Structuralism with an empiricist face?

Bas Van Fraassen's *Scientific Representation: Paradoxes of Perspective* is a rich, masterful study of a wide range of issues arising from the manifold ways in which scientists represent nature. Scientific representations or "models" of the world can themselves be physical objects, as in the case of "scale models", but more characteristically they are mathematical models. Under what conditions can such an abstract entity represent concrete objects, events , or processes? What does it take for a mathematical model to succeed in fitting the world? This question runs as a motif through Van Fraassen's book.

Consider, for instance, the equation (derived from Fourier's law of heat conduction) specifying how an insulated metal bar with no internal heat sources, having some initial internal temperature distribution, arrives over time at a uniform equilibrium temperature. Each model of the theory consists of a mathematical function TEMP(x,y,z,t) that solves the heat equation. Each model represents one way that such a bar's internal temperature could vary as a function of spatial location in the bar and time elapsed. How does the model succeed in capturing the temperature behavior of metal bars? The model does not represent a bar by having the same temperature as the bar. The model does not have any temperature; it is an abstract object. The model's success must involve various features of the model standing for various features of the bar. For example, the t coordinate stands for the elapsed time, whereas "TEMP" stands for the bar's temperature at a given location after a certain period of time has passed. Various relationships among the model's properties (such as TEMP's approaching uniformity as t increases) must be the same as various relationships among a bar's properties (such as its temperature's approaching uniformity as time passes). The model and the world must share some structure.

How can this condition for a model's success be cashed out? Apparently, if we lack an "interpretation" of the model connecting it to the world -- identifying which features of the model are supposed to represent which features of the world – then the model tells us nothing about the world. (I just suggested one such interpretation: "t" stands for the elapsed time, "TEMP" stands for a metal bar's temperature, and so forth.) The model itself, absent an interpretation, does not specify whether "TEMP(x,y,z,t)" stands for a bar's temperature or the reciprocal of its temperature – or its electric charge density, or its mass density, or some quantity not pertaining to a metal bar at all. The model taken alone fails to specify what it is supposed to be a model of. How an interpretation can latch the model onto the world is the "problem of coordination", which Van Fraassen traces through the work of philosophers including Ernst Mach, Moritz Schlick, Rudolf Carnap, and Hans Reichenbach.

An interpretation specifies how a model relates to the world, but how does an interpretation relate to the world? Here we run the risk of reproducing exactly the problem that we were trying to solve. To specify an interpretation, we must specify various features of the world (together with which features of the model represent them). But then we need to represent those features of the world – and thus we apparently again need an interpretation.

One way to try to avoid this problem is to return to the thought that a model is successful when it and the world share some *structure*. Rather than a scientist's having to have an interpretation that picks out which features of the world are purportedly represented by which features of a model, the scientist generates empirical predictions from the theory as "Ramsified". So formulated, the theory does not have terms that purport to represent various unobservable features of the world. Instead, the theory consists of "Ramsey sentences" to the effect that there exist various unspecified properties possessing various structural properties and standing in various structural relations to one another and to observable features. The theory represents the world as having some properties filling these various structural roles, but the theory does not purport to specify what those role-filling properties are in themselves. According to such "structuralism" about science, all we can know (or, at least, all that science cares about knowing) about these properties is that a certain structure is thus instantiated. Van Fraassen traces this view in thinkers including Heinrich Hertz, Pierre Duhem, Henri Poincaré, and Bertrand Russell. It is easy to sympathize with structuralism, since it is easy to find oneself thinking (to paraphrase Lord Kelvin) that from Maxwell's equations, we understand everything about electrodynamics – including when charges flow and what accompanies their flow -- except what it *is* that is then flowing.

However, in 1928, the mathematician M.H.A. Newman offered a now-classic objection to structuralism (or "structural realism", as Grover Maxwell later termed it). Newman showed that any domain of entities, as long as it is large enough, exhibits any structure whatever. Hence, if a theory's observational consequences are true, then (as long as the world contains sufficiently many entities) there is certain to exist some interpretation of the theory under which the world makes its Ramsey sentence formulation true. That the world exhibits a certain structure, specified by the theory, was supposed to be a substantial claim about the world. (As Newman said, the question of whether or not matter is atomic "is a real question to be answered by consideration of the evidence.") On structuralism, however, it amounts only to the claim that there are enough entities to exhibit that structure, since as long as there are enough, it is guaranteed that they will somehow exhibit it. Structuralism portrays a theory's truth as coming far too cheaply.

Let me put this point in another way, so as to better exhibit how devastating it is. Structural realism is supposed to be a variety of scientific realism, since it holds that in accepting a scientific theory, scientists believe something over and above its empirical adequacy – namely, that the world (perhaps in its unobservable features) possesses a certain structure. One supposed advantage of this view (emphasized especially by John Worrall) is that it can embrace the argument that a scientific theory's success in predicting novel sorts of phenomena is probably "no miracle" (that is, no mere coincidence, as it would be if the theory were utterly mistaken about the world) since the theory's success is likely a result of its having correctly described some feature of the world's structure. However, Newman has shown that for a theory to represent the world's structure amounts to nothing more than the world's having enough role players to fill all of the places in the structure that the theory posits. Accordingly, structural "realism" fails to do justice to the idea that it would probably be "no miracle" that a theory makes accurate novel empirical predictions.

Structural realism tried to avoid needing an interpretation to fix the way a theory represents the world; it construed theories as representing only the world's structure and thereby contented itself with the mere existence of *some* interpretation under which the theory represents the world. But as Newman showed, this form of structuralism is undone by its liberality in allowing the existence of any interpretation to suffice. We could try to steer a middle course between needing an interpretation and allowing the existence of any interpretation to suffice by imposing some non-structural but relatively liberal constraints on allowable interpretations – that is, on the respects in which the world has to realize the posited structure in order for the model to count as accurately representing the world. For instance, David Lewis suggested roughly that a theory purports to represent the structure exhibited by the world's *natural* properties and relations. That the world includes enough entities to fill out the structure posited by a theory fails to guarantee that the structure is realized in a unique way by *natural* properties and relations. It is thus no longer too easy for the model to count as an accurate representation of the world.

However, Van Fraassen rejects Lewis's gambit on the grounds that Lewis's constraint on allowable interpretations is empty unless we have some grip on what would make a property "natural" apart from its ability to serve in an allowable interpretation. Of course, I agree that Lewis's constraint presupposes that naturalness has some independent content. Indeed, Lewis puts the notion of a "natural" property to work in other parts of his metaphysics, such as in his account of natural law. Nevertheless, Lewis is prepared to countenance naturalness as a metaphysical primitive. Although there are some motivations for construing science as aiming to identify the world's natural properties and relations, Van Fraassen sees the formulation of comprehensive, empirically adequate theories as the unique goal of science.

Van Fraassen offers a different way around the "problem of coordination." As Newman showed, it is easy for there to exist an interpretation of some model under which it is true -- and it should be easy, since as long as we do not know which interpretation will do, the model tells us nothing about the world. For the model to be informative, we must give it an interpretation, and to specify an interpretation, we need to pick out features of the world (and then specify the features of the model that purport to represent them). Van Fraassen accepts that to pick out features of the world requires using a language that is already "coordinated". But in making observation reports, compiling them into what Van Fraassen calls "data models", and smoothing and massaging those models in turn (under the influence of some theory that we already accept) into "surface models", we are using a framework that we have presupposed to be coordinated. In other words, we regard ourselves as already having the means of picking out features of the world. What a theory confronts directly is itself a representation that we consider successful; we already have ways of picking out those representations (since we did so in constructing the surface model) and we have already recognized them as accurately representing the world (namely, in making our observation reports and turning them into our data models and surface models).

That we can interpret a model only by using representations whose relation to the world is not in question might seem to launch an infinite regress. But as Van Fraassen says, “we already have a language that we live in.” Van Fraassen characterizes his solution as “Wittgensteinian”, and his use of the term “interpretation” emphasizes this kinship. An interpretation of a rule is just another rule, so doesn’t it need an interpretation itself, just as an interpretation of a model can latch it onto the world only if its components are already latched onto the world? As Wittgenstein famously replied (in *Philosophical Investigations* §201), there is a way of grasping a rule that is *not* an interpretation. In making our observations and measurements, we have stabilized a practice that we take as yielding representations of some features of the world. In using them for that purpose, we do not need to specify an interpretation of them. Of course, we may at some other time find ourselves with reason to doubt whether they do indeed represent the world accurately, and we will then need evidence to support some interpretation of them. But the evidence to which we then appeal we will thereby use without interpretation. As Wilfrid Sellars put the point (in “Empiricism and the Philosophy of Mind” §38), “empirical knowledge, like its sophisticated extension, science, is rational, not because it has a *foundation* but because it is a self-correcting enterprise which can put *any* claim in jeopardy, though not *all* at once.”

In short, meaningful questions about a model’s capacity to represent the world (such as “Does ‘TEMP(x,y,z,t)’ in this model capture the behavior of that bar’s temperature?”) presuppose meaningful ways of talking about the world (e.g., talking about that bar’s temperature). Van Fraassen concludes that “structuralism finds its proper articulation only in an empiricist setting”. But I don’t see his response to the problem of coordination as available solely to empiricists in some narrow sense. The representations whose status is not being called into question could just as well concern unobservables, it seems to me. Scientific realists like Sellars are keen to emphasize the presuppositions about the world that lie behind making observations, data models, and surface models – especially when the theory used to massage data into surface models is justified only in the context of a certain conception of the unobservable world.

It seems to me that structuralist ideas can make important contributions – for instance, toward capturing the way science uses structural analogies among the fundamental laws governing physically dissimilar processes to explain why the phenomenological regularities in those processes exhibit various similarities that would otherwise have to be “miraculous” coincidences. Take these two phenomenological regularities:

> Consider a cylinder of length *L,* radius *a*, generating heat that keeps it at constant temperature T_1, surrounded by a uniform layer of thermal insulation, thickness (*b-a*), the outside of which is kept at temperature T_2. We find experimentally that in all such cases, the rate at which heat is

> generated inside the cylinder, thence to pass through the cylinder's surface, is proportional to $L(T_1 - T_2) / \ln(b/a)$.
>
> Consider a cylinder of electrically conductive material of length L, radius a, held at a constant voltage (electrical potential) V_1, surrounded by a uniform layer of electrical insulation, thickness (b-a), the outside of which is kept at voltage V_2. We find experimentally that in all such cases, the charge on the cylinder is proportional to $L\ (V_1 - V_2) / \ln(b/a)$.

Is it a coincidence that these two results are analogous? Physically, yes: after all, the electric field is not heat. But mathematically, it is *no* coincidence. The more fundamental laws of electrostatics and thermodynamics have precisely the same structure. (For instance, heat flow plays the same role as the electric field, temperature as electric potential, and heat generated as electric charge.) That is why there is a single mathematical proof deriving both of these phenomenological results from their respective more fundamental laws. It is "no miracle" that these two phenomenological laws take the same form – not because heat and electric charge share key intrinsic properties, but because they are structurally alike.