On Mathematics as the Language of Physics

Galileo wrote that "the Book of Nature is written in the language of mathematics". It is indeed often said that mathematics is the language of physics. But in what sense is mathematics a *language*?

Mathematics *has* a language: it is that language to which belong symbols such as '5', '=', '$\int dx$', and so forth. But this language describes mathematical objects, from numbers to differentiable manifolds. It does not describe 'Nature'. What does describe nature are the abstract structures defined by sentences (that is, equations) of this mathematical language.

A better question, then, is: in what sense are the *objects* of mathematics a language? The symbol '5' represents the number five, but what does that number itself represent? Or, to stay closer to physics: a differential equation describes a class of solutions – but does any such solution represent the state of a physical system? If so, how?

In a broad sense, the answer is undoubtedly 'Yes'; otherwise mathematical physics would not succeed so well. We may still ask whether this kind of representation also requires non-mathematical language: does mathematics *by itself* represent nature? To illustrate what I mean, note that the sentence "the cat is on the mat" represents the cat as being on the mat; so does a picture of the cat drawn on the mat. Moreover, both the sentence and the picture represent by themselves: once I have uttered the sentence, or drawn the picture, there is no further need for me to clarify what they mean. Of course, words do not have the miraculous ability to latch onto objects. If it turned out that some distant constellation of stars spelled out the word 'cat', we would not take them to refer to any cat. Given the way I have learnt English and the way my use of it is embedded within a community of speakers, however, I simply need utter the sentence "the cat is on the mat" to say that the cat is on the mat.

(Sometimes such utterances require further clarification. Someone may ask what 'cat' means, or whose cat I have in mind. The same is the case for pictures: is that blob with two triangles sticking out of its top and four lines sticking out of its bottom meant to represent a *cat*? It is also possible that the surface meaning of a sentence is clear yet its deeper meaning is not. In a poem about love, "the cat is on the mat" may have another sense entirely. For those subtleties to arise, however, we must assume that the representation in question already has a literal meaning that requires clarification or interpretation. This does not detract from my claim that sentences and pictures represent directly.)

There are then two views about mathematics I wish to contrast. The first is that mathematical models, just like sentences and pictures, represent immediately. This does not mean that representation is context-independent. It does mean that mathematical models represent by themselves. The second view is that mathematical models do not represent immediately. The role of mathematics in physics is to define certain models that *eventually* come to represent the world, but in order for those models to represent the world one has to endow them with content by means of another, typically linguistic, representation. To invert the terminology introduced by Wallace (2022), I will call the first view *maths-first-maths-last* (*maths-last* for short) and the second *maths-first-language-last* (*language-last* for short). What I call the 'maths-last approach' is of course intended to resemble Wallace's 'math-first approach'. But *both* views start from mathematically formulated theories. The question that divides them is whether one must subsequently appeal to language or whether the mathematics can stand on its own.

This terminology of course suggests the possibility of *language-first-maths-last* and *language-first-language-last* approaches. I will briefly discuss these in the next section.

I thus define the maths-last approach as follows:

> **Maths-Last Approach**: mathematical models can represent the world by themselves, without mediation by another, non-mathematical representation, in the same way that linguistic or pictorial representations can.

The language-last approach is the denial of this claim.

Which of these approaches is correct is of interest in itself. But the stakes are higher. For example, Wallace (2022) claims that the *content* of models may depend on whether mathematical representation is immediate or not. In particular, he believes that on a maths-last approach (in particular: a structural realist one), (i) questions of ontology and ideology are 'not applicable' (367), and (ii) the problem of underdetermination of theory by data is 'bypassed' (Wallace 2022).

Similarly, Weatherall's (2018) putative resolution of the hole argument implicitly on a maths-last approach in its insistence that the interpretation of a theory is guided by its mathematics – and mathematics alone. This is made explicit by Bradley & Weatherall (2022), who hold that the interpretation of a theory should take place from 'within' its mathematical formalism: "the metatheory is representationally irrelevant" (1231). Weatherall's approach to the hole argument indeed promises to avoid (i) questions of spacetime ontology and ideology and (ii) the underdetermination of a theory's diffeomorphism-related models by data.

My aim here is not to question whether maths-last representation is possible, but rather whether the transcendence of metaphysics and the avoidance of underdetermination are part of the spoils of the maths-last approach. I will defend three theses: (A) the language-last approach is implicit in most if not all contemporary theories of scientific representation, those that explicitly adopt the so-called 'semantic view' of theories included; (B) the maths-last approach can only transcend metaphysics and avoid underdetermination if mathematical and linguistic representations are semantically incommensurable; (C) there is no available metasemantic theory for mathematical representation that lends any plausibility to such incommensurability. In the conclusion, I will sketch an alternative to the maths-last approach that promises to take both mathematics *and* language seriously.

Semantics and Mathematics

I start with thesis (A): that the language-last approach is the orthodoxy in the contemporary literature on scientific representation – even on the semantic view of theories. Recall that while on the syntactic view a theory is presented as a collection of (interpreted) sentences, on the semantic view it is presented as a collection of mathematical models. The syntactic view is thus a *language-first-and-last* approach: it says that we should start out and end with linguistically formulated theories. The maths-first-language-last approach, on the other hand, acknowledges that a theory is presented as a collection of models, but adds that a collection of sentences must accompany those models for them to successfully represent. From now on, I will not consider the language-first-and-last (i.e. syntactic) approach and use the term 'language-last approach' exclusively to refer to the maths-first-language-last approach. (There remains the possibility of a 'language-first-maths-last' approach: but I cannot think of any views that would count as such.)

While the syntactic view just is the language-first-and-last approach, the semantic view is *not* identical to the maths-first-and-last approach. The semantic view is compatible with the view that a theory is presented as a collection of models that require further interpretation. One may complain that this construal of the semantic view is unfairly weak: the semantic view should say that a theory is presented as *only* a collection of models *without* any accompaniment. But that is certainly not the semantic view as usually understood. To show this, I will consider three standard conceptions of scientific representation on the semantic view: the similarity conception, the structuralist conception, and the inferentialist conception (Frigg & Nguyen 2020). Each of these supplements models with linguistic explications. Chakravartty (2001, 330-31) summarises the situation well: “a model can tell us about the nature of reality only if we are willing to assert that some aspect(s) of the model has a counterpart in reality. That is, if one wishes to be a realist, some sort of explicit statement asserting a correspondence between a description of some aspect of a model and the world is inescapable. This requires the deployment of linguistic formulations, and interpreting these formulations in such a way as to understand what models are telling us about the world is the unavoidable cost of realism.”

On the *similarity conception*, a model represents a system in virtue of some similarity between them. This conception is most obvious for mechanical models, such as a scale model of an aeroplane. But its advocates believe that the same applies to mathematical models. The well-known problem with this account is that anything is similar to anything else in *some* respect. The traditional solution is to let model-users propose *theoretical hypotheses* that specify the particular respect in which the model is similar to the system: “It is not the model that is doing the representing; it is the scientist using the model who is doing the representing. One way scientists do this is by picking out some specific features of the model that are then claimed to be similar to features of the designated real system to some (perhaps fairly loosely indicated) degree of fit” (Giere 2004, 747-8). These theoretical hypotheses “are linguistic entities” (Giere 2010, 273), so the similarity conception presumes a language-last approach.

On the *structuralist conception*, a model represents a system in virtue of some morphism (isomorphism, partial morphism, etc.) between them. Like similarity, however, (iso)morphism is too flexible: almost any system instantiates *some* structure that is isomorphic to a chosen model. This is related to Putnam’s infamous paradox. There are two responses possible. The first is to use an already-understood language to specify the *intended* isomorphism (van Fraassen, 2008). The second is to describe the system itself in a certain way in order to pick out the intended structure. Both clearly involve appeal to further linguistic resources. French & Saatsi (2006, 557), advocates of the structuralist conception, admit that “it is indeed the linguistic complement of the semantic framework that allows us to sidestep the problem of unintended models.” The structuralist conception also presumes a language-last approach.

On the *inferentialist* conception, a model represents a system in virtue of the inferences about the system the model allows its user to draw. But in order to draw such inferences, some link between model and world is necessary. Thus, on Contessa's (2007) version of the inferentialist conception an 'analytic interpretation' is introduced by which a user takes a model's elements to denote objects, properties and relations. The inferentialist-expressivist account of Khalifa et al. (2022) also specifies that "model elements are given interpretations using the resources of language." The inferentialist conception is language-last, too.

This survey is not exhaustive, but other options such as Frigg & Nguyen's (2020) 'DEKI-account' similarly supplement models with linguistic descriptions. Wallace is thus mistaken when he says that "the move from a 'syntactic' to a 'semantic' conception of theories is itself more or less sufficient to turn standard realism into structural realism" (2022, 347). Although in a later paper (Wallace 2024) he acknowledges that the appeal to language "plays a central, underacknowledged role in much of the literature on scientific representation" (16), he nevertheless presents 'mathematics-first' as simply another name for the semantic view (2). This conflates distinct views, however, as a survey of the literature shows. The semantic view of theories is maths-first, but not necessarily maths-last.

Maths-Last Structural Realism

Before I defend theses (B) and (C), I need to explain why the maths-last approach is supposed to avoid questions of ontology and ideology and the problem of underdetermination. I have already drawn a contrast between language-last and maths-last approaches. This distinction concerns *how* models represent, but not *what* they represent. Wallace believes that the maths-last approach entails a form of structural realism. In this section, I define this maths-last structuralism. In the next two sections, I question the claim that the maths-last approach entails such structuralism.

The maths-last approach is neutral about our epistemic attitude towards the contents of mathematical models. It is possible to supplement the maths-last approach with:

> **Maths-Last Scientific Realism**: the mathematical models of our theories succeed, at least approximately, in representing the systems which they are used to model, including features of those systems that are unobservable. (Wallace 2022, 361)

Notice that, while standard formulations of scientific realism appeal to *truth*, maths-last scientific realism appeals to *successful representation*. The reason is that models are not truth-apt. In this sense, the relation between model and world is "more akin to the relation between map and territory than that between word and object" (Wallace 2022, 350). In an informal sense we may say that a picture is true to its subject, and likewise the maths-last scientific realist can informally speak of a theory's being true.

Yet maths-last realism is silent about the content of mathematical models. For all I have said so far, it is possible that the language-last realist and the maths-last realist merely use a different medium to represent the world as being exactly the same way. But this is denied by maths-last *structuralism*:

> **Maths-Last Structuralism**: the mathematical models of our theories represent only the structure of the systems which they are used to model.

The conjunction of maths-last realism and maths-last structuralism yields a form of structural realism. I will call this combination of views *maths-last structural realism*.

Notice, firstly, that maths-last realism does not entail maths-last structuralism nor vice versa; and secondly, that whereas structuralism is usually presented as an epistemic claim (our theories are correct only in what they say about structure) and/or a metaphysical claim (structure is all there fundamentally is), maths-last structuralism is a semantic claim. Wallace (2022) uses the term 'maths-[last] ontic structural realism' for the conjunction of maths-last structuralism and the metaphysical claim that the structure represented by mathematical models is all there fundamentally is; and 'maths-last epistemic structural realism' for the conjunction of maths-last structuralism and the *denial* of the latter claim. The truth of the metaphysical claim does not matter in what follows.

It is famously difficult to articulate the distinction between 'structure' and 'nature'. I will approach the issue indirectly: we will get a grip on the structure that models are supposed to represent by reflection on structuralism's supposed virtues. These are the transcendence of metaphysics and the avoidance of underdetermination. With respect to the former, the idea is that languages invariably describe the world in terms of objects, their properties and the relations between them. This is certainly true for first-order logic, and plausible (but incomplete) for much of English. Yet models, it is claimed, are not concerned with objects, properties and relations. (Although this claim seems somewhat questionable once we explicitly analyse models as set-theoretic structures.) Whatever kind of structure mathematics represents, then, it is silent on such issues. This is beneficial because it enables us to sidestep "our ordinary talk of objects, properties and relations" (Wallace 2022, 366) as well as "overly-fine metaphysical distinctions" (361) that have troubled scientifically minded philosophers. (Of course, many philosophers are perfectly happy with very fine distinctions!)

With respect to the second, maths-last structuralism answers the problem of underdetermination in the usual way: putative examples of inequivalent theories are actually equivalent as far as structure is concerned. And here of course the second benefit links up to the first. For the difference between these putative examples of inequivalent theories are supposed to differ, the maths-last structuralist says, is exactly in what they say about objects, properties, and relations. If we were to strip them of any commitments with respect to ontology and ideology, the theories that remain would have the very same content, namely their shared structure.

Let me digress a little on the issue of theoretical equivalence. It is helpful to distinguish between three types of equivalence: empirical, theoretical and mathematical. Theories are *empirically equivalent* whenever they make the same empirical predictions; and *theoretically equivalent* whenever they represent the world as being exactly the same way. The notion of *mathematical equivalence* is murkier: proposed definitions in terms of isomorphism, categorical equivalence or intertranslatability are all up for debate. The maths-last structuralist claims both that empirical equivalence entails mathematical equivalence, and that mathematical equivalence entails theoretical equivalence (Wallace 2022, 362–3). I will not say much about the first implication because it does not concern maths-last structuralism specifically, although it is controversial (Bradley, 2021; Norton, 2003). The second implication is supposed to follow specifically from maths-last structuralism: the implicit assumption is that since mathematically equivalent theories *have* the same structure, they therefore *represent* the same structure.

Granted the first implication and a particular definition of mathematical equivalence, however, theoretical equivalence does not yet seem to follow. For example, suppose that Newton-Cartan theory and Maxwell Gravitation are mathematically equivalent. It remains the case that they are formally distinct: the models of the former contain *curved* covariant derivates, those of the latter equivalence classes of *flat* covariant derivatives. It is unclear why such differences, which by stipulation transcend whatever notion of mathematical equivalence is in play, should not have representational significance even for a structuralist. For example, taken at face value it seems that Newton-Cartan spacetime is curved whereas Maxwell spacetime is not (North, 2021) – and if this is a physical difference, then surely it is one in structure!

For another example, consider the AdS/CFT correspondence. Despite their presumed mathematical equivalence, these theories differ over the number of spacetime dimensions. One does not need a subject-predicate account to spell out this difference: dimensionality is well-defined mathematically. It is therefore unclear why maths-last structuralism would entail that "The question 'what is the dimension of spacetime at the most fundamental level' has no answer" (Wallace 2022, 368).

Let me put these issues with theoretical equivalence aside, however, and focus on the broader picture. We can take it as characteristic of structuralism that it transcends object-property-relation metaphysics and, in this way, avoids underdetermination. Now, Wallace (2022, 363) believes that

> we should see [maths-last] scientific realism as *already* a form of structural realism, without any need for further selective scepticism: the move from [language-last] to [maths-last] formulations of theories […] is itself sufficient to make this form of realism 'structural'.

It is *this* inference, from maths-last realism to maths-last structuralism, that I wish to dispute in the remainder of this paper. So, I concur with (or at least concede to) Wallace (2024) that mathematical models can represent reality without the existence of an antecedently understood linguistic representation of reality. The claim I question is that this mathematical representation is necessarily (that is, in virtue of its form) structural.

Maths-Last Semantics

So far, I have accepted the claim that mathematical models can represent the world by themselves. The question remains: what *does* a model of a theory say the world is like? If Wallace is correct that maths-last representation is already a form of structuralism, then models represent the world in a way that does not make reference to objects, properties and relations. I will argue that this is only the case if mathematical and linguistic representations are semantically incommensurable. This is thesis (B).

What we would like, of course, is a *theory of meaning* for mathematical models. Such a theory has two parts (Speaks, 2025). Firstly: a *semantic theory*, that is, a theory that allows one to determine the meaning of the expressions of a representational system. Secondly: a *metasemantic theory*, that is, a theory of what makes it the case that those expressions mean what they mean. In this section, I discuss the former. The next section will ask whether metasemantic theories can support the kind of semantic incommensurability that maths-last structuralism requires.

The classical form of a semantic theory is akin to what one finds in the semantics chapter of a logic textbook: it assigns an extension to terms and lays out compositional rules for how to interpret sentences that contain those terms. In the first instance, one would expect a maths-last semantic theory to take the same approach. This might consist of a map, specified in a language, from the elements of a model to objects in the world. The accounts of scientific representation surveyed above typically take such an approach. But this would sneak language in through the back door. Although it would not entail that language is required for models simply to represent – in the same way that a speaker of English need not have any knowledge of English's semantics – it would ultimately explicate mathematical representation in linguistic terms. This directly leads to a non-structuralist metaphysics of objects and properties.

Wallace (2024) therefore opts for an alternative, broadly Davidsonian semantic theory, on which to know what an expression means is to know its truth conditions (Davidson 1967). Thus, a semantic theory for a language L should entail what is called a *T-sentence* for every sentence S of L:

'S' is true in L iff P.

The idea then is that someone understands L whenever they know all of L's T-sentences. There are well-known problems with this type of theory, such as whether the material biconditional is sufficient, which I set aside here.

The reason Davidson's theory may seem helpful to the maths-last approach is that it can explicate one type of representation in terms of another representation *of the same type*. Consider first two different natural languages, say, English and Dutch. For a native English speaker to understand Dutch, she would need to know T-sentences such as "'Sneeuw is wit' is true iff snow is white". Those sentences are clearly informative to someone who has no prior knowledge of Dutch. But now consider what it takes for a native speaker of English to understand *English*: she would need to know T-sentences such as "'Snow is white' is true iff snow is white". Sentences of this form seem trivial: the distance between the *explanans* and the *explanandum* is small. This is not a problem for Davidson's theory, however. It is rather a consequence of the fact that it is impossible to escape our representational system to clarify it from the outside.

In order to apply this theory to the maths-last approach, we can think of mathematical and linguistic representations as akin to English and Dutch. Thus, one way to provide a semantic theory for mathematical models is to explicate what each model means in terms of a linguistic meta-representation. For a native ‘speaker’ of mathematics, however, this is unnecessary. The analogues of T-sentences on the maths-last approach simply express the semantics of one model in terms of another, possibly identical model. It is true that such expressions may seem akin to the uninformative statement that a model M represents whatever it represents. But the same is true for linguistic representation, so the maths-last approach is at no disadvantage here.

There are questions about how to extend Davidson’s semantic theory, with its reliance on linguistic notions such as truth and quotation, to non-linguistic representation. Wallace’s (2024) TT (for ‘theories and theorising’) account offers one possibility. The analogue of a T-sentence on his account would look like:

> a mathematical structure X accurately represents a class of systems iff [assertion of Y],

where Y is of course an appropriately chosen already understood model of the same class of systems. Just as the object language and metalanguage of a T-sentence may coincide, so X and Y may be drawn from the same theories. It is left open what exactly it takes to ‘assert’ a model.

Does a Davidsonian semantics for maths-last representation entail structuralism? There is one situation in which it clearly does not: if it is possible to faithfully translate a mathematically formulated theory into a linguistically formulated one. For in that case, mathematical models are committed to objects, properties and relations insofar as subject-predicate sentences are; and hence cases of underdetermination would reappear. The semantic incommensurability of mathematics and language is thus a necessary (but not sufficient) condition for maths-last structuralism. It is not enough to show that maths-last representation does not *need* a linguistic interpretation (as Wallace (2024) does): the maths-last structuralist must also show that it *cannot* have one.

Wallace endorses incommensurability: “physical theories are not given in natural language and cannot losslessly be translated or interpreted that way” (2024, 27–8). I will first show that Wallace’s remarks on this issue are insufficient to establish incommensurability and then proffer one reason in favour of commensurability.

Firstly, Wallace (2024) likens learning a new language to learning a new, superior theory. Based on that analogy, it would seem that translation back into the old language/theory is indeed impossible. But the analogy fails: theories make claims about the world; languages merely provide the framework within to make such claims. It is conceivable that distinct theories expressed in the same framework are commensurable, yet that formulations of the same theory in different frameworks are not. For example, it is not possible to express relativistic spacetime in terms of a classical conception of space and time, but both are expressible within the same framework of differential geometry.

Secondly, Wallace (2022) points out that the same theory may admit of more than one linguistic explication. One could take this to mean that such explications include extraneous content. But that would be to mistake a case of the indeterminacy of translation for a case of incommensurability. It is well-known that the map from one language to another is underdetermined by the evidence, as Quine's gavagai example shows (Quine, 1960). The lesson is not that translation is impossible, however, but that reference is indeterminate. If there is more than one explication of the same (by maths-last standards of equivalence) theory, we should at most conclude that the content of that formalism is indeterminate between whatever is expressed by those explications. For instance, AdS/CFT leaves the dimensionality of spacetime indeterminate.

Thirdly, and most seriously, Wallace (2024) says that native speakers mathematics themselves believe that no English set of sentences can fully express the content of a mathematical model. Physicists might try to answer subject-specific questions in words, but "push sufficiently hard on that answer and sooner or later they will write an equation on the blackboard and explain that *really* what they mean is *this*" (2024, 2). And he likens this to the way in which a translation of Plato supposedly cannot communicate his thoughts as well as the original Greek. But there are many reasons why physicists might prefer mathematics short of incommensurability, such as conciseness, precision or familiarity. Furthermore, we may in the process of translation extend and augment our language; and even then, translation is often approximate and holistic. Perhaps some particular Greek word has no English equivalent; that does not mean a more-or-less faithful translation of Plato is impossible, certainly not if we are allowed to introduce loanwords into our vocabulary. Likewise, admittedly there is no snappy phrase to express asymptotic freedom, but that does not show that one cannot translate the whole of quantum field theory into an expanded version of English. There is more to be said on this issue, however; I hope that the view presented in the final section of this paper does justice to the intuition that mathematics is necessary to physics.

I therefore see no reason to accept incommensurability. Given that it is a radical thesis, the burden of proof lies with its advocate. Nevertheless, let me sketch one reason against incommensurability. The reason is essentially Davidson's (1973). Consider once more, from our vantage point as English speakers, a foreign language such as Dutch. Insofar as the T-sentences provide a semantic theory for Dutch, I can only understand that language *as* a language – as a system of symbols that makes truth-apt claims about the world – if I can translate it into *my* language. Otherwise, Dutch sentences would for me fail to have truth conditions at all. In the words of Malpas (2024): "inability to translate counts as evidence, not of the existence of an untranslatable language, but of the absence of a language of any sort." The same applies to mathematics: I can only understand mathematical models as accurate-or-not representations of the world if they are translatable into my language. The condition of intertranslatability is symmetric. The maths-last realist could insist that mathematics *is* their language: but then English and Dutch, if they are not translatable into mathematics, will fail to count as languages. So, if the maths-last realist wants to espouse incommensurability, they will first have to formulate that thesis as a mathematical model!

Criticisms of Davidson's theory have focused on the connection between translation and truth (Aune, 1987; Hernández-Iglesias, 1994). But Wallace embraces this connection: "What, from [an external observer's] perspective, makes it *true* that the student understands general relativity? I suggest that the answer needs to be a mathematized version of radical interpretation" (2024, 13). The most obvious response is not available to him.

To sum up: if a Davidsonian semantic theory for mathematical models succeeds, maths-last representation is possible. But in order to show that this is a maths-last *structuralism*, a thesis of semantic incommensurability must hold. If, to the contrary, mathematical and linguistic representations are intertranslatable, the former are committed to a metaphysics of objects, properties and relations just as much as the latter, and maths-last realism will have to face the problem of underdetermination just as much as language-last realism.

Maths-Last Metasemantics

While a semantic theory is interested in what an expression means, a metasemantic theory is interested in what makes it the case that an expression means what it means. In virtue of what does a mathematical model represent the world as being a certain way?

We might try to look for evidence of incommensurability not at the level of semantics, but at that of metasemantics. The shortest path is through theoretical equivalence. If mathematically equivalent theories have the same content while that of their linguistic formulations differs, then it is plausible that the latter are an unfaithful translation of the former. For the sake of argument, I assume that explications of mathematically equivalent theories *are* theoretically inequivalent. The question then is whether the maths-last metasemantics implies that such theories are theoretically equivalent. I will argue – here is thesis (C) – that metasemantics provides no reason to think this.

I am not aware of any explicit work on a metasemantic theory for mathematical models, but some recent accounts of scientific representation are easily modifiable for this purpose. I will discuss three such proposals: mentalism, pragmatism, and internalism. I cannot rule out the possibility that another metasemantic theory is better suited to maths-last structuralism, but currently such a theory is not available.

Mentalism

The first type of foundational theory I discuss is *mentalism*. On this view, every kind of representation is ultimately reducible to mental representation. Although many have proposed theories of this form, I will focus on the 'General Griceanism' coined by Callender & Cohen (2006). (Note that Lewis' (1956) theory of language, which Wallace (2024) favours, is also mentalistic, since it analyses meaning in terms of belief; but Wallace disavows this particular aspect of Lewis' account.)

Callender and Cohen set out to show that scientific representation – which includes representation by means of mathematical models – is in principle no different from linguistic representation. Callender and Cohen introduce a doctrine they call 'General Griceanism', which consists of two theses: (1) all representations derive their representational force from a small set of fundamental representations of the same kind, namely mental representations; (2) facts about mental representation are eventually reducible to facts that are not about representation at all. What they call 'Specific Griceanism' is a particular way to fill out General Griceanism based on speaker-intentions, but this specific theory will not concern us here.

In order for General Griceanism to provide a maths-last foundational theory, it should entail that mental representation does not involve objects, properties and relations, and that mathematically equivalent representations reduce to mentally equivalent representations. The latter condition is necessary to ensure that mathematically equivalent representations have the same representational content as per maths-last structuralism.

It is highly unlikely that these conditions are satisfied. The reason is that, if General Griceanism is correct, linguistic representation is *also* reducible to mental representation. Since linguistic representation does involve objects, properties and relations, it is impossible that mental representation does not. Moreover, since mathematically equivalent representations may correspond to inequivalent linguistic representations, it is just as plausible that they correspond to inequivalent mental representations.

It is possible that although the mental representations that underlie linguistic representations involve objects, properties and relations, the mental representations that underlie mathematical representations don't. This would involve an implausible distinction between two kinds of mental representation. It is unlikely that we have evolved specific representational capacities for representations of, say, fibre bundles or tensor fields. Moreover, such a distinction would take away the parsimony and unificatory force of mentalism. Indeed, it is questionable whether mentalism can still function as a foundational theory for maths-last representation at all if the latter is reduced to a unique and equally mysterious form of mental representation. I therefore concur with Wallace (2024) that mentalism does not support maths-last structuralism.

Pragmatism

The second type of metasemantic theory bases the meaning of models in their *use*. I will call such views 'pragmatist'. Wallace seems to prefer a pragmatist approach when he takes scientists to use a model M to represent a system whenever this "best makes sense of their behavior" (2024, 14). So far, however, the proposal is imprecise. How does use contribute to meaning? I will focus on one recent proposal, namely Menon's (2024) 'inferential scientific vindicationism', inspired by Robert Brandom's pragmatist-inferentialism.

Menon's proposal does not concern models, but declarative sentences such as 'electrons have negative charge'. I will first summarise his account and then consider how to apply it to the maths-last approach. The account starts with the idea that our communicative practises are constrained by certain *norms of inference*. This includes (i) rules for which assertions I am entitled to base on other assertions, and (ii) 'entry' and 'exit' rules for which assertion I am entitled to on the basis of some course of action or which course of action I am entitled to on the basis of some assertion. Consider the assertion that my desk is square. The first type of norm entitles me to infer that my desk is not circular. The second type of norm entitles me to try to fit it into the corner of my room. The meaning of an expression is then equated to the contribution that that expression makes to inferences that conform to the community's norms. The ultimate arbiter of those norms is the world itself: a norm becomes entrenched when its users are appropriately rewarded. Menon's specific aim is to show how an inferentialist account is consistent with scientific realism: what makes a certain discourse realist is that it is embedded in a rich network of inferences, entry rules and exit rules and thereby latches onto the world.

It is fairly easy to transpose this account to mathematical models. The norms of inference include, for example, mathematical derivations and idealisation assumptions; the entry and exit rules correspond to the construction and application of models; and success is constituted by scientific prediction and explanation. The content of a model is based on the inferences and applications in which it features within the relevant community.

Does this metasemantic theory mean that models represent the world in purely structural terms? Not necessarily: Menon shows that inferentialism does not prohibit a robust notions of reference and predication. Potentially the physicist's norms of inference are such that no reference is actually achieved. It would require an empirical study of the communicative behaviour of physicists to find out, but a shortcut proceeds via the notion of equivalence. If physicists treat mathematically equivalent models as inferentially equivalent (once a suitable definition of mathematical equivalence is settled), then that is evidence that those inferences are only concerned with structure. This is consistent with the TT account's notion of theoretical equivalence: T_1 and T_2 are equivalent whenever it makes best sense of the behaviour of their users to take the theories to accurately represent the same phenomena (Wallace 2024, 22).

There are two questions we should ask at this point: which inferences, and which community? With respect to the first, physicists often treat mathematically equivalent models rather differently. Some models are simply easier to use, or more familiar, or more perspicuous. One may discount these differences as merely practical, but such a response is implausible for a metasemantic theory on which practice constitutes content. There are in any case more substantive differences. Sometimes one mathematical model is more likely to survive theory change than another, for example if it is more amenable to quantisation. Other times one model is more explanatory than another: consider Galileo's heliocentric model, which was (at least at the time) empirically inferior to its competitors but more explanatory due to its simplicity.

There is an easy response available to the maths-last structuralist: only those inferences that directly contribute to the theory's empirical accuracy are relevant to the aims of science. This immediately excludes both factors of convenience and theoretical virtues. But it leaves us little more than a thinly veiled empiricism. This is, perhaps, enough to achieve a maths-last structuralism, but certainly not a maths-last structural *realism*.

With respect to the second question, we may wonder how much to defer to physicists. Grant that physicists treat mathematically equivalent models as entirely on a par; philosophers certainly don't! On one prominent view on the role of philosophy of physics, it is a 'continuation of science by other means' (Chang 1999). Both the physicist and the philosopher aim to find out what the world is like, although they focus on different questions and use different methods. If we include philosophers in our community, mathematically equivalent models can accrue different content in virtue of their distinct use by philosophers. Thus, even if the first question is adequately answered, maths-last structuralism requires that we fully defer to physicists in the task of interpretation. This is implausible: it would mean that the status of maths-last structural realism depends on whether physicists themselves are realists.

Internalism

The previous pair of metasemantic theories both tried to explain mathematical representation in terms of external factors, such as mental representation or use. What I will call *internalism* rather recovers mathematical representation in terms of the semantic relations *between* models. This proposal is reminiscent of Saussure's structuralist semantics in which the content of a sign is constituted by its relation to other signs.

Dewar (2023) has recently proposed a theory of exactly this type. On what Dewar calls the 'external' approach, one first assigns models an interpretation and then determines whether they are synonymous. The external approach is neutral with respect to metasemantics. On Dewar's alternative, 'internal' approach, one first decides which models are synonymous, which *implicitly defines* an interpretation of them. This view does have metasemantic import: a model represents the world as being a certain way in virtue of relations of synonymy it bears to other models. (At least partially: for a model's empirical content, Dewar seems to prefer a pragmatist approach.)

The next question is where these synonymy relations come from. There are two options: either they are determined by us, or by the models themselves. Dewar opts for the former. For example, he argues that we should *choose* to consider boost-related models of classical mechanics as synonymous because otherwise we are committed to unobservable absolute velocities. This leaves us with much leeway. We could decide that mathematically equivalent models are synonymous, but we could also deny it. This metasemantic account makes maths-last structuralism a matter of convention.

Incidentally, it seems that the very considerations that could weigh in favour of some synonymy convention are anathema to the maths-last approach. For example, Dewar's argument relies heavily on anti-haecceitism, which is exactly the kind of metaphysical thesis that the maths-last structuralist hopes to avoid entirely.

The other option is that models have some in-built standard of synonymy, namely mathematical equivalence. Here we run up against the problem of how to define mathematical equivalence. I see no *a priori* reason to believe that mathematical equivalence is even extensionally equivalent to theoretical equivalence, or in other words, that theories with the same mathematical structure represent the world as being exactly the same way.

To make this concrete, think of the difference between Newton-Cartan theory and Maxwell Gravitation or that between theories on either side of the AdS/CFT duality. Grant that these theories are mathematically equivalent: are they synonymous? This depends on whether the differences between them – the fact that particles in free fall are accelerated in one but not the other, or the fact that spacetime has N or N-1 dimensions – are representationally relevant. The maths-last structuralist wants to say that they are not, but to build this into a metasemantic account of mathematical representation by fiat begs the question.

Therefore, no extant metasemantic theory of scientific representation, when applied to the maths-last approach, entails that mathematically equivalent models are theoretically equivalent. If they are not, there is little reason to deny that maths and language are commensurable – hence little reason to believe that the maths-last approach already embodies a form of maths-last structuralism.

An Egalitarian Alternative

There is a popular picture of theory interpretation on which models are considered as mere tools, which philosophers endow with content by means of a commentary. Rickles (2016) describes the division of labour as follows: "If the primary task of the physicist is to construct models and theories of the world, the primary task of a philosopher of physics is to interpret these products of physics." Maudlin (2018, 5–6) presents a similar vision: "No amount of staring at the mathematics per se can resolve questions like: which mathematical degrees of freedom in the representation correspond to physical degrees of freedom in the system represented? […] The only way is by a *commentary* on the mathematical representation." This picture corresponds to the intuition that the uninterpreted formalism of a theory is hard to understand at best, and altogether devoid of content at worst.

The maths-last approach rejects this picture. It asserts that mathematics can speak for itself. But this does not assuage our worry that we haven't quite grasped what it is that the mathematics tries to say. Any attempt to clarify this, in our language, is supposed to result in distortion and confabulation. There is a sense in which we cannot *say* what a theory means on the maths-last approach. In this, it takes its inspiration from Wittgenstein: "Wittgenstein's later work, if right, puts paid to an alluring pipe-dream that philosophers of language have long had. This is the dream of being able to state what the words and phrases in our language mean, so as to reveal, in some philosophically substantial way, what it is to understand them. […] In order to remain faithful to the spirit of Wittgenstein's later philosophy, one can legitimately theorize about language only by commenting descriptively on various actual practices in which we engage, practices that are already permeated with meaning and that can be described only by one who already knows how to participate in them. That is, one can legitimately theorize about language only from within" (Moore 1985, 138). Although the quote concerns linguistic representation, on the maths-last approach the same would apply to mathematical representation. We can only show, not tell, what a mathematical model means.

Concretely, this means that interpretation does not consist of a commentary on a theory's formalism, but of work within that theory's formalism. There is much invaluable work in this second tradition. For an example, consider Wallace's (2014) defence of the Everett interpretation of quantum mechanics, which consists of results that show (i) the emergence of approximately classical branches from decoherence, and (ii) the emergence of probabilities from quantum amplitudes. These are almost purely mathematical results which are nevertheless philosophically significant. (That said, not entirely uncontroversial philosophical principles such as 'Dennett's criterion' enter at crucial moments; see (Mulder, 2024).) The more prosaic work of linguistic explication, however, is essentially reduced to pedagogy and popularisation: with some violence to the 'true' theory we can make it more accessible to those who don't speak the maths.

This is a coherent alternative to the traditional picture of theory interpretation. But it is unnecessarily stark. The maths-first-language-last approach is authoritarian: language has dominion over mathematics. Both the language-first-language-last and the maths-first-maths-last approach, meanwhile, are totalitarian: they countenance only one mode of representation.

They all enforce a hard distinction between mathematics and language. But this distinction is contrived, and it is time to drop the artifice. In practice, we use both kinds of representation to say what the world is like, as well as others such as mechanical models, pictorial diagrams and hand gestures. It is true, as Wallace says, that when you ask a physicist to explain their theory the final word is often an equation on the blackboard. In the process up to this equation, however, they will have used plenty of other representational resources; for example to specify what their theory is about in the first place, or to clarify which aspects of the mathematics to take seriously and which aspects to consider as a mere scaffold.

I therefore want to propose an egalitarian alternative to both totalitarianism and authoritarianism. We should not see mathematics and language as systems in competition, but as complementary ones. In other words, we should see mathematics and language as only parts of the vocabularies of the language of Nature. Just as English stripped of all colour terms is expressively poorer than English as a whole, so mathematics and language by themselves are less resourceful than mathematics and language combined. Or maybe it is better to think of mathematics and language as if they describe colours in different terms between which there is no systematic relation; they slice up the colour spectrum in incompatible ways (cf. Hacker, 1996)). This observation does justice to the intuition, mentioned earlier, that physics requires mathematics. The view of language and mathematics as complementary entails a weaker form of incommensurability: since mathematics and language have disjoint expressive resources, one cannot say with one what one can say with the other. But this is not the kind of incommensurability that leads to distinct conceptual schemes. For it is possible to use both jointly – and this simply enables us to say more.

This view does not contradict the maths-*first* injunction "to foreground, when doing metaphysics, that the language of physics is mathematics" (McKenzie 2024, 2). But neither mathematic nor language has the last (let alone only!) word. This view does therefore depart from the demand that one should always stick to 'the language of the theory' in the interpretation of that theory (Bradley & Weatherall, 2022), or rather it asserts that the language of a theory is broader than just its mathematical formalism. The language of a theory consists of any and all of the resources we use to eventually describe the world.

The advocate of the maths-last approach may respond that my irenic alternative is nevertheless biased: as soon as we reintroduce language into our descriptions of reality, we are committed to the existence of objects, properties and relations at the fundamental level; and we are forced to treat mathematically equivalent theories as theoretically distinct. This is certainly a possibility on my approach – but unless mathematics and language are truly incommensurable, it is also a possibility on the maths-last approach. On the other hand, it is not the *only* possible outcome. I contend that we can still maintain a structural realism.

How? That is a question I cannot hope to answer here. It will in any case require us to take language less literally and use it more flexibly. To illustrate what this means, it will suffice to survey a few of the available strategies:

Metaphor. We can use language metaphorically to express certain ideas indirectly. Physics is no different: think of an electric *current*, a magnetic *field* or a gravitational *well*. Metaphors are less committal than literal sentences. To borrow an example from Dewar (2015): when someone says "that rock looks like a dragon's head", they are not thereby committed to the existence of dragons. For all we know, it might be possible to reconstrue our talk of objects, properties and relations as mere metaphor.

Weaselling away. It is also customary to explicitly take back what has been said. Melia (2000) calls this 'weaselling away'. For instance, consider the sentence "every dog is good except for Fido". Literally, this is a contradiction, but everyone will understand it as a succinct way to say that every dog that is not identical to Fido is good. In the same way we might take back our commitments to certain objects or properties. The claim that structural realism is committed to 'relations without relata' is an example: we are asked to represent a structured network of objects, and then to take away our commitment to the objects themselves.

Structures as objects. There is no reason to take constant and predicate terms to refer to objects and properties respectively. We can use constants and predicates to refer to structures and quantities, too. The latter are acceptable from a structuralist point of view. For example, the elements of a state space are physical states, rather than concrete objects. Put differently, we can retrofit our syntax to our desired semantics.

Alternative formalisms and novel vocabulary. We are not bound to natural language or first-order logic. There are other formalisms that are better suited to structural realism. Two recent examples are predicate functorese, which Diehl (2018) argues is a language that is not committed to the existence of individual objects, and the calculus of relations, which Carr (2025) proposes as a perspicuous language for relations without relata. Alternatively, philosophers might introduce new vocabulary into English, such as the concept of 'grounding', which McKenzie (2017) uses to reformulate structural realism. We should moreover feel free to translate back and forth between those various formalisms.

These strategies are not without risk: they might yield nonsense at worst and a rather indirect description of reality at best. (In particular, it seems to me that these strategies will not generally satisfy the recent demand for 'perspicuous' representations (Jacobs, 2022; Møller-Nielsen, 2017). What I am inclined to say here is that while the mathematical formalism itself should be as perspicuous as possible, the theory as a whole – linguistic commentary included – need not be.) But that is just to say that some care is required. The point is that the surface form of our language is not a good guide to metaphysics. Just as the truth of the sentence "wisdom is a characteristic of Socrates", to use an example due to Ramsey (1925), should not lead us to believe that universals such as wisdom exist, so the truth of the sentence "Socrates is wise" need not lead us to believe that objects such as Socrates exist. The same holds for mathematics: the fact that general relativity is formulated in terms of a spacetime manifold does not, by itself, mean that it carries a commitment to substantivalism. We are always free to clarify, add to, take back from, correct, nuance, emphasise, improve upon and call into question what we have said.

I therefore cannot subscribe to Galileo's oft-quoted statement: mathematics is not *the* but *a* language of nature. If contemporary physics is any indication, it is the first word on what the world is like, but certainly not the last.

Bibliography

Aune, B. (1987). Conceptual Relativism. *Philosophical Perspectives*, *1*, 269–288. https://doi.org/10.2307/2214148

Bradley, C. (2021). The Non-Equivalence of Einstein and Lorentz. *British Journal for the Philosophy of Science*, *72*(4), 1039–1059. https://doi.org/10.1093/bjps/axz014

Bradley, C., & Weatherall, J. O. (2022). Mathematical Responses to the Hole Argument: Then and Now. *Philosophy of Science*, *89*(5), 1223–1232. https://doi.org/10.1017/psa.2022.58

Callender, C., & Cohen, J. (2006). There is No Special Problem About Scientific Representation. *Theoria: Revista de Teoría, Historia y Fundamentos de La Ciencia*, *21*(1), 67–85. https://doi.org/10.1387/theoria.554

Carr, L. (2025). A style guide for the structuralist. *Noûs*, *59*(4), 873–901. https://doi.org/10.1111/nous.12537

Chakravartty, A. (2001). The Semantic or Model-Theoretic View of Theories and Scientific Realism. *Synthese*, *127*(3), 325–345. https://doi.org/10.1023/a:1010359521312

Chang, H. (1999). History and Philosophy of Science as a Continuation of Science by Other Means. *Science & Education*, *8*(4), 413–425. https://doi.org/10.1023/A:1008650325798

Contessa, G. (2007). Scientific Representation, Interpretation, and Surrogative Reasoning*. *Philosophy of Science*, *74*(1), 48–68. https://doi.org/10.1086/519478

Davidson, D. (1967). Truth and meaning. *Synthese*, *17*(1), 304–323. https://doi.org/10.1007/BF00485035

Davidson, D. (1973). On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association*, *47*, 5–20. https://doi.org/10.2307/3129898

Dewar, N. (2015). Symmetries and the Philosophy of Language. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *52*(Part B), 317–327. https://doi.org/10.1016/j.shpsb.2015.09.004

Dewar, N. (2023). Interpretation and equivalence; or, equivalence and interpretation. *Synthese*, *201*(4), 119. https://doi.org/10.1007/s11229-023-04102-9

Diehl, C. (2018). A Language for Ontological Nihilism. *Ergo: An Open Access Journal of Philosophy*, *5*, 971–996. https://doi.org/10.3998/ergo.12405314.0005.037

French, S., & Saatsi, J. (2006). Realism about Structure: The Semantic View and Nonlinguistic Representations. *Philosophy of Science*, *73*(5), 548–559. https://doi.org/10.1086/518325

Frigg, R., & Nguyen, J. (2020). *Modelling Nature. An Opinionated Introduction to Scientific Representation*. Springer.

Giere, R. N. (2004). How Models Are Used to Represent Reality. *Philosophy of Science*, *71*(5), 742–752. https://doi.org/10.1086/425063

Giere, R. N. (2010). An Agent-Based Conception of Models and Scientific Representation. *Synthese*, *172*(2), 269–281. https://doi.org/10.1007/s11229-009-9506-z

Hacker, P. M. S. (1996). On Davidson's Idea of a Conceptual Scheme. *Philosophical Quarterly*, *46*(184), 289–307. https://doi.org/10.2307/2956442

Hernández-Iglesias, M. (1994). Incommensurability without Dogmas. *Dialectica*, *48*(1), 29–45. https://doi.org/10.1111/j.1746-8361.1994.tb00103.x

Jacobs, C. (2022). Invariance, intrinsicality and perspicuity. *Synthese*, *200*(2), 135. https://doi.org/10.1007/s11229-022-03682-2

Khalifa, K., Millson, J., & Risjord, M. (2022). Scientific Representation: An Inferentialist-Expressivist Manifesto. *Philosophical Topics*, *50*(1), 263–292.

Lewis, D. (1956). Languages and Language. In H. Feigl & M. Scriven (Eds), *Minnesota Studies in the Philosophy of Science* (pp. 3–35). , Vol.

Malpas, J. (2024). Donald Davidson. In E. N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Fall 2024). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2024/entries/davidson/

Maudlin, T. (2018). Ontological Clarity via Canonical Presentation: Electromagnetism and the Aharonov–Bohm Effect. *Entropy*, *20*(6), 465. https://doi.org/10.3390/e20060465

McKenzie, K. (2017). Ontic Structural Realism. *Philosophy Compass*, *12*(4), e12399. https://doi.org/10.1111/phc3.12399

McKenzie, K. (2024). Structuralism as a Stance. *Philosophy of Physics*, *2*(1). https://doi.org/10.31389/pop.77

Melia, J. (2000). Weaseling Away the Indispensability Argument. *Mind*, *109*(435), 455–479. JSTOR.

Menon, T. (2024, January 23). *The inferentialist guide to scientific realism* [Preprint]. https://philsci-archive.pitt.edu/23011/

Møller-Nielsen, T. (2017). Invariance, Interpretation, and Motivation. *Philosophy of Science*, *84*(5), 1253–1264.

Moore, A. W. (1985). Transcendental Idealism in Wittgenstein, and Theories of Meaning. *The Philosophical Quarterly (1950-)*, *35*(139), 134–155. https://doi.org/10.2307/2219340

North, J. (2021). *Physics, Structure, and Reality*. Oxford University Press.
Norton, J. D. (2003). Must Evidence Underdetermine Theory. *The Challenge of the Social and the Pressure of Practice*, 17–44.
Quine, W. V. O. (1960). *Word and Object*. MIT Press.
Ramsey, F. P. (1925). Universals. *Mind*, *34*(136), 401–417. https://doi.org/10.1093/mind/XXXIV.136.401
Rickles, D. (2016). *The Philosophy of Physics*. Polity.
Speaks, J. (2025). Theories of Meaning. In E. N. Zalta & U. Nodelman (Eds), *The Stanford Encyclopedia of Philosophy* (Winter 2025). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2025/entries/meaning/
van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford University Press.
Wallace, D. (2014). *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford University Press.
Wallace, D. (2022). Stating Structural Realism: Mathematics-First Approaches to Physics and Metaphysics. *Philosophical Perspectives*, *36*(1), 345–378. https://doi.org/10.1111/phpe.12172
Wallace, D. (2024). *Learning to Represent: Mathematics-First Accounts of Representation and Their Relation to Natural Language* [Unpublished manuscript].
Weatherall, J. O. (2018). Regarding the 'Hole Argument'. *The British Journal for the Philosophy of Science*, *69*(2), 329–350. https://doi.org/10.1093/bjps/axw012