So, first of all: thanks to the organisers for putting this all together, and especially to Sam Fletcher for inviting me to take part in this session.

As Hans remarks in his preface, the main aim of the book is to introduce a wider philosophical audience to some key formal and technical ideas, that both have been used historically to make certain philosophical claims, and which are useful to making precise other philosophical ideas. By doing so, Hans has done all of us interested in formal philosophy an extraordinary service. The book brings together material that was previously not available in a compact and philosophically sensitive form; it also (perhaps more importantly) serves as an entry point to some powerful new tools, both those developed by Hans and his students (such as Morita equivalence) and those well-known in the mathematics community but still unfamiliar to most philosophers (such as the categorical approach to set theory and logic). Certainly, there's still a lot of material in the book that I haven't digested, and I suspect that my copy is destined to become even more well-thumbed and dog-eared than it already is.

For this session, I felt that the best thing might be to focus more on the philosophical consequences that Hans draws. Since the most sustained discussion of these consequences is in the final chapter—chapter 8—I've decided to use the structure of that chapter as a scaffold. So, what follows is some commentary on the issues Hans discusses in the first four sections of that chapter: Ramsey sentences, the relation between models and possible worlds, Putnam's paradox, and the realism debate as a debate about equivalence. A lot of what I say will be programmatic, hopefully in a way that meshes with the ambitions of the book: at a lot of places I don't have a concrete plan or proposal, but rather a sketch of where I think an interesting project might lie.

1 Ramsey sentences

Concerning the first section (on Ramsey sentences), I want to start by picking up on Hans' comparison of two theories: the theory T (in signature $\{m\}$) asserting that $\exists x(x = m)$, and its "Ramseyfication" T^R, the theory in the empty signature which asserts only that $\exists y \exists x(x = y)$. So a model of T^R is just a set X; and for every $p \in X$, there is a model X_p of T taking p as the extension of m. Now, Hans has the following to say about this case:

> We can see that T and T^R are not intertranslatable (or definitionally equivalent). Nonetheless, there is a sense in which mathematicians would have no qualms about passing from T^R to the more structured theory T. Indeed, once we've established that the domain X is nonempty ... we could say "let

> m be one of the elements of X." This latter statement does not involve any further theoretical commitment over what T^R asserts. (p. 248)

This is a rather intriguing little comment, since it allows that we may desire a notion of equivalence more liberal than definitional equivalence (and which is more liberal in a sense distinct from the sense in which Morita equivalence is more liberal). I want to spend a bit of time exploring this idea, and its connection to Ramsey sentences (in the process, picking up on some of Holger's discussion); especially so, since reflecting on this has made me realise some gaps in my own previous thoughts on Ramsey sentences.

First, as Hans points out, there seems to be some connection to the notion of symmetry going on here. After all, although a permutation of X_p which doesn't fix p isn't an automorphism, it still (in some sense) "preserves the theory": in particular, if we transfer the extension of m to the image of p under the automorphism, then we get another model. But permutation symmetries are tricky, and I think some of the issues might come out a little more clearly if we take a different example; this example will also let us discuss "true" Ramseyfication (where we Ramseyfy second-order, not just first-order, names).

So, consider the following example theory T_H:

$$\forall x(Lx \vee Rx)$$
$$\forall x \neg(Lx \wedge Rx)$$

We can think of this as a primitive theory of handedness, which asserts that everything is either left- or right-handed (but nothing is both). Intuitively, this theory exhibits a symmetry between the notions of left- and right-handedness; and this can be made precise using Hans' definition of a "syntactic symmetry" (§5.5): the translation $F : T_H \to T_H$ given by

$$FL = R$$
$$FR = L$$

is an equivalence of theories. (Indeed, this example is essentially the same as Hans' Example 5.5.12.)

Since the theory T_H treats left- and right-handedness identically, there is an intuitive sense in which it says nothing more than a theory which asserts the existence of an equivalence relation (a relation of "congruence") which has exactly two equivalence classes. We could formalise this latter theory as the following theory, T_C:

$$\forall x Cxx$$
$$\forall x \forall y (Cxy \to Cyx)$$
$$\forall x \forall y \forall z (Cxy \wedge Cyz \to Cxz)$$
$$\forall x \forall y \forall z (\neg Cxy \wedge \neg Cyz \to Cxz)$$

However, the two theories T_H and T_C are not intertranslatable: there is no appropriate way of translating L (or R) from T_H into T_C. (In terms of the dual functor, there is no uniquely privileged way of assigning the labels "L" and "R" to a model of T_C.)

By analogy to the case Hans discusses, we might hope that the notion of Ramseyfication proves helpful in discussing this case. So let's define the following two second-order theories:

$$T_H^* := \{\forall x(Xx \vee Yx) \wedge \forall x \neg (Xx \wedge Yx)\}$$
$$T_H^R := \{\exists X \exists Y \left[\forall x(Xx \vee Yx) \wedge \forall x \neg (Xx \wedge Yx)\right]\}$$

This lets us talk formally about the implications of the theories T_C and T_H for the existence of certain properties. To start with, the theory T_C intuitively entails the existence of (exactly) two equivalence classes. This observation can be formalised by noting that in full semantics, T_C entails the following sentence:

$$\exists X \exists Y (T^* \wedge \forall x \forall y [(Xx \wedge Xy \to Cxy) \wedge (Yx \wedge Yy \to Cxy)]) \tag{1}$$

In other words, T_C entails the existence of a 2-element partition where the two cells of the partition are equivalence classes of C. Note that on full semantics, the first component of this assertion—that there exists a 2-element partition, i.e. T^R—is trivial; it is only the further claim that this partition is a partition *into equivalence classes of* C that requires T_C as a premise.

On Henkin semantics, however, even the first part of the assertion is nontrivial: T^R is not a theorem of Henkin semantics. Nevertheless, one might have thought that it should be a consequence of T_C, since (speaking intuitively) in any model of T_C there are indeed such a pair of equivalence classes (just look at a picture of such a model!). But this isn't the case. For, if it were, then that would mean that on any model of T_C, the two equivalence classes are both definable. But there are models of T_C in which an automorphism maps one equivalence class to the other (namely, those models of

T_C in which the two equivalence classes are equinumerous), and hence neither class is definable. (Recall that definable sets are always invariant under automorphism.)

In other words, what's missing is a good grip on the notion of *applying arbitrary labels*. The problem with Henkin semantics, in this context, is that (speaking roughly) we might think that a theory or model is committed to the existence of some property if such a property is "constructible" in the model, not just if it's definable—where the notion of constructibility weakens that of definability, by not insisting on uniqueness in the way definability does.

How could we analyse this notion of constructibility formally? I'm not sure. But I suspect that Hilbert's epsilon operator will be useful. (I hadn't thought of the epsilon-operator's implications for this context before reading Holger's comments.) After all, what the epsilon-operator does is let you move from having a collection of things with a given property to talking about a specific thing with that property—regardless of whether there is any way of singling out one of the things in particular. So what we might need, at least for this example, is a kind of second-order epsilon-operator, that would allow us to pick out—and then assign labels to—the two equivalence classes of C.

2 Counting possibilities

The next section is about the relationship between possible worlds and models. Hans identifies a debate here between what he calls (following Belot) the "shifty" philosophers, who think that isomorphic models represent distinct possible worlds, and the "shiftless" philosophers who think that isomorphic models represent the same possible world. Now, I am unabashedly in the shiftless camp; so it naturally concerns me that Hans wishes a plague on both our houses! (Well, to be fair, Hans puts such a curse in his own fashion, as "all parties to the dispute have adopted a questionable presupposition.")

Hans gives an example of how he thinks the shiftless philosopher goes wrong, which I think will provide a useful starting-point. We are invited to consider a pair of theories: the theory T_0, in the empty signature, which says that there are exactly two things; and the theory T, in the signature $\{P\}$ (where P is a unary predicate-symbol) which says that there are exactly two things, of which exactly one is P. Next, we take a pair of models M and N of T, both of which have the domain $X = \{a, b\}$; but while P's extension in M

is a, its extension in N is b. Hence, the map $h : M \to N$ such that

$$h(a) = b$$
$$h(b) = a$$

is an isomorphism from M to N.

The problem now comes when we consider the (trivial) translation $I : T_0 \to T$, and its induced functor $I^* : \mathrm{Mod}(T) \to \mathrm{Mod}(T_0)$. I^* maps both M and N to the single model X of T. It also maps the isomorphism $h : M \to N$ to the nontrivial automorphism (let's call it i) from X to itself. But, Hans says, this shows that the shiftless story cannot be right:

> Recall, though, that functors map identity morphisms to identity morphisms. Hence, if the isomorphism $h : M \to N$ is considered to be an identity (as the shiftless seem to do), then it would follow that I^*h is the identity morphism. Thus, contra the shiftless, we cannot identity M and N and forget that there was a nonidentity isomorphism $h : M \to N$. If we do that, then we won't be able to see how the theory T is related to the theory T_0. (p. 261)

I find this argument a little compressed, so I want to suggest the following way of fleshing it out.[1] I think that Hans will be amenable to my reconstruction; if not, I apologise! Consider the following pair of categories, $\mathcal{C}$ and $\mathcal{D}$, of models of T (with the morphisms in $\mathcal{C}$ and $\mathcal{D}$ being elementary embeddings). The objects of the category $\mathcal{C}$ are just the pair of models M and N: as such, the morphisms we have in $\mathcal{C}$ are the isomorphisms Id_M, Id_N, h, and h^{-1}. In the category $\mathcal{D}$, we have just the model M; consequently, the only morphism in $\mathcal{D}$ is Id_M.

Now, here is (my reading of Hans' reading of) the shiftless philosopher's position: $\mathcal{C}$ is, in some sense, just an imperfect representation of what is really going on; although it presents two models, these are (like Hesperus and Phosphorous) merely different labels for one and the same thing—the (unique) possible world represented by the isomorphic models M and N. So actually, we ought to be using the category $\mathcal{D}$ for the purposes of metaphysical theorising. This representation has carved nature at the joints, and trimmed the excess fat into the bargain (to butcher everyone's favourite butchery metaphor): it shows us how things really stand, at the level of worlds rather than models.

But, says Hans, this is a bad idea. For if we devote our attention exclusively to $\mathcal{D}$,

[1] Thanks to John Dougherty for helping me work through these ideas.

then we risk missing out important information about the relationship between T and T_0. In particular, consider now the category $\mathcal{E}$, whose sole object is the model X of T_0 with the elementary embeddings (indeed, automorphisms) Id_X and i. All functors from $\mathcal{D}$ to $\mathcal{E}$ (or rather, the unique functor $H : \mathcal{D} \to \mathcal{E}$) have the property that the only morphism they map to is the identity morphism Id_X: unsurprisingly, since the only morphism in $\mathcal{D}$ is the identity morphism Id_M. But this isn't true of $\mathcal{C}$. From $\mathcal{C}$, there are *two* functors F and G: although F and G agree on objects (sending both M and N to X), F sends h (and h^{-1}) to Id_X whilst G sends h (and h^{-1}) to i. (Hence, $G = I^*$.) If we myopically fixate on $\mathcal{D}$, then we don't see these further, interesting facts—such as the fact that there exist functors to $\mathcal{E}$ which send isomorphisms to i, not just to Id_X (and, indeed, that the dual functor to the translation is an example of such a functor).

One has to be a bit careful here, of course. Hans doesn't want to claim that this shows that in fact, we ought to have been using $\mathcal{C}$ rather than $\mathcal{D}$: the point is just that $\mathcal{C}$ and $\mathcal{D}$, as equivalent categories, need to *both* be embraced by the aspiring philosopher (along with all their many equivalent counterparts). By doing so, we can see which features of the formalism are mere artefacts of that formalism—namely, by seeing which features fail to be invariant under categorical equivalence.

(Incidentally, one might be worried about whether this example really gibes with Hans's claim that only things which are invariant across equivalent categories should be taken seriously: isn't it a problem if $\mathcal{C}$ and $\mathcal{D}$ disagree over how many functors there are to $\mathcal{E}$? The answer is no, for although F and G are distinct functors, they are *naturally isomorphic* (in the sense that there is an invertible natural transformation between them); and so the *category* of functors from $\mathcal{C}$ to $\mathcal{E}$ is equivalent to the category of functors from $\mathcal{D}$ to $\mathcal{E}$. It's turtles all the way down.)

Now, there is a lot here I agree with. In particular, I think Hans is right in advocating two morals. The first (which he allows that shiftless philosophers might be "fumbling their way toward") is this:

> A theory T is indifferent to the question of the identity of its models. In other words, if M and N are models of T, then T neither says that $M = N$ nor that $M \neq N$. The only question T understands is: are these models isomorphic or not?

And the second is the "positive proposal" that

> the philosopher of science shouldn't say things about $\mathrm{Mod}(T)$ that are not invariant under categorical equivalence, nor should they argue over questions—such as "how many models does T have?"—whose answer is not invariant

> under categorical equivalence.

But I don't think either of these morals require that I give up my shiftless ways! After all, the shiftless philosopher's claim was that isomorphic models represent the same possible world: and that claim is indeed independent of whether there is just one isomorphic model, or two, or many.

However, I do think that Hans' critique forces us to give up a certain picture of what such a position involves. That picture is one according to which the "right" way to engage with the modal structure of a theory is to always insist on talking directly, or immediately, about the worlds described by the theory; and to repudiate analysis of the models of the theory as metaphysically impious. We've already seen the role that this picture plays in Hans' argument: it was this picture that was driving the idea that because $\mathcal{D}$ has a one-to-one correspondence between models and worlds, it is *better* than $\mathcal{C}$ in all respects—so much better, in fact, that once one has formulated $\mathcal{D}$, the use of $\mathcal{C}$ should be abjured.

In other words, I think that shiftless philosophers should combine their metaphysical parsimony (about the number of possible worlds) with representational profligacy. Indeed, I think that the shiftless position should be reconceived, along the lines that Hans suggests in §8.4: as a proposal for what we *mean* by possible world. On this proposal, a possible world just is the invariant content across a number of representationally equivalent models; so our understanding of possible worlds is posterior to our judgment that certain models are equivalent, not prior to it. As Hans says (in §8.4), about the suggestion that the equivalence of sentences means the identity of the propositions they express:

> Now, I don't disagree with this claim; I only doubt its utility. If you give me two languages I don't understand, and theories in the respective languages, then I have no way of knowing whether those theories pick out the same propositions. (p. 271)

In making this kind of move, we can also make contact with certain traditions of neo-Kantian thought, which hold that the mistake of "dogmatic" metaphysics is always to seek the one, privileged, transcendental account of reality:

> It is only if we resist the attempt to press together the totality of forms (which here offer themselves to us) into a final *metaphysical* unity, into the unity and simplicity of an absolute "world-basis", that their authentic concrete content and their concrete abundance reveal themselves to us. ... The

> naive realism of the ordinary worldview ... always, of course, falls into this mistake. From the totality of possible reality-concepts, it separates some individual one out and erects it as norm and exemplar for all the rest.[2]

It's right to recognise that genuine, objective fact is the invariant core across different languages; but a mistake to take this as a reason to be a monoglot.

3 Putnam's Paradox

Third, Hans looks at what his formal tools have to make of Putnam's famous model-theoretic argument against metaphysical realism. In his terms, this argument turns on the observation that for any consistent theory T, we will (modulo concerns about cardinality) be able to find a model W of T whose domain is that of the world. Now, Hans says that he doesn't have a great deal of time for Putnam's argument:

> . . . let me be completely clear about my view of the argument: it is absurd. This version of Putnam's argument is not merely an argument for antirealism, or internal realism, or something like that. This version of the argument would prove that all consistent theories should be treated as equivalent: there is no reason to choose one over the other. (p. 263)

But I have to admit, I'm not entirely sure why he takes such a dim view of Putnam's argument. After all, the conclusion of the argument is *meant* to be absurd: one can't have a reductio without an absurdum! What seems to me to be missing from Hans' discussion of Putnam is sufficient recognition of the fact that the assumptions Putnam makes are not assumptions he takes to be correct, but rather to be assumptions *of the metaphysical realist*. (So, in this regard, I agree with Holger's analysis of Hans on this point.)

Consider, for instance, Hans' admonition (on p. 264) "that nobody here—including Putnam—is free from language and theory", and compare it Putnam's characterisation of metaphysical realism:

> *Metaphysical realism* . . . is, or purports to be, a model of the relation of any correct theory to all or part of THE WORLD. . . . Let us set out the model in its basic form. In its primitive form, there is a relation between each term in the language and a piece of THE WORLD (or a *kind* of piece, if the term is a general term). . . . there has to *be* a determinate relation of *reference* between

[2] (Cassirer, 1921)

> terms in L and pieces (or sets of pieces) of THE WORLD, on the metaphysical realist model . . . What makes this picture different from *internal* realism (which employs a similar picture *within* a theory) is that (1) the picture is supposed to apply to *all* correct theories at once (so that it can only be stated with "Typical Ambiguity"—i.e., it transcends complete formalization in any one theory); and (2) THE WORLD is supposed to be independent of any particular representation we have of it—indeed, it is held that we might be *unable* to represent THE WORLD correctly at all (e.g. we might all be "brains in a vat", the metaphysical realist tells us).[3]

In other words, it is precisely because the metaphysical realist assumes himself to be free from both language and theory (as asserted by points (2) and (1) respectively) that he opens himself up to the "absurd" conclusion of Putnam's argument.

In further support of this, consider Hans' discussion of the idea that "for Putnam's argument to go through, he needs to be able to make distinctions in W that simply cannot be made by users of the theory T." (p. 264) In other words, Hans observes that it is only from the perspective of a metatheorist that we can raise the sorts of worries about malfunctioning reference that underpin the arguments of Putnam, and goes on to ask

> But how does the metatheorist's language get a grip on the world? . . . Now, Putnam might claim that it is not he, but the realist, who thinks that the world is made of things, and that when our language use is successful, our names denote those things. So far I agree. The realist does think that. But the realist can freely admit that even he has just another theory, and that his theory cannot be used to detect differences in how other people's theories connect up with the world. (pp. 264–265)

I contend that this is exactly what the realist (as envisioned by Putnam) cannot admit: after all, Putnam is the one who notoriously argues that any appeal to trans-linguistic factors (e.g. causal or natural structure) to determine reference is "just more theory"! Hans has, I think, underestimated how immodest a thesis metaphysical realism is.

To finish off this piece of Halvorson exegesis, it only remains to make sense of Hans' claim that he can "neither affirm nor deny" the charge that he has "simply affirmed Putnam's conclusion", on the basis that there is too much "unclarity in the meaning of 'internal realism'." (p. 265) Towards an answer, consider the joke that David Foster Wallace tells in his 2005 commencement speech at Kenyon College:

[3] (Putnam, 1977, pp. 483–484)

> There are these two young fish swimming along and they happen to meet an older fish swimming the other way, who nods at them and says "Morning, boys. How's the water?" And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes "What the hell is water?"

I think that there might be something similar going on here: I think Hans' difficulty in making sense of internal realism is that its central precepts—the language- and theory-dependence of our concepts and utterances, and the impossibility of a transcendental semantics—don't even register as the precepts of a viewpoint, so much as background preconditions of making sense of representational practice.

To close out this section, I want to pick up on a different part of the book (which Professor van Fraassen already discussed), where Hans gives a rather deflationary account of the role of logical semantics:

> . . . the picture typically presented to us is that logical semantics deals with mind-independent things (viz. set-theoretic structures), which can stand in mind-independent relations to concrete reality, and to which we have unmediated epistemic access. Such a picture suggests that logical semantics provides a bridge over which we can safely cross the notorious mind-world gap.
>
> But something is fishy with this picture. How could logical semantics get any closer to "the world" than any other bits of mathematics? . . .
>
> In what follows, we will attempt to put logical semantics back in its place. The reconceptualization we're suggesting begins with noting that logical semantics is a particular version of a general mathematical strategy called "representation theory". . . . In all such cases, there is no suggestion that a represented mathematical object is less linguistic than the original mathematical object. (pp. 164–165)

Similarly, Hans claims in the introduction (again, in a passage already discussed by Professor van Fraassen) that

> logical semantics is ... wait for it ... just more mathematics. As such, while semantics can be used to represent things in the world, including people and their practice of making claims about the world its means of representation are no different than those of any other part of mathematics.

It seems to me that this program of cutting logical semantics down to size is remarkably consonant with some of Howard Stein's remarks on the realism-antirealism debate, which (besides being apposite, given we're in Chicago) can also help illuminate the connection to Putnam's paradox:

> I want to describe ... the other side of the trouble with the appeal to reference. It is simply that one can "Tarski-ize" any theory that is constructed in the standard way—that is, with the help of a metalanguage of suitable strength, and also containing the extra-logical resources of the language of the theory T under discussion, one can supplement T with a theory of reference for it, T'; and can do so in such a way that whenever θ is a theorem of T, the statement that θ is true is a theorem of T'. Thus, ... the Tarskian theory of reference and truth in a rather serious sense trivializes the desideratum put forward by the realists. ... The point can be restated this way: The semantics of reference and truth (for a given theory) is itself a theory.[4]

4 Realism and Equivalence

But if the affinity with Stein holds good, then it's a consequence of the views Hans expresses here that we should be dissatisfied with "standard" ways of stating the realism-antirealism debate (at least, the semantic portion of that debate)—after all, the quotation above occurs in the context of a deconstruction of the realist-instrumentalist distinction. But Hans doesn't take that debate to have been vitiated; rather, he thinks it needs to be reconceptualised as a debate about which *other* debates are worth having: that is, as he puts it, how liberal or conservative a notion of equivalence one ought to employ.

I think the large and interesting question to discuss here is the ways in which we can see this as connecting to broader pragmatist traditions. But because I don't have the time, space, or ability to discuss that question, I'm going to talk instead about a small and boring issue. Essentially, I just want to discuss one of Hans' examples: specifically, the example of how the debate over "quantifier variance" could be put into these terms. To remind you, Hans discusses this issue in Chapter 5, and introduces his characters of Niels the Nihilist and Mette the Mereological Universalist. Niels' theory is formulated in the empty signature, and says merely that there are exactly two things; Mette's theory has a signature with a binary relation symbol P (to represent parthood), and consists of axioms asserting that P is a strict partial order, that there are exactly two atoms of P,

[4] (Stein, 1989, pp. 50–51)

and exactly one thing of which the two atoms are each parts. For convenience, we also introduce the unary predicate A ("is an atom"), defined by

$$\forall x(Ax \leftrightarrow \neg\exists y Pyx) \tag{2}$$

To translate from Niels' theory to Mette's, we (in effect) interpret Niels' use of the term "thing" or "object" by Mette's term "atom"—formally, by taking the "domain formula" in the translation to be Ax. Translating from Niels' theory to Mette's is not so straightforward, but provided that Morita equivalence is admitted, this is also doable. In effect, Niels takes Mette's "thing" to mean "two-element multiset of things" (recall that a *multiset* is like a set, except that the same object can recur multiple times), by thinking of them as quotients of pairs: two pairs define the same multiset if they are related by a permutation. So, Niels is able to agree with Mette that there are three Mettean things: if his things (Mette's atoms) are a and b, then Niels will recognise that there are three multisets $\{a, a\}$, $\{b, b\}$, and $\{a, b\}$. Moreover, it turns out that these two translations play nice with one another (i.e. they're almost inverse), so we can indeed regard Niels and Mette as advocating equivalent theories.

Now, I'm quite happy with this conclusion. But my worry is that it doesn't go far enough. I wanted a way to resolve the universalism and nihilism debate quite generally, not just in this special case! And the strategy outlined here will, clearly, not permit us to translate from a nihilist theory asserting that there are *three* objects to the corresponding universalist theory (the one that asserts that there are three mereological simples, from which a further four composites can be made). I think one could come up with an extension of the strategy used for Niels and Mette: Niels will treat composites as quotients of triples, with two triples regarded as equivalent if they feature the same objects (so, for instance, $\langle a, b, b\rangle$ will be equivalent to $\langle b, a, b\rangle$); and one will similarly be able to cook up a strategy for any case where our nihilist asserts the existence of some fixed finite number of objects (and the universalist asserts the existence of that same number of composites). But this is all still pretty limited.

I don't have a recipe for how to escape this limitation, although it's reasonably clear where the story will have to go: evidently, if Niels wants to be able to match Mette's expressive resources, then we will need to be able to give him the ability to talk about *collections* of mereological simples. So the question becomes: is there a plausible notion of theoretical equivalence, relative to which a first-order theory could be regarded as equivalent to a theory with second-order quantification, or plural quantification, or set-theoretic resources, or mereological resources? It does seem to me that such a view should, at least, admit of articulation—although it will be well to the left of Morita equivalence. Indeed, this whole discussion is really just an illustration of Hans' observations (p. 275) that "Morita equivalence isn't all that liberal", and that "the relation 'being a logical construct of' differs from the mereological parthood relation."

Let me close this section by heading off one worry that you might have about this proposed criterion of equivalence: it is generally accepted that the logicist program is untenable, i.e. (in Hans' terms) that the empty theory—even in second-order logic—is not equivalent to second-order arithmetic. But given that one can construct a model of arithmetic in set theory, then if we can augment any theory with set-theoretic resources and thereby obtain an equivalent theory, won't we be able to make the something of arithmetic out of the nothing of pure logic? (Incidentally, this also suggests an interesting project using Hans' book: to evaluate the debate over logicism and neo-logicism in terms of equivalence.)

However, this overestimates just how liberal the criterion of equivalence would need to be. Note that in order to regard the nihilist and universalist as equivalent, we don't need to allow the nihilist *arbitrary* or (better) *iterative* set-theoretic resources: Niels only needs to be able to talk about sets of simples, not sets of sets of simples. (Indeed, there's a sense in which this is the key formal difference between mereology and set theory.) Given that set theory's enormous power derives essentially from that iterative ability, then we need not allow that Niels' theory is equivalent to full ZF or ZFC. Of course, it would follow that Niels' new theory, with quantification over sets of simples, would indeed be equivalent to a theory which quantifies over sets of sets of simples; but provided we insist that only finite chains of these equivalences are permitted, then (hopefully!) we can keep a check on things.

5 Conclusion

As I said at the start, there's still a lot of the book that I'm digesting. In particular, I would consider myself to still be very much a novice in the kinds of categorical methods that Hans uses so successfully in the book (and certainly, working through the book has been a good reminder of that fact). But what I'm excited about is the extent to which the book feels like a starting-point: each of the four sections above naturally suggests a project that I think would be interesting and worthwhile. Holger has already mentioned how the works of Carnap or the Munich structuralists can be seen as forebears of Hans' work. If I may, I'd like to suggest another parallel. The development of possible-world semantics for modal logic was, in many ways, *the* technical development that led to the flourishing of modern analytic metaphysics. To me, it feels that the methods Hans is seeking to proselytise are at least as fruitful, and at least as suggestive for future philosophical inquiry, as this. It's also true that the kind of project suggested is very, very different; but that's all to the good, insofar as philosophical progress is made possible by successive paradigms of inquiry (the cynic might say, waves of fashion). So I'm excited to see where the paths Hans has opened up might lead.

References

Cassirer, E. (1921). *Zur Einsteinschen Relativitätstheorie*. Bruno Cassirer, Berlin.

Putnam, H. (1977). Realism and reason. *Proceedings and Addresses of the American Philosophical Association*, 50:483–498.

Stein, H. (1989). Yes, but... Some Skeptical Remarks on Realism and Anti-Realism. *Dialectica*, 43(12):47–65.