Reading 'Writing the Book of the World'

Cian Dorr

Analytic philosophy is suspicious of jargon words unless introduced by explicit definitions or for purposes of disambiguation. But this healthy suspicion must not be allowed to degenerate into a knee-jerk refusal to admit any conceptual innovations. The heart of *Writing the Book of the World* is an extended plea for the intelligibility, and importance, of a certain technical use of 'structural', a close cousin of Lewis's technical use of 'natural'. In this central aim, the book is in my view almost entirely successful. Setting aside certain exotic constructions involving '$\mathcal{S}$' (the formal counterpart of 'structural') which even Sider recognises as straining intelligibility 'to the breaking point' (p.257), I am convinced that 'structural' is not only intelligible, but a fruitful addition to the philosopher's idiolect, which allows us to raise questions that are interesting both intrinsically and for their bearing on other topics.

Does accepting this make me what Sider would call a 'realist about structure'? I am not sure: there are several ideas which seem central to Sider's vision which I do not accept. Let me mention some of them, to illustrate that 'realism about structure' need not be a package deal:

- The idea that people with highly unnatural concepts are 'making a mistake' (p.2). (As I see it, their only problem is that they are missing out on some interesting, important truths.)
- The idea that representations—or a special subclass of representations, the 'fundamental theories'—must use structural concepts if they are to be 'fully successful' (p.vii). (All that is going on, I think, is that people who propound theories in a certain important tone of voice sometimes *implicate* that their undefined words have structural meanings; but when the implicature is false, it is a failing in the theorist, not the theory.)

- The claim that 'structural' is *precise,* so that the question 'Is F structural?' is never vague unless F is. (As Sider acknowledges, there are questions, such as whether disjunction or conjunction is structural (§10.2), where it would be appealing to invoke vagueness as a reason for refusing to answer.)[1]

However, these points of disagreement seem minor compared to the point on which Sider and I agree: that *What is structural?* is a good and important question.

In the rest of my allotted space, I want to address Sider's Quine-inspired proposal about the right methodology for answering this question:

> The familiar Quinean thought is that we search for the best—simplest, etc.— theory that explains our evidence. My addition… is that this search is ideological as well as doctrinal; we search simultaneously for a set of concepts and a theory stated in terms of those concepts. We solve for the best and most explanatory pair $\langle I, T_I \rangle$ of ideology I and theory T_I in terms of that ideology.

Sider attempts to put this methodology into practice at many points in the book: when he argues that certain concepts are or are not structural (e.g. that no modal concepts are structural, or that some quantifiers are structural), he does so on the basis of a comparison of goodness/explanatoriness between theories which use those concepts and theories which do not.

Arguments of this form will be hard to assess until we know more about the relevant notion of goodness, and on this point Sider doesn't provide much guidance. My worry is that it is quite hard to make sense of the notion in such a way as to vindicate even ob-

---

[1] Sider rules this out on the grounds that it conflicts with his thesis (7.12) that 'no special-purpose vocabulary that is distinctive of indeterminacy carves at the joints', and 'fundamental languages obey classical logic'. I see no conflict here. However, vagueness in 'structural' would be hard to combine with Sider's thesis (7.13) that *structural* is structural. For further discussion see §4 of Cian Dorr and John Hawthorne, 'Naturalness' (forthcoming in *Oxford Studies in Metaphysics* vol. 8, ed. Karen Bennett and Dean Zimmerman).

viously justified beliefs about structure. For example, here are three pairs that we clearly need to find a way to count as bad:

(i) Suppose T is some state-of-the-art physical theory expressed in terms of some concepts P one of which is *mass*—or to be precise, the two-place predicate 'the rest mass of x in Planck units is n'. Let me introduce the concept *mass** by stipulating that whenever an object's mass is n, its mass* is $^{17}\sqrt{n}$. Let P_1 and T_1 be the results of replacing 'mass' in P and T with 'mass*'. (Note that T_1 is inconsistent with T: e.g. if T says (truly) that the ratio of the mass of the proton to that of the electron is 1836, T_1 will say (falsely) that the ratio of the mass* of the proton to the mass* of the electron is 1836.)

(ii) Let 'Q' abbreviate some true sentence not entailed by T—e.g. 'the moon and the sun look approximately the same size from Earth'. Let me introduce the concept *Q-mass* by stipulating that for any object x and real number n, n is the Q-mass of x iff either Q and n is the mass of x, or not-Q and n is zero. Let P_2 and T_2 be the results of replacing 'mass' in P and T with 'Q-mass'.

(iii) Let me introduce the concept *T-friendliness* by stipulating that for any object x, x is T-friendly iff T is true. Let P_3 consist of standard logical vocabulary together with 'T-friendly', and let T_3 be the sentence '$\exists x(x$ is T-friendly$)$' together with its consequences in first-order logic.

If we know anything at all about structure, we know that *mass**, *Q-mass* and *T-friendliness* are not structural. So $\langle P_1, T_1 \rangle$, $\langle P_2, T_2 \rangle$ and $\langle P_3, T_3 \rangle$ had better be considerably worse than the best $\langle$ideology, theory$\rangle$ pairs. But it is not obvious how to understand the notion of goodness in such a way as to distinguish these pairs from $\langle P, T \rangle$, which is presumably quite good by Sider's lights. Considerations of syntactic simplicity will not help, since T_1 and T_2 are both syntactically isomorphic to T, while T_3 is syntactically ultra-simple. And considerations like familiarity and ease of use can hardly be relevant, since they would count heavily against the austere, physics-and-mathematics-based

pairs that Sider favours. We need to take something else into account. What could it be?

One tempting idea is to attribute the badness of $\langle P_1, T_1 \rangle$ and $\langle P_2, T_2 \rangle$ to the fact that mass* and Q-mass are, in fact, much less natural than mass. Because of this, the propositions expressed by T_1 and T_2 are, plausibly, much less natural than the one expressed by T.[2] And Sider's discussion of the role of structure in inductive epistemology (§3.3) suggests that this difference is significant—simpler explanations of our evidence are ceteris paribus more worthy of belief, and the degree of naturalness of a proposition is closely connected to its degree of simplicity.

But the best-pair methodology would lose its point if goodness had to be understood in terms of facts about what is structural/natural. If we want to use the methodology in giving non-question-begging arguments for conclusions about structuralness (and Sider certainly does), we need to be able to make comparisons of theoretical merit with some confidence even when we are still unsure what is structural—otherwise, the methodology will be unusable in the same way as the advice 'believe the truth'. Our assessment of the goodness of an $\langle I, T_I \rangle$ pair must thus somehow prescind from the question how natural the concepts in I actually are. If we want to appeal to some notion of explanatory satisfaction, our question must be something like 'How satisfying an explanation of the phenomena would T_I provide if the structural concepts were those in I?'.[3]

(In fact I see no easy way to reconcile the best-pair methodology with the putative epistemological role of naturalness. Since that role relates the facts about what it is rational to believe to the facts about how natural certain properties or propositions *actually are*, it suggests that when we are uncertain about naturalness, we should normally be uncertain what rationality requires. By contrast, according to the best-pair method-

---

[2] This strategy does not naturally generalise to T_3: while T-friendliness is plausibly less natural than anything in P, this is counterbalanced by the greater syntactic simplicity of T_3.

[3] On satisfaction, see Roger White, 'Explanation as a Guide to Induction', *Philosopher's Imprint* 5: 1–29.

ology, uncertainty about naturalness will arise whenever there are multiple best pairs (p.221), and need not involve uncertainty about the goodness ranking itself.[4])

So can we point to some phenomenon of which T would be a better or more satisfying explanation, if the structural concepts were those in P, than T_1, T_2 or T_3 would, if the structural concepts were those in P_1, P_2 or P_3? Not obviously. For example, if *mass** or *Q-mass* were structural, it seems that T_1 or T_2 would be just as satisfying as explanations of facts about the shapes of planetary orbits as T would be if mass were structural. Indeed, it is rather mysterious how *any* fact not about mass could count as being better explained by $\langle P,T\rangle$ than by $\langle P_1,T_1\rangle$ or $\langle P_2,T_2\rangle$. Once we turn our attention to facts about mass, the situation becomes more delicate. Of course we know many things about mass that are inconsistent with, and a fortiori not well explained by, T_1—for example, that protons are 1836 times as massive as electrons. But the Quinean methodology is surely supposed to explain why it is rational to have beliefs like this, rather than having to treat them as independently given.

One idea worth exploring is that $\langle P,T\rangle$ might be superior because it does a better job of explaining the known fact that 'mass' refers to mass. Let us see how this might play out in the comparison with $\langle P_1,T_1\rangle$. Structural meanings are easy to refer to; so if the structural concepts were those in P, it would not be surprising that physicists should have a word referring to *mass*. On the other hand, if the structural concepts were those in P_1, *mass* would be a rather unnatural quantity (*mass** to the 17th power), making the fact that physicists introduced a word for *mass* quite remarkable.[5]

In evaluating the suggestion that the required discriminations among $\langle I,T_I\rangle$ pairs can be attributed to differences in their capacities to explain facts about reference, we need

---

[4] Brian Weatherson ('The Role of Naturalness in Lewis's Theory of Meaning', forthcoming in *Journal of the History of Analytic Philosophy*) notes a similar tension between the epistemological role of naturalness and the claim that *electronhood* is perfectly natural.

[5] This fact becomes even more surprising when we add that if T_1 were true, an interpretation on which 'mass' means mass* would be more charitable than the correct interpretation as well as being more eligible.

to be clear whether the relevant sets I include the concept *structure*. Ultimately, Sider wants to argue that *structure is structural*, on the grounds that the best $\langle I, T_I \rangle$ is one where I includes '$\mathcal{S}$' and T_I includes '$\mathcal{S}(\Phi)$' for each $\Phi \in I$. The suggestion is that this $\langle I, T_I \rangle$ is better than more austere alternatives which leave out '$\mathcal{S}$' because of its superior capacity to explain facts about reference, similarity, lawhood, and so forth. If so, then since P does not include '$\mathcal{S}$', we cannot award $\langle P, T \rangle$ credit for explaining why 'mass' refers to mass. But perhaps this is not a problem: if $\langle P, T \rangle$ is too austere to be the best pair in any case, there is no obvious need to score it as any better than $\langle P_1, T_1 \rangle$, $\langle P_2, T_2 \rangle$ or $\langle P_3, T_3 \rangle$. To avoid generating an unacceptable level of scepticism about structure, it is enough if our measure gives a higher score to an enriched pair $\langle P^+, T^+ \rangle$—where $P^+ = P \cup \{\mathcal{S}\}$, and T^+ enriches T by adding '$\mathcal{S}(\Phi)$' for all $\Phi \in P^+$ and '$\neg\mathcal{S}(\Psi)$' for all complex expressions Ψ built from vocabulary in P^+—than to correspondingly enriched versions of $\langle P_1, T_1 \rangle$, $\langle P_2, T_2 \rangle$, or $\langle P_3, T_3 \rangle$. However, the cost of denying that $\langle P, T \rangle$ is any better than the other pairs is that it makes it harder to get any intuitive grip on the relevant notion of goodness.

I have two specific worries about Sider's use of the best-pair methodology to argue that *structure* is structural. The first is that Sider's favoured $\langle I, T_I \rangle$ pairs, where I involves only first-order quantifiers, are not actually equipped to provide the relevant explanations of facts about reference, similarity, etc. To give a satisfying explanation of the fact that a word means Φ, it is not enough to say that Φ is natural and 'fits with usage'; we need to say that Φ is more natural than any other concept that fits with usage. Such explanations thus turn essentially on quantification over 'meanings' or 'concepts', and identity among them, neither of which is available in Sider's favoured ideology.

The second worry concerns the contrast between Sider's attitude to facts about similarity and reference and his attitude to, e.g., facts about economics. It is notoriously hard to provide cognitively satisfying explanations of such facts using only the vocabulary of physics. But Sider does not want to conclude on these grounds that concepts like *demand shock* are structural. This pushes us to conceive of the relevant kind of explana-

tory merit in a way that somehow abstracts from considerations of cognitive accessibility. What then is the relevant difference between facts about reference and economic facts, such that the need to explain the former justifies expanding our ideology beyond that of physics while the need to explain the latter does not? Sider addresses this question in §7.13. As I understand it, the response is that if Sider is right about the list of structural concepts, the concepts of the special sciences are not 'highly disjunctive', and are thus explanatorily useful, so that the list can garner some credit for the explanatory successes of those sciences. By contrast, if *structure* is not structural, it *is* 'highly disjunctive', and therefore explanatorily impotent. This argument raises many more questions than I can address here.[6] Here is one worry: if the denial that structure is structural entails that *structure* is highly disjunctive, it plausibly also entails that reference, similarity, and lawhood are highly disjunctive in the same sense; and whatever truth there may be in the idea that highly disjunctive concepts are explanatorily impotent, it is hard to see how there could be anything wrong with invoking a highly disjunctive concept in explaining a phenomenon that is itself highly disjunctive.

Fortunately, we do not need to decide whether structure is structural in order to accept the present suggestion that the belief that mass rather than mass* is structural is justified by its role in explaining why 'mass' refers to mass rather than mass*. Can this strategy be generalised to $\langle P_2, T_2 \rangle$ and $\langle P_3, T_3 \rangle$? It is difficult to say, since Sider's discussion of comparative naturalness (§7.11.1) does not suggest any straightforward way to answer the question how natural *mass* would be if the structural concepts were those in P_2 or P_3. For example, it is unclear how to apply the Lewisian thought that a concept's degree of naturalness depends on the simplicity of its definition in structural terms— mass cannot in any intuitive sense be 'defined' in terms of the concepts in P_2 or P_3, and Sider gives no further account of the relevant notion of 'definition'.

_____

[6] See also Dorr and Hawthorne, *loc. cit*.

A different strategy for explaining why we can reasonably dismiss the hypotheses that the structural concepts are those in P_2 and P_3 is to appeal to something like Lewis's principle that 'truth supervenes on being'—e.g. the claim that all facts supervene on facts expressible in structural terms. Since it seems obvious that there are many facts—about mass, for example—that do not supervene on facts expressible using P_2 or P_3, such a principle could be used to rule out these candidate lists of structural concepts. (One could integrate this with the best-pair methodology by including facts of non-supervenience among the 'phenomena' to be explained.) But Sider is not well-placed to endorse this argument: one central theme of the chapter on modality is that we should, in general, resist arguments against theories about structure which are based on modal intuitions.[7] And he *needs* to take this attitude, since a methodology that accords more weight to modal intuitions will tend to undermine Sider's favoured short, mathematics-and-physics-driven list of structural concepts: many of us intuit that two things could fail to be duplicates even though no predicate of actual-world mathematics or physics applies to either.

The claim that everything supervenes on the structural is one way to precisify the slogan 'the structural is complete'. Sider prefers a different precisification: the demand that 'every language has a metaphysical semantics', in which 'meanings are to be given in purely joint-carving terms' (§7.4). For a non-context-sensitive and purely 'factual' language, a metaphysical semantics takes the form of a theory which includes, for each sentence S, a theorem 'S is true in L iff Φ', where Φ is constructed out of structural expressions. Can this demand take over the dialectical work that we wanted to get out of the supervenience principle, by allowing us to rule out the likes of P_2 and P_3 as candidates to be the set of structural concepts? I find it hard to say, because Sider says little

---

[7] Sider argues that this attitude towards modal arguments follows from his favoured 'Humean' view of modality: modal intuition cannot 'be regarded as probative in matters of fundamental metaphysics' unless it is 'somehow probative concerning the *actual* falsity of rivals; but then there would be no need to bring in possibility; one could argue directly against the rivals' (p.278). But even if the Humean view is true, why couldn't the most effective argument for a nonmodal conclusion have modal premises?

about the criteria for success in metaphysical semantics. So long as a candidate set of structural concepts allows us to construct at least one true sentence and at least one false one, there is certain to be a true biconditional of the required form for every sentence S and language L. What more might be required? If a metaphysical semantics had to be recursively axiomatisable, we might be able to get more mileage out the demand, but Sider imposes no such requirement. And even if he did, it would not help to distinguish $\langle T, P \rangle$ from $\langle T_2, P_2 \rangle$, since we can turn a collection of true biconditionals given in terms of P into one given in terms of P_2 just by replacing 'mass' with 'Q-mass' everywhere.

Sider's discussion suggests that the criteria for success have to do with explanation: the metaphysical semanticist seeks to explain certain aspects of linguistic behaviour, such as why 'English speakers will point to the salient horse, rather than the salient car, when they hear the sounds "Point to the horse!"' (113).[8] But it is unclear how such explanations are supposed to work, given that metaphysical semantics is not, according to Sider, required to integrate with 'theories of action and rationality'. (If we are trying to explain the pointing behaviour in physical or biological terms, expressions like 'true in English' are useless; on the other hand, a satisfactory 'high-level' explanation must surely also advert to the psychological factors which dispose particular hearers to be obedient on particular occasions.) Thus, pending further clarification of the enterprise, it is going to be hard to assess arguments of the form 'Those cannot be the only structural concepts, since in that case certain languages would have no adequate metaphysical semantics'.[9]

Having surveyed various options, then, I still feel that I don't really understand what it means for an $\langle$ideology, theory$\rangle$ pair to be 'good' or 'explanatory' in the sense

---

[8] See also Dorr and Hawthorne *op.cit.*, §4.

[9] Such arguments will, moreover, be rather easy to shrug off until we are told more about the *source* of the demand for metaphysical semantics. Why should the explanation of these particular behavioural facts should be subject to a constraint—that it yield a system of biconditionals in which only structural expressions occur on one side—that does not apply to explanations in general?

Sider has in mind. Pending further clarification of this notion, the dialectical force of Sider's invocations of the best-pair methodology in arguing for and against specific structuralness claims will inevitably be limited.

Since Sider initially develops the best-pair methodology as a response to the worry that facts about structure are 'epistemically inaccessible', I should stress that I am not bothered by this worry. One straightforward way to find out about structure is to rely on the connections to other subject matters (similarity, laws, reference…) that form part of structure's 'inferential role' (§2.1). Using these connections, we can exploit our knowledge of those subject matters to gain knowledge about structure. For the moment, I expect that arguments of this general form will provide the most dialectically effective way to address the question which concepts are structural.