[I plan to expand this reply soon.]

How can we form reasonable beliefs about structure, about where reality’s joints lie? Cian Dorr’s preferred strategy is to exploit the connections between structure and other subject matters. Much of my book argued for specific connections between structure and: reference, similarity, fundamentality, necessity, laws, explanation, substantivity, objectivity, reasonable credence. As Dorr puts it, “using these connections, we can exploit our knowledge of those subject matters to gain knowledge about structure”. (This strategy is not undermined if structure is more metaphysically fundamental than reference, similarity, and so on. To repeat what has become almost a cliché in recent metaphysics: the order of epistemic priority need not match the order of metaphysical priority.)

My book recommended a different strategy:

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 $\langle I, T \rangle$ of ideology $I$ and theory $T$ in terms of that ideology. (p. 13)

I agree that Dorr’s strategy is a reasonable one, but I think that mine is too.

My best-pair strategy has little content until the sense of ‘best’ (and ‘most explanatory’, ‘simple’, etc.) is specified, and Dorr rightly complains that I said little about this. I said little because I wanted to remain as neutral as possible, because I thought that existing accounts of explanation, simplicity, and so forth—largely from the philosophy of science—would be applicable, and because I don’t have detailed ideas about the matter. But keeping the account at such a high level of abstraction is admittedly worrisome, and Dorr argues that when we bear down on specific cases, we can see that the best-pair strategy does not work.¹ I appreciate Dorr pushing me on this, and hope to deepen my account by addressing this issue in more detail than I did in my book.

¹Along the way he raises a number of other important and challenging concerns about my book. Unfortunately I cannot address them more fully than to say: i) in reply to Dorr’s first concern about my argument that structure is structural: set-theoretic constructions of meanings, including of structured meanings, are available; ii) in reply to his second concern:
Before considering Dorr’s specific cases, let me say in slightly more concrete—though still highly abstract—terms what I meant by ‘best’. At the very least, the best ideology-theory pair \( (I, T_I) \) ought to predict as much of the data (what we observe) as possible, and it ought to contain as simple and powerful laws as possible, where simplicity is measured in terms of \( I \).

Syntactically simple generalizations using predicates like ‘mass’ and ‘charge’ have been found which, together with bridge laws and descriptions of background assumptions, imply much of the data; thus the criterion favors the ideology and theory of physics. Predictive generalizations using zodiac-theoretic ideology, on the other hand, have not been found; the criterion rules against astrology.

But now consider the first of Dorr’s three specific cases. Dorr compares a “state-of-the-art physical theory” \( T \), containing the predicate ‘mass’ (measured in Planck units), with a theory \( T_1 \) that results from \( T \) by replacing ‘mass’ everywhere with the intuitively non-joint-carving predicate ‘mass*’, where the mass* of an object is defined as the seventeenth root of its mass. The challenge is to say why \( T_1 \) is a “worse” theory. The challenge is supposed to be hard to answer since the theories are syntactically identical.

But can’t we simply reply that \( T_1 \) is worse because it is known to be false? Supposing \( T \) to contain the sentence ‘the ratio of the mass of the proton to that of the electron is 18\( \frac{3}{6} \)’ (this is in fact the correct ratio), \( T_1 \) contains the false sentence ‘the ratio of the mass* of the proton to that of the electron is 18\( \frac{3}{6} \)’.

Dorr notes all this, but replies that my best-pair strategy “is surely supposed to explain why it is rational to have [the background physical knowledge that the first rather than the second sentence is true] rather than having to treat [this knowledge] as independently given”. The best-pair strategy is indeed supposed to be part of a general epistemology that explains the possession of the background knowledge, but this general epistemology accounts for both the background knowledge and the knowledge of where the joints lie. We may think of how physicists arrived at \( T \) as follows: after many false starts, they realized that if they posited a certain array of quantities—which they called ‘mass’, ‘charge’, and so on—and hypothesized that these quantities obeyed certain laws—as described in \( T \)—then the data could be explained in a simple manner. This realization justifies both the belief that \( T \) is true (and hence that

---

2 The way in which physics underlies other successful theories is another part of the case for physics.
$T_1$ is false) and also the belief that the posited quantities carve at the joints. Dorr describes his two other cases as follows:

(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...$T_2$ be the [result] of replacing ‘mass’ in...$T$ 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...$T_3$ be the sentence ‘$\exists x (x$ is $T$-friendly)’ together with its consequences in first-order logic.

Unlike $T_1$, the defects of $T_2$ and $T_3$ do not include falsity. Their defects concern, rather, their laws and how those laws explain. Take any datum that is entailed by sentences in $T$ together with appropriate background assumptions. Since $Q$ is not implied by $T$, there is no guarantee that the corresponding sentences in $T_2$ plus the same background information entail the datum. Granted, if we add to the background information the additional fact that $Q$ is true (plus the definition of ‘$Q$-mass’), then the entailment goes through. But the fact that we need to add this additional fact to each explanation invoking $Q$-mass detracts from the simplicity of those explanations. (Given how often this same piece of “background information” must be cited, $T_2$ in effect treats $Q$ as an extra law.) Furthermore, explanations using $T_2$ are fragile in the sense that they work only on the assumption that $Q$; and this fragility surely detracts from $T_2$’s explanatory value. As for $T_3$, the problem is that its laws do not entail the data (not even together with appropriate background information). Only by adding the additional assumption that $T$ is true do they entail the data; but then the theory is no simpler than $T$. (There are, admittedly, delicate questions about how to evaluate simplicity for theories employing a truth predicate.)