Supervenience, Logic, and Empirical Content: Commentary on Hans Halvorson, *The Logic in Philosophy of Science*

(Contribution to the APA Central 2020 'Author Meets Critics' Symposium Febr. 2020)

Bas C. van Fraassen

Abstract

Halvorson's book's real achievement is that it is both a source and a challenge, and not just for philosophers of science. I will begin with some notes to add to Halvorson's discussion of supervenience and definability. Then secondly I will engage the book's way of dealing with empirical content. Extension of formal methods to the relation of theory to world, as mediated by experiment and measurement, seems to me crucial to its value, and I will make three suggestions for this. Then thirdly I will turn to the tantalizing hints Halvorson gives us of an overall view of logic and language, and speculate about how that would answer questions about scientific representation and more specifically about the object language / metalanguage relation.

Halvorson's book's real achievement is that it is both a source and a challenge, and not just for philosophers of science. Analytic philosophers generally should be drawn in by his discussions of Carnap, Quine, and Putnam. Once there they will be confronted with quite a few waves to rock their boat. The part on implicit definability, reduction, and supervenience could raise some issues in philosophy of mind to a new level of sophistication, and if they do the work to understand Morita equivalence they'll realize how simplistic ideas about theoretical equivalence have been heretofore.

The book presents itself at first blush as a textbook; it is more than that, but its first mission could be to educate a new generation of philosophy of science students. Hans means to give them cutting edge formal methods for the study of theory structure. Way back when, in 1970, I had a similar hope.[1] But 2020 is not 1970. Today the technical competence required for entry into the philosophy of science is at a much higher level. Hans is continuing the formal approach to philosophy of science. This continuation includes rightful correction as well as innovation, and I am grateful for both (I will include a *mea culpa* below). We have moved on, in these last fifty years, and there is a whole new world waiting for the new, powerful, formal methods that Hans presents.

I will begin with some notes to add to Halvorson's discussion of supervenience and definability. Then secondly I will engage the book's way of dealing with empirical content. Extension of formal methods to the relation of theory to world, as mediated by experiment and measurement, seems to me crucial to its value, and I will make three suggestions for this. Then thirdly I will turn to the tantalizing hints Halvorson gives us of an overall view of logic and language, and speculate about how that would answer questions about scientific representation and more specifically about the object language / metalanguage relation.

1] On supervenience

Let's begin with the section "Beth's Theorem and Implicit Definition" in Chapter Six, which highlights the relevance to discussions of reduction, supervenience, and the possibility of non-reductive physicalism. It gives a rude shock to the familiar slogan of *supervenience without reduction* in philosophy of mind.

Halvorson focuses on what I would call local supervenience claims, which concern the relation of a single factor described in one vocabulary to a finite set of factors described in another vocabulary. I imagine examples like the claim that intelligence supervenes on neurological features, or perhaps that an untoward sense of humor supervenes on the constellation of the four cardinal humours in the body. Put in the formal mode it is a claim that relates certain terms and Hans takes this to be precisely the claim of implicit definability. By Beth's theorem that implies explicit definability – aha! there is no sense to 'supervenience without reduction'.

Global supervenience

I would like to elaborate on this, for the task is not complete without attention to the idea of global supervenience, which Halvorson broaches briefly in his note about Petrie (pp. 203-204). As Halvorson does, I want at once to paraphrase Petrie's talk about properties in the formal mode, switching to the context of first-order logic, rather than immerse myself in analytic metaphysics. And my paraphrase would be as follows.

Let Σ^1 and Σ^2 be two disjoint signatures, and Σ^+ their union. Assume that the models of Σ^+ are just those that are 'allowed' in a specific context (not ruled out by some background theory formulated in Σ^+). Perhaps the first signature is the vocabulary of folk psychology and the second that of physics, or – to make contact with the section on Ramsey sentences – the first is a theoretical vocabulary and the second an observation term vocabulary.

[1] Rueful note: I wrote a textbook, *Formal Semantics and Logic,* to introduce especially philosophy of science students to places beyond the logic they could find in Carnap, Hempel, Nagel and their kin. So I had a similar aim to Hans' today, and there were other similarities too because I had fallen in love with topological methods, using ideals, filters, ultraproducts for completeness and compactness proofs, and really wanted students to get to know about them. Well, the book did not become a best-seller. It did not educate a whole new generation of philosophy of science students, not even when the semantic approach to theories took hold.

A claim of global supervenience allows for the possibility that the pertinent relation does not hold between finite texts, and perhaps not even between infinite long explicitly or recursively definable sets of sentences. Perhaps a psychological variation could correlate only to a variation, impossible to specify explicitly, affecting every single physical phenomenon! Informally, or rather half-formally, I would phrase it like this:

> **Global supervenience of Σ^1 discourse on Σ^2 discourse:** for any text F formulated in Σ^1, if F is true then the world could not be different in that respect unless something formulated in Σ^2 would also have a different truth-value.

To make this precise, let's think of the relation between these two parts of discourse as follows. The models of Σ^+ can be partitioned into the family $Q^{(2)}$ = {MOD(T): T a maximal consistent set of sentences in signature Σ^2}.

Suppose that F, a text formulated in Σ^1, is true in model M and false in model M'. Then these two models cannot belong to the same member of $Q^{(2)}$, or supervenience would be violated. So if M does lie in MOD(T), a member of $Q^{(2)}$, then all of MOD(T) is part of MOD(F). And similarly, if M' is in MOD(T'), a member of $Q^{(2)}$, then MOD(T') is entirely disjoint from MOD(F).

From this it follows that MOD(F) is the union of cells in the partition in question, that is, the union of the family $F^{(2)}$={MOD(T) in $Q^{(2)}$: MOD(T) is part of MOD(F)}.

If we can think of union as the infinitary form of disjunction then, as far as truth-value is concerned, any text F in Σ^1 is a disjunction of texts in Σ^2. So just as Hans does, we arrive at the conclusion that there is here no supervenience without reduction.

It is different, because 'infinitary disjunction' is not something expressible in the language. In fact it could not be in general, for there is no implication here that the family $F^{(2)}$ is even definable by any criterion of definability. The 'reduction' or 'translation' of the former discourse into 'disjunctions' of the latter is, in general, not a humanly graspable translation. It is not, in general, a translation sentence by sentence, paragraph by paragraph, book by book, recursively definable set by recursively definable set, nor …. continue as you like. It is, as Pascal would have phrased it so well, a translation or reduction *at the far end of infinity*, not for us.

Should this qualify Hans' eyebrows-raised presentation of the physicalism story? I do not think so. Global supervenience could consist in a matching between indefinable sets of sentences in the two signatures, and yes, that matching could exist in Cantor's paradise or its category-theoretic equivalent. But to assert such a thesis of supervenience in support of physicalism turns physicalism into a thesis designed to be irrefutable. And what's the value of that?

Modified concepts of supervenience and reduction

Discussion with the others in the symposium lead me to add two remarks about this. The first is a reaction to Holger Andreas' well taken point that in Carnap's view of the relation between observational and theoretical vocabulary we need to do justice to the admissibility of less than total constraint. On Andreas' view, the truth-values of the observation base sentences are definite, and those constrain the truth-value assignment to the theoretical sentences, but the constraint is not total, and does not yield a unique classical valuation. Andreas then handles that by introducing supervaluations, which assign just what is common to all the classical valuations that satisfy the constraint.

Does my argument above extend properly to that insight? What is required is to take the case in which the truth-values of sentences of Σ^2 are definite, but not necessarily those of Σ^1. The informal definition of global supervenience can be verbally the same. But we have to switch now from the models of Σ^+ to supervaluations constructed from them. Although the constraints on values assigned to Σ^1 sentences are not specified, we do know this. For each supervaluation s there is a set C(s) of Σ+ models, such that s assigns T (respectively F) if and only if all members of C(s) do so, and secondly, that the values of Σ^2 sentences are definite, that is,

> If G is a Σ^2 sentence then all members of C(s) assign the same truth-value to G.

This implies that each supervaluation assigns either the value T or the value F to each Σ^2 sentence.

Because of this, the supervaluations of language Σ^+ can be partitioned into the family $Q^{(2)}$ = {SUP(X): X a maximal consistent set of sentences in signature Σ^2}, where of course by SUP(T) I mean the set of supervaluations that assign T to all members of X.

Consider now a sentence or theory F formulated in Σ^1, and the set SUP(F). Suppose SUP(Y) is a member of $Q^{(2)}$ and that there is an overlap between SUP(F) and SUP(Y). So there is a supervaluation that assigns T to all of F and all of Y – now if there are also members of SUP(Y) which do not assign T to F, we have a violation of supervenience. For in that case, the full description of the world in Σ^2 is the same, but the value of F is not.

Hence to have no violations of supervenience, SUP(Y) must be a proper part of SUP(F) … or else entirely disjoint of course. But then it follows, since this pertains to all members of $Q^{(2)}$ that SUP(F) is the union of a family of cells of the partition $Q^{(2)}$. So the conclusion is the same as above.

The second remark is in response to what Nick Huggett pointed out about Fodor's classic discussion of supervenience. It is clear that Fodor appreciated that a case of supervenience might amount to equivalence to a disjunction in the basis language. What bothered him was not any question about whether that would be expressible in the language, and perhaps he was only thinking of local supervenience. What bothered him rather is that the disjunction might be of a gerrymandered set of disjuncts with little unity. If I understand him correctly, he wanted the notion of reduction to include clauses about lawlike connections.

With the bar raised for the notion of reduction there would indeed be a reasonable sense in which one could be claiming supervenience without reduction. Agreed. But the conclusion remains: supervenience amounts to a logical equivalence, whether visible in the object language or only in the metatheory. It remains, as I see it, that so-called non-reductive physicalism is still just the old familiar materialism with good PR.

2] The treatment of empirical content

Halvorson does not just teach the new concepts and methods now available; he does so in the course of presenting a powerful new analysis of theory structure, with constant attention to its connection to theories in the natural sciences.

Nevertheless the book's attention to the question of empirical content is inconclusive. We would like a clear insight into the relation of formal analysis to the mire and blood of scientific theorizing. To achieve that, the analysis needs to be extended to an analysis of the relation, between theories and the world, that is characteristic of the empirical sciences, and that means, a relation constituted by experimentation and measurement.

In section "Empirical theories" of Chapter 4 Halvorson takes up Carnap's attempt to isolate the empirical content of a theory through a division of the vocabulary into observational and theoretical terms. He is fully aware of the critique this received in the sixties, takes us through some of the history of this critique, and then relates the conviction reached in the seventies that Carnap's attempt had failed. But the discussion is so fair and even-handed, and so apparently reluctant to take up the arguments between Carnap and his critics, that it remains inconclusive. We are left with is a sense that something hadn't worked, though some of it might not be so bad in retrospect, but … .

Halvorson turns back to this subject much later in the book, with a discussion of selective scientific realism and Ramsey sentences. And here, in this last chapter devoted to more general philosophical issues, we find Halvorson simply going with Carnap's scheme. Scientific theories are represented as theories T_i written in a signature $\Sigma \cup \Sigma_i$, the former part, common to all, being apparently the 'observational' or 'old' terms as before.[2]

Perhaps this is not out of order if one wants only to connect with some older literature on Ramsey sentences and the like. The question is what more could be done.

Formal methods can have the impact and influence in philosophy of science that they richly deserve to have but, I submit, only if they be extended to the relation between scientific theories and what those theories are about, *as mediated by experimentation and measurement*.

And to be positive, I am in fact convinced that formal methods can be extended to an analysis of the practices of experimentation and measurement, and thereby to the empirical relation between scientific theories and what those theories are about.

To add this positive note to my critical remarks, I'll add three suggestions which could perhaps provide some directions for doing this, which could perhaps, I only say perhaps, offer a way into this continuing project.

---

[2] On page 129 Halvorson discusses formalizing a theory in a many-sorted language, with distinct sort symbols for observable objects and theoretical objects. In this language there would be no well-formed expression corresponding to "No theoretical entity is an observable entity". But there is no real division between entities introduced by postulate for theoretical purposes (some of which are observable and some not observable) and other things.

2-1. Theoreticity of measurement

A measurement is a procedure with an outcome, but that does not by itself define what a measurement is. The questions of what counts as a measurement procedure, and what it measures if it is a measurement, are in the main *theoretical questions,* answered on the basis of a theory. While in any theoretical context, certain quantities count as directly measurable, others are theoretical in a precise sense that relates to how they are measured.

This was shown originally by Mach and Poincaré, but was introduced into the later literature by Joe Sneed:

> a quantity is theoretical, in or for a given theory, if and only if its values are determined on the basis of direct measurement data via equations provided by the theory itself.

So for example, measurement of Newtonian mass on a given system involves calculations based on Newton's laws, hence assumes or presupposes that the system in question is a Newtonian system. That is what it means to say that mass is a theoretical quantity of Newtonian mechanics.[3]

This point, the theoreticity of measurement, needs to be kept as central to our entire discussion. It was exploited specifically around 1980 by Clark Glymour in his theory of relevant evidence, which drew on the account by Hermann Weyl in his *Philosophy of Mathematics and Natural Science* (Glymour 1975, 1980).[4]

To show clearly how Glymour's account relates to that account of theoreticity, I will here put it in the form that takes the theory and hypotheses to be equations, whose solutions are assignments of values to the quantities in those equations. Think of a theory as thus representing its subject matter with three ingredients: a state-space *S,* an algebra of physical quantities *Q,* and a dynamics *D*.

Then, with the data E, E' being sets of outcomes of direct measurement, Glymour's central empirical relation between theory and world is the following (using my own words for it):

E *bears out* H *relative to* theory T:

<u>exactly if</u>

E has some alternative E'

and T has some sub-theory T'

such that:

(1) T + E + H has a solution.

(2) T' + E' has a solution.

---

[3] For the theoreticity of measurement, crucial to the whole story after the demise of all forms of naive empiricism, see van Fraassen 2012.

[4] If Glymour's work was not as influential as it should have been, that was due in part to the lack of a clear distinction between evidential support and confirmation. I am using my own words to avoid that. Some of the unclarity was perhaps unavoidably due to the felt obligation to present the account in terms familiar from Hempel's writings, dominant at the time.

(3) All solutions of T’ + E are solutions of H.

(4) No solutions of T’ + E' are solutions of H.

Closeness to actual practice increases if we concentrate on the case in which H concerns the value of a particular theoretical quantity ***q***, and we grade the informational value of the data set by asking which of the following is the case:

(1) All solutions of T’ + E assign a value to ***q*** in the same specific interval I;

(2) All solutions of T’ + E assign the same value to ***q***.

Viewed in this way, the procedure is a measurement of that quantity, it makes explicit the form of measurement of a theoretical quantity.

This can certainly all be cast in a first-order language, with the data and hypotheses replaced by corresponding sentences, the theory and sub-theory cast as sets of sentences, all with the same signature, and the solutions as models keyed to that signature. But the division of that signature to delineate the theoretical part of the vocabulary could not be there at the outset. It is a demarcation line that can shift from theory to theory, even if those theories are written in the same vocabulary.

The form of verbal representation, whether as equations or in some other way, is not the problem. The challenge is to extend the formal analysis, in whatever form, to measurement and its relation to theory in ways that will be recognized as fruitful in the larger philosophy of science community.

2-2. Toward an abstract logic of experimentation and measurement

Drawing on Glymour’s account I stipulated that the data are not just random bits of information but outcomes of direct measurement. That distinction is, however, outside the account; the concept of data, derived from measurement outcome, remains there as a ‘black box’, with no look inside.

To go beyond this, and still with the intent to find fruitful application of formal methods, I suggest a look at work also done around that same time, but mainly by mathematicians and physicists who were exploring alternatives in the foundations of physics.[5] Their motivation appeared to be this conviction, to which we cannot now agree:

> that there must be a structure that any physical theory must have, and this structure must be determined by the structure of possible experiments and measurements belonging to the theory’s domain or subject matter.

From an outsider’s point of view there was a detectable circularity in the way they carried out this project. There is very little that can be determined a priori about the structure of possible

[5] An early introduction to this approach is found in Emch and Jauch (1965), and expositions of how this was continued in Foulis and Randall (1974, 1978). Chapters 16, 19, and 20 of Beltrametti and Cassinelli (1981) places this in a context relating to the further work by, for example, G. W. Mackey. My exposition here is not meant to match precisely what any of these did. I mean to convey in a general way the project that seemed to be shared in a plethora of conferences and articles encountered in that period.

experimentation or measurement, and to fill in more, they took cues from what measurement looks like in quantum mechanics, their main target theory.

But if we go to this project with the insight that what counts as a measurement, and how measurements must be related to each other, are in the main theoretical questions, then I think this project could open the black box of data and measurement results in the epistemology of science.

Characterizing measurement set-ups

The project begins by characterizing a measurement as consisting of a procedure plus an outcome. A *measurement set-up* encompasses all the measurements with a specific procedure.

For a given procedure we take there to be a finite set of discernible basic results, such as the numbers from 1 to 10. But when the procedure is carried out, the result may not be entirely definite, it might be something like "greater than 5". So while the set-up is individuated by its procedure, we can represent the encompassed measurements with their outcomes by the sets of basic results that thi s procedure could have. That is a finite Boolean algebra of sets, which we may call the *events* in that set-up.

Speaking naturally we would actually say that it is a Boolean algebra with the empty set missing – which, as Tarski pointed out, is exactly Lesniewski's mereology. But it is convenient to include the *impossible measurement* -- that is, the impossible case where the procedure is carried out successfully but there is no outcome. That is then represented by the empty set of possible outcomes, so as to complete the algebra.

To illustrate how logical relations enter, suppose the procedure consists in feeding coins into a coin sorting machine like they have in supermarkets and banks. It is an obvious move to say that the measurement with outcome "dime or quarter" is *implied* by the one with outcome "quarter" and is *orthogonal* to the one with outcome "nickel". That parallels set-theoretic relations, and this example remains within a single set-up.

So far, then, for the representation for a single measurement set-up. The subject becomes interesting, and difficult, only when we start asking about structure that cuts across different measurement set-ups. The question makes sense, for those logical relations can be explained, though in an extraneous fashion, in modal terms. Then they can apply across set-ups as well:

> <u>Effective implication:</u> **if the former would result** (when the procedure is implemented) **then the latter would result** (if the procedure were implemented).
>
> <u>Orthogonality</u>: **the former would result** (when the procedure is implemented) **if and only if the latter would not result** (if the procedure were implemented)
>
> <u>Compatibility</u>: there is a measurement **that effectively implies both.**

So understood, these relations can hold between measurements in different set-ups. And so understood, we infer at once that implication is reflexive and transitive while orthogonality is irreflexive and symmetric. The entire world of measurements in a certain domain can thus be represented as a partially ordered set with additional structure, but with the partial order relation connecting elements belonging to different Boolean subalgebras.

But how can questions be answered about just which of these relations hold?

Theory classification in terms of the algebra of measurement operations

A metaphysician would presumably look to necessities in nature to answer questions about the structure imposed by these relations, of implication and orthogonality, on the world of measurements. In practice, however, those are theoretical questions answered by theories when the measurement procedures are in their domain of application.

Writers in this area tended to take some such answers for granted, and that was a flaw. Instead of trying to provide the answers by fiat, we should rather say that *scientific theories can be classified by how they answer them*.

For example, it is typical to describe the difference between classical and quantum physics by saying that in the former, all measurements are compatible. That is not wrong, as long as we keep in mind that it is the theory itself which determines what counts as a measurement.

That said, we need to have ways into the subject if this classification is to become insightful. Remember the abstract schema I used for the discussion of Glymour: a theory has three ingredients: a <u>state-space</u> *S,* an algebra of <u>quantities</u> *Q*, and the <u>dynamics</u> *D*, which is a semi-group or group of <u>transformations</u>. A bridge is needed between that theoretical structure and the structure we see in the world of experimentation and measurement.

(a) Transformations

One way is this: think about the special case of measurement procedures that have only one possible outcome. They are of no use for gathering information, obviously, and the outcome of any of them could just be "It's done!"

But what was done? In general a measurement procedure, being a physical interaction, has an effect on the object or system measured, it changes the state, it is a *transformation.* For example, it might be a rotation of the object. So we can identify physical transformations with single-outcome measurement procedures. Call this family T^e, with the "e" for "empirical".

There are three roles for these transformations: with respect to measurements in general, to quantities, and to states.

The first is that of transformations induced on the world of measurements. For given any measurement procedure P we get a new measurement if we first apply transformation T and then P. And we can add to this that this induces also transformations on measurement set-ups, for a transformation combined with a measurement procedure transforms all the encompassed measurements 'in the same way', so to speak, to yield another measurement set-up.

Some subfamilies of the transformations, such as rotation through different degrees, form groups or semi-groups and will preserve structure, some will even be symmetries of the measurement landscape (of a specific theory), and some will systematically alter the structure of implication and orthogonality relations.

(b) Quantities

On the side of the measurements we can define a notion of empirical quantities Q^e. To define them we need to distinguish significant sets of measurements from gerrymandered sets. That distinction is also theory-laden. But there is a general condition to impose: the significant sets have to include all the elements of the Boolean algebras of sets of measurements that belong to a single measurement set-up if they include any of those elements.

Let us call these significant sets the *measurement propositions.*

Then an empirical quantity is a function that maps the Borel subsets of the real line into measurement propositions. Intuitively, if E is a Borel set and q one such function, then q(E) is the set of measurements whose outcome determine the value of q to lie in E.

We see here a second role played by the family of transformations T^e. For if **t** is an *appropriate* such a transformation then the function q', defined by the condition that for all Borel sets E, q'(E) = {**t**M: M in q(E)}, is another empirical quantity. The distinction signaled here by "appropriate" is also theory-laden; it is not determinable a priori.

An example would be the dynamics in quantum mechanics presented in the 'Heisenberg picture'. where the temporal transformation is not of the states but of the quantities.

(c) States

What is a state? It is something that has two roles: a state is something in which quantities have values and which determines the probabilities that certain values will be found if quantities are measured. In the most standard presentation of quantum mechanics these two roles are connected by the 'eigenstate-eigenvalue link': q has value r in state φ exactly if the probability that a measurement of q, when made in that state, would find value r with probability 1. That is controversial in matters of interpretation of the theoretical states and quantities, but as long as we focus just on the measurements, and not on the theory, there is no discernible leeway.

Thus we have a notion of the empirical states, S^e, which are to be distinguished from the theoretical states (even though so much of what went into the identification of measurements and measurement results depended on the theory).

We define such a state to be a function φ which assigns values ("probabilities") in the interval [0,1] to the measurement propositions, with the following constraints:

- restricted to the Boolean algebra of measurements in a single set-up, φ is a probability function
- φ respects effective equivalence and orthogonality

It is clear from these constraints that φ is assigning *probabilities conditional on measurement*. Any other properties such states have to have will come from structure that we cannot delineate a priori.

The third role that the empirical transformations can play is to induce a systematic (e.g. dynamic or temporal) change of state: if **t** is an appropriate such a transformation then the function φ', such that for all measurements M, φ'(M) = φ(**t**M) is another empirical state. Once again, the distinction signaled by "appropriate" is also theory-laden; it is not determinable a priori.

(d) The main matching: probabilities conditional on measurement

The probability that a value in E will be found upon a measurement belonging to set-up *S*, of quantity q, in state φ, equals φ(∪{A ∩ q(E): A in **B**}, where **B** is the Boolean algebra of events belonging to *S*.

That this should be placeable in a matching relation to what the theory says with respect to the theoretical quantities and states is, from a practical point of view, the main requirement of harmony between the theoretical landscape and the world of experimentation and measurement.[6]

2-3. Intensional logic as path toward articulation of subject matter

Thirdly, I want to highlight an approach to the relation between theory and its subject matter that does not relate directly to measurement, but may hold the promise of a larger theoretical framework for what I have mentioned so far.

Halvorson indicates from the beginning of the book that he will not be taking up modal logic, but that modality and intensionality are present in the wings, so to speak, throughout (see pages vii, 258). It may be possible, however, to put intensional logic to use in the area where we are at present looking for some clarification by means of formal methods.

Though the concepts and techniques also go back for a half century or more, what I want to point to is a recent paper by Jeremy Butterfield, "On Dualities and Equivalences Between Physical Theories". This paper includes, not coincidentally, disagreements with Halvorson on the concept of theory equivalence. What is of interest here is rather Butterfield's characterization of what is to be understood as the *subject matter* of a theory.

In the usual possible world semantics for modal logic, the semantic content of a sentence is represented as a set of possible worlds. Although this was developed for an analysis of language, it can be straightforwardly adapted, as Butterfield notes, to our above presentation of physical theories:

> Roughly speaking: systems are objects, i.e. references of singular terms; quantities are relations of objects to numbers (relative to a system of units); a state is a value-assignment to all the quantities that apply to a given object; and a possible world is the total histories of the states of all the objects in the world. (Butterfield, 15)

But then, what, in these terms, is the subject matter of a given such theory? Butterfield goes with David Lewis' analysis of the concept of subject matter. Lewis gives an informal example:

> We can say that two worlds are exactly alike with respect to a given subject matter. For instance two worlds are alike with respect to the 17th Century iff their 17th Centuries are exact intrinsic duplicates (or if neither one has a 17th Century). (Lewis 1988: 111-112; quoted Butterfield, 16)

---

[6] There was a great deal of literature to build on in this area. Both embedding and no-go theorems were proved, especially in the relation between partially ordered sets with a well-defined orthogonality relation and propositions as conceptualized in quantum logic.

Formally speaking, that means thinking of the subject matter as an equivalence relation on worlds, or equally well, as the partition of the worlds into the corresponding equivalence classes.

> The equivalence relation on worlds partitions the worlds into equivalence classes. The equivalence classes are propositions, ways things might possibly be. An equivalence class is a maximally specific way things might be with respect to the subject matter. So a third way to think of a subject matter, again general, is as the partition of equivalence classes. (ibid.)

So if a theory were to completely specify what the 17th century was like, it would in effect be the proposition that the actual world is one of those in a specific element of the partition, one of its 'cells'. Theories are not usually complete in this way, so they would only single out a family of such cells as where the actual world could be located.

Thus, what Butterfield, following Lewis, offers as explanation of "about", for a theory or proposition, is this:

> a proposition is about, i.e. entirely about, a subject-matter if its truth-value is determined by the facts about the subject-matter: that is, if the set of worlds at which it is true is a union of cells of the subject-matter. (Butterfield, 17)

In the adaptation to the talk about physical theories, we can imagine a series of theories T_i , as living on a common large state-space, but having their subject matter specified by equivalence relations E_i on trajectories in that state-space. The subject matter of given theory T could perhaps be the set of quantities Q, defined on that common state space, the equivalence relation being that of *having the same value* for all members of Q.

Butterfield expands on this in ways that refine the ways in which theories can be related to each other through their subject matters, as thus understood.

Note well that this is not an empirical but theoretical representation of the theory's subject matter. The equivalence relation in on the trajectories in the state space that satisfy the dynamics: each of these three are entirely internal to the theory. The theoretical representation of its subject matter is theory's own way of viewing what it relates to in the world.

To conclude this section, I don't know whether these specific suggestions can bear fruit, whether they can bring us to work with formal methods that will be of value to philosophy of science generally. But I am convinced that impact in our field will depend on whether there are going to be extensions of formal studies to the relation between theories and what they are about, as mediated by experimentation and measurement.

[3] Speculation: What Is Halvorson' Fundamental Project?

Occasionally Halvorson offers us tantalizing passages that point to deep philosophical problems. Tantalizing, for me anyway, for they suggest an overall view of science that lies behind Halvorson' work on formal methods and the structure of theories.[7]

[7] I am using Sartre's term "fundamental project" but, I hasten to add, mean to restrict it here purely to projects in philosophy.

To be selective, I will focus first on *representation* as it relates to scientific theories, and their relation to mathematical entities such as languages and models. Then secondly I will turn to the concept of a *meta-language*, that is, turn to *representation of language in language itself*. So both parts pertain to the general subject of representation, but it quickly appears that this term may be far from univocal.

Representation: stepping out of mathematics

Halvorson poses the main challenge clearly in his critique of our reception of Carnap's *Problematik*:

> In fact, philosophers have been so focused with epistemological questions that they seem to have forgotten the puzzle that Carnap faced, and that we still face today: how do the sciences use abstract mathematical structures to represent concrete empirical reality? (page 110)[8]

Mea culpa: I'll admit at once that around, say, ca. 1980, I was one of the guilty. I spoke quite blithely about isomorphism between phenomena and mathematical structures. That was wrong. For a certain kind of realism, the kind that speaks of 'carving nature at the joints', that would be just fine. But it is not at all fine for me, aspirant empiricist. It took me some twenty years to arrive at view that an empiricist can offers as response to Carnap's puzzle.

Taking Carnap's puzzle seriously, as Halvorson clearly exhorts us to do, the immediate task for the reader is now to find out how Halvorson himself would face this challenge.

On scientific representation

In some important passages Halvorson talks about scientific representation in just the way philosophers do who are not into formal analysis. He introduces the topic very straightforwardly in the section where he poses Carnap's puzzle:

> Here we use the phrase "empirical theory" or "scientific theory," to mean a theory that one intends to describe the physical world. You know many examples of such theories: Newtonian mechanics, Einstein's general theory of relativity, quantum mechanics, evolutionary biology, the phlogiston theory of combustion, etc. (page 107)

In some other places too, representation of things in the world by scientific theories or their models does not sound in any way problematic. For example, he discusses without qualms Gordon Belot's distinction between 'shifty' and 'shiftless' philosophers concerning whether two space-times (models of GTR) do or do not represent the same possible world (page 258).

What we have to look for is the connection between this intuitive grasp of science as representing things and the putatively related notions in Halvorson' formal conception of theories formulated in languages.

[8] This puzzle was certainly salient in Carnap's early surroundings; for example it was explicitly posed by Reichenbach (1920/1965: pp. 37-38), and Carnap refers to Reichenbach's point in *The Logical Structure of the World* (section 15).

On interpretation

In the Introduction that sounds rather straightforward as well, though emphasizing certain claims on behalf of semantics that need to be corrected:

> Throughout this book, we argue for a fundamental duality between logical syntax and semantics. To the extent that this duality holds, it is mistaken to think that semantic accounts of concepts are more intrinsic, or that they allow us to transcend the human reliance on representations, or that they provide a bridge to the "world" side of the mind-world divide.
>
> To the contrary, 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. (pp. 14-15, my italics)

Note well, in the first paragraph, his "the human reliance on representations " and the phrase I italicized in the second. This passage includes both a strongly negative reaction to what a philosopher might propose, and a positive endorsement of the fact (however puzzling it might rightly be): that "abstract mathematical structures [can] represent concrete empirical reality".

Halvorson' words have at least the implicature that there is a common, understood sense of *representation* in which bits of mathematics, like models, represent things that are not just other bits of mathematics.

But this becomes seriously qualified when Halvorson reflects on how the terms "interpretation" and "representation" appear to play different roles in formal studies, and the story becomes at once more intriguing and more disturbing.

To begin with interpretation, Halvorson writes that the word "interpretation" is highly suggestive, and might suggest that an interpretation of the set of sentences of a signatures endows the symbols with a meaning. But what is called an interpretation in this context is just a mathematical function, for example a mapping into the set $\{0,1\}$. The set of sentences **Sent**(Σ) and the set $\{0,1\}$

> are both mathematical objects; neither of them is more linguistic than the other, and neither of them is more "concrete" than the other. (page 22)

He might have added: the same holds for functions that map one into the other. It holds also for the signature itself, which is "a collection of items" that "have no independent meaning" (page 19).[9]

[9] And Hans goes further: "This point becomes even more subtle in predicate logic, where we might be tempted to think that we can interpret the quantifiers so that they range over all actually existing things. To the contrary, the domain of a predicate logic interpretation must be a *set,* and a set is something whose existence can be demonstrated by ZF set theory. Since the existence of the world is not a consequence of ZF set theory, it follows that the world is not a set. (Put slightly differently: a set is an abstract object, and the world is a concrete object. Therefore, the world is not a set.)" (page 22)

> We believe that it is a mistake to think that there is some other (mathematically precise) notion of interpretation where the targets are concrete (theory-independent) things. (p. 26)[10]

But is there no clue at all then to how the words in a theory's vocabulary "stand for things'? It seems not:

> There is also a tendency among philosophers to think of a signature as the vocabulary for an **uninterpreted language.** The idea here is that the elements of the signature are symbols that receive meaning by means of a semantic interpretation. Nonetheless, we should be careful with this kind of usage, which might suggest that formal languages lie on the "mind side" of the mind-world divide, and that an interpretation relates a mental object to an object in the world. In fact, formal languages, sentences, and theories are all *mathematical objects* - of precisely the same ontological kind as the models that interpret them. (p. 97)

In fact, logical semantics is just a special case of the *mathematical strategy of representation theory*. That is a strategy which relates different mathematical objects to each other, and it has nothing to do with talk of how scientific theories represent the phenomena:

> logical semantics is a particular version of a general mathematical strategy called "representation theory." There is a representation theory for groups, for rings, for C ∗ - algebras, etc., and the basic idea of all these representation theories is to study one category C of mathematical objects by studying the functors from C to some other mathematical category, say S. (page 165).)[11]

Enlarging on this in the context of remarks about the semantic view of theories, Halvorson laments the idea 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 bit of mathematics? (page 164/165)

"Representation" is therefore not univocal. In formal studies it denotes, as "interpretation" does, functional relations between mathematical entities or categories, while in the puzzle Carnap faced and we still face today it refers to a quite different relationship. Regrettably, the same word is used in two different roles.

---

[10] Hans may be taking issue here with some specific writings on modeling in the sciences. Thus in R. I. G. Hughes DDI account of modeling, the letter "I" stands for interpretation, which names the final step of relating features of a model to claims about the target system.

[11] Hans adds "In all such cases, there is no suggestion that a represented mathematical object is less linguistic than the original mathematical object. If anything, the represented mathematical object has superfluous structure that is not intrinsic to the original mathematical object." (ibid.)

Thus Carnap's puzzle, about how the science's use of mathematical structures represents concrete empirical reality, is not going to be addressed by any part of the formal analysis of theory structure that Halvorson is presenting in this book.

Somewhat later on Halvorson crosses the *t*s and dots the *i*s:

> It is of crucial importance that we do not think of a Σ structure *M* as representing the world. To say that the world is isomorphic to, or even partially isomorphic to, or even similar to, M, would be to fall into a profound confusion. (page 174)

Of course this is not all there is to it, even in the topics of Halvorson' concern. Attention to the relation between scientific theories and what they are about is not just dismissed. In fact, attention to that relation re-emerges recurrently in Halvorson' book, separately from this strong emphasis on the purely formal character of formal semantics.

So, can we tell from subsequent passages what Halvorson would see as a way to understand the relation between theory and world, that would not fall into a profound confusion?

Looking for a way out

There is definitely no direct answer to be found in the text, but there are two passages that point to a solution in pragmatics. And there is also a separate strand in the discussion that gives us, I think, an indirect access to how Halvorson may be thinking about this.

The two passages in question are first of all one I quoted above from the Introduction:

> it is mistaken to think that semantic accounts of concepts … allow us to transcend the human reliance on representations (page 14)

and the second appears at the very end:

> Just imagine two relativists – say, a cosmologist and a black hole theorist – sitting down to argue over whose model gives a more perspicuous representation of reality. They won't do that, because they are well aware that these models are accurate representations for certain purposes and not for others. (page 279)

These passages suggest – actually, do more than suggest -- that the relation of representation that goes from theories or models to things, phenomena, or processes in the world is not to be thought of as a binary relation standing on its own two legs. Instead, this is a relationship that depends on what is contributed by agents who can *use* and *choose*, who have *purposes* to be served. What is not addressed in the formal analysis, and is not meant to be addressed there, but is crucial to what scientific representation is, must be that *use* of the mathematical structures in the service of representation of reality.

Even if this subject is not further developed in this book, at least not explicitly, there is still, as I said, something else, that is possibly an indirect route to a view behind all this. It is to be found, I think, in Halvorson' various reflections on special case of the representation of language withi language, that is, on the object language/ meta-language distinction.

Representation of the world in language, and of language in language

In the section called "Putnam's Paradox" that Halvorson seems to be going for closure on Carnap's puzzle.

As I understand Halvorson's discussion, the main traditional answer to Putnam's (in)famous 'model-theoretic argument' is ruled out. By that I mean the answer based on a metaphysics of the sort typically signaled by the carnivore phrase "carve nature at the joints". Examples would be Stathis Psillos' form of scientific realism, or David Lewis' 'New work for universals'. Within that form of realism, it makes literal sense to speak of a functional relationship such as isomorphism between a mathematical structure and some part of the world, because that part has a 'privileged' structure itself. Here Halvorson not only dismisses that as a response but add his own telling critique of Lewis' position.[12]

What is really most interesting about this, though, is the surrounding text in which Halvorson reveals some of his own thought about language. In his example, Mette and Niels have very similar languages except that the denotations of terms in Mette's language are not the same as in Niels.

The first point is that Mette and Niels have no way to explain their difference in the way that we do, for they can refer to the denotata only by using their own terms:

> It is the metatheorist who can say: "Mette uses *c* to denote *a,"* and "Niels uses *c* to denote *b."* But how does the metatheorist's language get a grip on the world? How can he tell what Mette and Niels are denoting, and that they are different things? (p. 264, typo corrected)

This is followed by something that, as Halvorson says, might look like he is affirming Putnam's (apparent skeptical) conclusion:

> 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 these 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. All of us - Mette, Niels, Hilary, you, and I - are on the same level when it comes to language use. None of us has the metalinguistic point of view that would permit us to see a mismatch between language and world. (page 264/265)

But no, disaster does not follow from this admission. The initial explanation of this point we must read as Halvorson not speaking entirely in his own voice. Since it includes some things that Halvorson has earlier taken to be absurd, we must read it as still subordinate to assumptions of Putnam's argument, but nevertheless serving to show that those assumptions need not lead to Putnam's conclusion:

> Consider another scenario, where now I, rather than Putnam, get to choose the rules of the game. In other words, I have my own theory of which I believe the world W is a model. Then along comes Putnam and says that any consistent theory can be interpreted into the world W. But if my background theory T_o is not ZF, then I don't see W as a set, and Putnam's argument cannot even get started. In particular, I don't necessarily grant that there is an isomorphism $f : M \rightarrow W$ between a model M of T and this model W of my

[12] Hans argues that even if the premises about reference attractors and natural predicates were granted, there would still be too many options to block the unacceptable consequences of Putnam's model-theoretic argument.

> theory T_o. For one, what would I even mean by the word "isomorphism"? I, the user of the theory T_o, know about isomorphisms between models of my theory. However, M is a model of a different theory T, written in a different signature Σ, and so there may be no standard of comparison between models of T and models of T_o. (p. 265)

So, no disaster follows when the idea of someone's 'own language', 'home language', is taken seriously. For what is crucial in this passage, I think, is not its main point but the crucial role of the assertion that I have *my own language*, and my own theory formulated in that language. This appears more straightforwardly below:

> In real life, this is the sort of criterion we do actually employ. If I hear someone else speaking, I judge that what they are saying "could be true" if I can reconstrue what they are saying in **my language**.
>
> […] As Otto Neurath pointed out, and as Quine liked to repeat, we cannot start the search for knowledge from scratch. Each of us already has a theory, or theories. (page 266, my boldface)
>
> […] Lewis' background theory T_1 has little to recommend it, even if we are inclined to accept that there are "natural properties." (And anyone who uses first-order logic implicitly does accept the existence of natural properties - they are precisely the properties that are definable in **her language**.) (page 268, my boldface)

I am not going to object to Halvorson's move here, introducing the view that we each have a language that is our own. On the contrary, the phrase "her language", in that last passage, I am more than willing to take on as understood.[13]

Two problems to be addressed

Still there are two possible doubts as to whether Halvorson has achieved closure by this move. The first is that this assignment of a language as *her* language or *my* language is subject to a significant problem of identification. The second problem is that, after assuming the first problem to be solved or settled, we need to face the question of what the relation could be between *that* language and the formal languages studied in the book.

What is her/my language?

There is a wonderfully challenging passage in one of David Lewis' papers about what *his* language is. He has just asserted that we can quantify over our counterparts in other worlds:

> I can. Some say they can't. They say their understanding is limited to what can be expressed by modalities and world-restricted quantifiers. I have no help to offer these unfortunates, since it is known that the expressive power of a language that quantifies across worlds outruns that of the sort of language they understand. (Lewis 1979, 517 ftn. 2)

We do understand very well, when reading this, that Lewis is at the same time inviting us to learn his language, to adopt it as our own, and thus to reap its intellectual bounties.

---

[13] Quine made this move at one point, introducing as language with special status our *home language*, but in his case the upshot was distinctly unsatisfactory (cf. van Fraassen 1990)

But here is the open question: by what criterion would I have actually done so?

I am sure I could learn his language very well, and I can even imagine that I could weekly teach two seminars, in my own (prior?) voice on Mondays and in Lewis's voice (my posterior voice?) on Wednesdays. I could do this so well that no one listening, no fly on the wall, could find any clue as to which day had me speaking in my own (current) voice. We could arrange a sort of Turing test, and demonstrate that.

So, if that is so, there is no behavioral criterion. What is it then, that makes one language my language rather than the other language that I speak as well as a native?

I am not saying that this question is unanswerable, but will say that the answer will look very different within diverse philosophical positions. From the point of view of naturalism it might be suggested that what is my language is determined by my dispositions for playing language games of various sorts in different contexts. But basing the distinction in my dispositions for different sorts of behavior will, I think, run headlong into the sort of problem signaled by Kripke's discussion of the chess players, one of whom had the unrealized disposition to go berserk if the other castled. Hans does not seem to favor naturalism as philosophical position, and is quite scathing about attempts to improve it by recourse to functionalism, supervenience, and Ramsey sentences. But if the question of what is my language does not admit of behavioral, dispositional, or functional criteria, we will need something more.

Could the question be dismissed? I don't think, at least not here. For it seems that the status of home language is of importance to Halvorson. One example appears when he discusses "those hard-core realists – like David Lewis or Ted Sider – who pin their theoretical hopes on natural properties and reference magnetism" (p. 257). He comments that

> adopting a first-order signature Σ already goes some way to picking out natural properties. When we specify a Σ-structure M , we get a natural property M (ϕ) for each formula ϕ of Σ. It's not clear then why a theorist who has adopted a first-order signature Σ would need to additionally specify a notion of natural properties. (ibid.)

The word "adopted" signals clearly that the signature in question would have a special status for the theorist. In this remark Halvorson must at least be pointing to some significance of that selection of predicates in order for this remark to be apt.

We should also read this passage with the sort of bilingualism in mind that I introduced above. If I am equally adept in two languages, with different signatures, then the predicates in those languages will 'pick out natural properties', to use Halvorson' phrase, in different ways. Cross-cutting a carcass would not be cutting it at the joints – if I may elaborate that bloody metaphor once again – so it would be important that just one, only one, of the two languages was *adopted* by the speaker *as their own* in the sense in which Halvorson speaks here of adoption.

That the status of one's own language is of importance is reinforced in the next section of this chapter. Here we have the example of Berit's language, in which she formulates the theory that there are just two things in the world, one and only one of which is P. Two models could have the same domain, but differ precisely in which element falls in the extension of P. Do these represent the same possibility or two different possibilities? Halvorson shows up multiple confusions that could beset this question. But what is important for us at this juncture is just what he explains here about representation:

> The confusion here is somewhat similar to Skølem's paradox (about the existence of uncountable sets in models of ZF set theory), where we run into trouble if we don't distinguish between claims made in the object language and claims made in the metalanguage. In the present case, one might be tempted to think of the theory T as saying things such as
>
> In model M, a is a P.
>
> Of course, T says no such thing, since it doesn't have names for models or for elements in models.
>
> The other problem here is in the way that we've set up the problem – by speaking as if the representation relation holds between M (or N) and the world. To the contrary, *the representation relation holds between Berit's language and the world*, and we (the metatheorists) are representing Berit's theorizing using our own little toy theory (which presumably includes some fragment of set theory, because that's a convenient way to talk about collections of formulas, etc.). Berit herself doesn't claim that M (or N) represents the world – rather, the metatheorist claims that M and N represent ways that Berit's language could represent the world. (page 261, my italics)

Note well that this passage is not about representation in the sense of representation theory in mathematics, it is not about functional relations between mathematical entities. Rather, it is about "the representation relation [that] holds between Berit's language and the world".

The language that is her language is itself something real in the world, something real that serves to represent something real. Given that, it makes sense to ask what, formulated in this language, is true and what is not true – I mean, true simpliciter, as opposed to true in a model. That is then not a question of a relation between the syntax and the models, not a question about a mapping into the 2-element Boolean algebra, in fact, it is not a question within mathematics at all.

If Halvorson is to have warrant for the introduction of the notion of *her own* language, of a *home language*, we need to have his full answer to what he introduced as Carnap's puzzle.

Let's be clear: it is not incumbent on Halvorson, within the project of this book, to provide this answer, or to settle all the fundamental questions that his project brings to light. *Bien sûr que non!*

Relation to formal languages

The one conclusion I do draw *pro tem* is that this distinction, between someone's own language and languages in general, is a crucial part of Halvorson' response to Putnam's paradox, which is where his general views on logic and language appear most saliently

Suppose, to advance our narrative, that we can speak freely, without equivocation or indeterminacy, about what is *my language*. Then that language is something real and concrete in the world. Never mind how it is identified within different philosophical views. Perhaps it is constituted by a plethora of dispositions and propensities, or more concretely by a neural state over time, or less metaphysically and less naturalistically, exists as the intensional correlate of my philosophical orientation. Fine, it is something in the world, my language is as much in the world as I am in the world.

But then what is the relation of any formal language, such as a first-order language of the sort studied in Halvorson' book, to my language?

The latter is a mathematical entity: the signature Σ is a mathematical entity, so is Sent(Σ), so is any theory T in that signature, so is MOD(T), and so forth. So to ask how the signature and its ancillary constructs represent my language is just: *how do we in formal studies represent the concrete empirical reality of my language by a mathematical structure?*[14]

So it seems we have come right back to the same question that Halvorson designated as Carnap's puzzle. No, not quite: we have come to it in a more special form, the question as it arises for a special case. The very question that would seem to be the one always hanging in the air since Tarski: what is the metalanguage / object-language relation?

What is meta-language, or what could it be?

We must certainly start with this quotation where Halvorson confronts us with an implicit challenge:

> The distinction between object language and metalanguage is one of the most interesting ideas in twentieth century logic and philosophy – and it remains one of the least well understood. (page 262)

Halvorson does not follow this up with a crisp, concise take on the proper or correct understanding. Almost immediately afterward, however, there is a revealing note on Quine:

> Quine seems to reject the idea that there is an important difference of status between object and metalanguage. He seems to propose, instead, that the ascent to metalanguage should be seen as an extension of one's object language – and so assertions in the metalanguage have exactly the same force as assertions in the object language. (page 263)[15]

I can only read as in a very disapproving voice. So if we are to speculate about what is Halvorson' conception of the object language / metalanguage distinction, we should begin with what is here rejected: Quine's take on Tarski's hierarchy of metalanguages upon metalanguages, language without end.

There was always a puzzle, even a paradox, about Tarski's conception. If that hierarchy exists it is complete: it is a completed, actual infinity that encompasses all that language could be. At any level in that hierarchy, discourse can only be about what is either present or talked about in some lower levels. For that reason, this hierarchy, taken as a whole, cannot be described, or

---

[14] This question does not reduce to "*Do the symbols in* Σ *represent my words?*". Not in the sense of what is called an interpretation in this book, or what is called representation theory in mathematics. It is also no help to suggest that the members of Σ are in fact my own words. They could be inscriptions of those words, but that would not make them words in my home language: recall Putnam's example of a squiggle traced by an ant in the sand, in a galaxy far, far away and long, long ago, which has the exact form of "Coca Cola". It is clearly not our meaningful phrase denoting a soft drink.

[15] I have not found it easy to find any precise view on this in Quine's writings. Section 56 "Semantic ascent" in *Word and Object* takes us, on a theoretical level, no further than section 4 "Use versus mention" in his *Mathematical Logic*.

indeed even talked about, in any discourse that could have a place within it. So our present discourse does not fit in it: so, if the hierarchy exists it is incomplete.

Quine's reaction to Tarski: semantic ascent

Quine's reaction, if I understood it correctly, was not to assert the reality of this hierarchy but rather to assert that we are always, at any time, in a position to practice *semantic ascent*. That is, we can always augment our current language, or I should say, *our language so far*, with linguistic devices to speak about that language. This act would thus extend *that language so far* into a language that includes as part a metalanguage, for which *that language so far* is the object language.

But how does this work, how is it possible for us, speaking the language that is our language so far, to practice semantic ascent? What must the *language-so-far* be like, or have in it, to make it possible for it to surpass itself in this way?

It really cannot be anything like the languages we study in formal logic or semantics. We must especially take to heart Halvorson' point about the essential riches for expressive power:

> The metalanguage describes the world in finer-grained language than the object language. (page 262)

Although I can only speculate on how Halvorson sees this, we can look to the important passage about Mette's and Niels' language.[16]

Given what Mette and Niels cannot do, how could they possibly lift themselves by their bootstraps to get into the metatheorist's bootstraps?[17]

What could that transition be like, if we were to think of semantic ascent as the change-over from a prior language to a posterior language with those greater resources, and thought of these languages as just of the sort studied in formal semantics? Would that not be a truly Nietzschean *Überwindung*? It would be one of those small miracles which seem so often, in philosophy, conveniently ready to save a theory.

We can take this aporia in one of two ways. We can take it as a direct critique of Quine's attempt at a re-conception of Tarski's hierarchy, as being about what we can do. Or alternatively we can grant that we can practice semantic ascent or something very much like it – but then deny that a person's language could possibly consist only in what can be formulated with the vocabulary-so-far by the rules of grammar-so-far. (Well, maybe both!)

---

[16] "To make these ideas more concrete, let's consider an example: Let $\Sigma = \{c,d\}$, where *c* and *d* are constant symbols. Let *T* be the theory in Σ that says $c \neq d$, and $\forall x((x = c) \mathbf{v} (x = d))$…. Of course, there is only one model of *T* up to isomorphism. And yet, a skeptical worry arises! Imagine two people, Mette and Niels, both of whom accept *T,* and both of whom think that the world is the set $\{a,b\}$. And yet, Mette says that *e* denotes *a,* whereas Niels says that *e* denotes *b.* Do Mette and Niels disagree? The answer is yes and no.
We have already misdescribed the situation. Mette cannot say that *"c* denotes *a,"* because *a* is not a name in her language. Similarly, Niels cannot say that *"c* denotes *b."* It is the metatheorist who can say: "Mette uses *c* to denote *a,"* and "Niels uses *d* to denote *b.*" "(page 264)

[17] We should also add here the passage on page 265: "All of us – Mette, Niels, Hilary, you, and I – are on the same level when it comes to language use. None of us has the metalinguistic point of view that would permit us to see a mismatch between language and world**."**

Most of all, I hope to have Halvorson' own reaction to this. But again it would be just too negative a commentary to leave the point here without any suggestion about how there could be a better understanding, that might or might not be in accord with Halvorson's.

Curry: object languages, metalanguages, U-language

What I want to introduce here as counterpoint is the arch rival to Tarski's way of thinking about languages and their metalanguages, or Quine's for that matter. That is the conception of Haskell B. Curry ((cf. Curry 1950, pp. 348–349; 1963, Ch. 2 Sects. A1, A2).

In Curry's view there is only one language, within which we are doing all that we can do in language, the language we live in, the language he calls our *language in use*. It is within this language that we do logic and mathematics, and that activity includes the creation of object languages, and of metalanguages for those object languages.[18]

To distinguish this from constructed or abstracted languages Curry calls this the U-language. Note well that the word "language" has here clearly become equivocal.

It is possible, according to Curry, to isolate *some part* of our language in use, a part that is already actual, to recast it in some canonical form, and to construct an object language that represents that part, in some way, for example by a first-order language of the sort studied in Halvorson'.

But there is no way in which we could construct or have an adequate such representation of our language in use. The term U-language does not refer to an object of the sort to which constructed languages, or languages abstracted from actual discourse, belong. To purport to have such a representation of our language in use, or to assume that there is one, is to land in very familiar but devastating semantic paradoxes.

Questions about *what denotes what*, for example, if asked about texts in our language in use, must be being asked in our language in use itself, and hence can only have trivial answers, 'stuttering answers' as McDowell would say: "snow" is a word for "snow" and "snow is white" is true iff snow is white.

Nor is our language in use a metalanguage for anything. To represent it as a metalanguage for an actual language in use would just mis-describe the situation. It would misdescribe the activity of isolating some part of our discourse-so-far and representing it by something that logicians call a language. However, we can construct a meta-language, in which to formulate assertions about some constructed object languages, whether actual or envisaged. And we could construct a meta-meta-language, and so forth.

Our language-in-use is not static. We can always make our formulations more precise, eliminate ambiguities, add quotation devices, enrich the vocabulary …. Curry's analogy is to precision tools and measuring instruments. Some higher level is achievable, for any so far achieved level of precision. What we cannot have at the same time is tools more precise than any tools we have, to determine the limits of precision of what we have.

[18] It is tempting to identifyCurry was a formalist, but with a view quite close to intuitionism in some respects. Brouwer and Curry both regard construction as the prime mathematical activity. But while Brouwer thinks of that as mental activity with language only there for (always essentially inadequate) communication of the construction to others, Curry thinks of it as linguistic activity in the broad sense of symbol creation and manipulation.

That seems right, but the reference to tools seems rather too shallow as an analogy. I would like to offer a different one. The analogy I propose is between logic, at two levels, and geometry at two historical stages. In Curry's formalist view, both geometry and logic are creative enterprises conducted within the language-in-use that focus on form.

Perspectival geometry, "Perspectiva", developed in the Middle Ages, was developed as a continuation from Euclid's *Optics*. Its topic was the construction of figures, *in proper perspectival form*, as pictorial representations of real worldly things. This discipline can be presented as pure geometry, but it was also the practical guide for 'realistic' painting in the Renaissance, starting in the 15th century.

The stage beyond that practice was projective geometry, created by Desargues in the 17th century. This introduced the study of *relations between* perspectival representations, that is, between the geometric objects that are constructed and studied in perspectival geometry. Eventually this turned into the study of transformations that relate such objects to each other, the projective transformation group.

The analogical two-stage pattern I would then see in the relation between elementary logic and meta-theory.

In elementary logic (whether sentential, quantificational, or modal) we teach students to construct verbal representations, *in a certain form*, *in logical form,* of worldly situations, and how to draw out implications of those representations.

Then, analogously to the turn to projective geometry, we turn in metatheory to a study of the *relations between* the constructions – theories in 'elementary' languages – that could function as direct verbal representations of real worldly things, events, and processes.

The analogy is thus between the levels of logical study pertaining, respectively, to object languages and the metalanguage(s) for them, on the one side, and on the other side, the levels of geometric study in, respectively, perspectival geometry and projective geometry.[19] It is not at all a perfect analogy. I have to qualify it at least because in actual elementary logic courses the study of language tends to turn 'meta' much more quickly than I imagine the teaching of Perspectiva does in a drawing class.

Nevertheless, this way of looking at it helps to highlight Halvorson' insistence that it is *not in metatheory* that we study the relation of language to the world. The object languages, or theories therein, are the putative vehicles for the representation of worldly situations, just as are perspectival figures. And metatheory is the study, not of relations to the world, but of relations between those representations, just as is projective geometry the study of relations between geometric representations.

CODA

Curry was famous for audacious efforts in combinatory logic to create ever richer object languages that skirted outright contradiction by demanding what he called 'combinatorial completeness'. Equally audacious was his understanding of our language in use, which is clearly

[19] Cf. van Fraassen 2008, 73-75; Kvasz 2008, 114-124.

not constituted by a vocabulary-cum-grammar, with or without an associated category of models, but something like the air within we live, and move, and have our being.

I do not know whether we can rest at peace within this vision bequeathed to us by Curry. But we need an answer to the questions he addressed, the very questions to which Halvorson pointed in his critical remarks.

BIBLIOGRAPHY

Butterfield, J. (Forthcoming) "On Dualities and Equivalences Between Physical Theories', in *Philosophy Beyond Spacetime*, ed. B. Le Bihan, N. Huggett and C. Wuthrich . Preprint http://philsci-archive.pitt.edu/14736/; https://arxiv.org/abs/1806.01505.

Curry, H. B. (1950). Language, metalanguage, and formal system. *The Philosophical Review* 59: 346–353.

Curry, H. B. (1963). *Foundations of mathematical logic*. New York: McGraw-Hill.

Emch, G. and J. M. Jauch (1965) « Structures Logiques et Mathématiques en Physique Quantique». *Dialectica* 19 : 259-279.

Foulis, D. J. and C. H. Randall (1974) "Empirical Logic and Quantum Mechanics". *Synthese* 29**:** 81–111. Reprinted in P. Suppes (ed.), *Logic and Probability in Quantum Mechanics.* Dordrecht: Reidel, 1976.

Foulis, D. J. and C. H. Randall, C. H. (1978) "Manuals, morphisms and quantum mechanics", pp. 105-126, in *Mathematical Foundations Of Quantum Theory*. Academic Press, NY.

Glymour, C. (1975) "Relevant Evidence". *Journal of Philosophy* 72: 403-426.

Glymour, C. (1980) *Theory and Evidence.* Princeton: Princeton University Press.

Halvorson, H. (2019) *The Logic in Philosophy of Science.* Cambridge: Cambridge University Press.

Kvasz, L. (2008) *Patterns of Change: Linguistic Innovations in the Development of Classical Mathematics*. Basel: Birkhauer Verlag.

Lewis, D. K. (1979) "Attitudes de dicto and de se". *The Philosophical Review* 88: 513-543.

Lewis, D. K. (1988) "Relevant implication". *Theoria* 54, 161-174. Reprinted in his *Papers in Philosophical Logic*, Cambridge University Press 1998; page references to reprint.

van Fraassen, B. (1990) Review of W. V. O. Quine, *Pursuit of Truth*. *Times Literary Supplement*, Issue 4558, August 10-16, 1990, page 853.

van Fraassen, B. (2008) *Scientific Representation: Paradoxes of Perspective.* Oxford: Oxford University Press.

van Fraassen, B. (2012) . "Modeling and measurement: the criterion of empirical grounding", *Philosophy of Science* 79: 773-784.