benefits that are the relevant ones to consider). I said above that the lower-level Principle of Tolerance then becomes the observation that being tolerant tends to bring pragmatic benefits. Such benefits might include: the fact that a more tolerant framework will permit more inferences (as we can "export" inferences between different theories); the fact, closely related, that we can switch between theoretical methods as the need may arise; and the time saved in not debating which theoretical method is true.

Nevertheless, such benefits are defeasible. If we identify theories too freely, then we may encounter pragmatic disadvantages that outweigh these considerations. I take it to be close to self-evident that there are pragmatic disadvantages to identifying theories which are not empirically equivalent. Even where two theories are empirically equivalent, there might be pragmatic arguments against identifying them. But if two representations are empirically equivalent *and formally intertranslatable*, then it is hard to see what the disadvantages to identifying them might be. (More subtly, it might be that some of the pragmatic advantages enumerated above will only apply when the theories are intertranslatable. For example, it is not clear to me how one might "export" a conclusion from one theory to another, without knowing how to express that claim in terms digestible by the second theory.)

References

Button, T. and Walsh, S. (2018). *Philosophy and Model Theory*. Oxford University Press, Oxford.

Coffey, K. (2014). Theoretical Equivalence as Interpretative Equivalence. *The British Journal for the Philosophy of Science*, 65(4):821–844. `http://bjps.oxfordjournals.org/content/65/4/821`.

De Haro, S. and Butterfield, J. (2018). A Schema for Duality, Illustrated by Bosonization. In Kouneiher, J., editor, *Foundations of Mathematics and Physics One Century After Hilbert: New Perspectives*, pages 305–376. Springer, Cham.

Gödel, K. (1995). *Kurt Gödel: Collected Works: Volume III: Unpublished Essays and Lectures*. Oxford University Press, Oxford, New York.

Marschall, B. (2021). *Carnap and the Ontology of Mathematics*. PhD thesis, University of Cambridge. https://www.repository.cam.ac.uk/handle/1810/324629.

Maudlin, T. (2018). Ontological Clarity via Canonical Presentation: Electromagnetism and the Aharonov–Bohm Effect. *Entropy*, 20(6):465. https://www.mdpi.com/1099-4300/20/6/465.

Price, H. (2011). *Introduction*. Oxford University Press, Oxford.

Putnam, H. (1962). The analytic and the synthetic. In Feigl, H. and Maxwell, G., editors, *Scientific Explanation, Space and Time*, number III in Minnesota Studies in the Philosophy of Science. University of Minnesota Press, Minneapolis.

Quine, W. V. O. (1936). Truth by Convention. In *Philosophical Essays for Alfred North Whitehead*, pages 90–124. London: Longmans, Green & Co.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, 60(1):20–43.

Quine, W. V. O. (1969). Ontological Relativity. In *Ontological Relativity and Other Essays*, pages 26–68. Columbia University Press, New York.

Sklar, L. (1982). Saving the Noumena. *Philosophical Topics*, 13(1):89–110. https://www.jstor.org/stable/43153911.

Teitel, T. (2021). What theoretical equivalence could not be. *Philosophical Studies*, 178(12):4119–4149. https://doi.org/10.1007/s11098-021-01639-8.

van Fraassen, B. C. (2014). One or two gentle remarks about Hans Halvorson's critique of the semantic view. *Philosophy of Science*, 81(2):276–283. http://www.jstor.org/stable/10.1086/675645.

Warren, J. (2020). *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. Oxford University Press.