The Churchlands and their Critics

Edited by
Robert N. McCauley

BLACKWELL Publishers

6

Paul Churchland and State Space Semantics

Jerry Fodor and Ernie Lepore

Introduction

A number of philosophers believe that the notion of identity of meaning should be replaced by some graded notion of similarity of meaning. With a graded notion of similarity of meaning, intentional generalizations would be viewed as subsuming individuals in virtue of the similarity of their mental states; translation and paraphrase would be viewed as preserving not identity but similarity of meaning. The prospects, however, for constructing a "robust" notion of similarity of meaning – one adequate to the purposes of semantics and cognitive science – are, we believe, remote. This prognosis isn't widely shared. That may be because friends of semantic similarity have generally been careful not to say what they take semantic similarity to consist in. Paul Churchland, however, has proposed a sketch of the sort of similarity theory that is required, a notion of mental (or, anyhow, neural) representation that "... embodies ... metrical relations ... and thus embodies the representation of similarity relations between distinct items thus represented" (1989b: 102; barring notice to the contrary, emphases are Churchland's throughout).

Since Churchland's attitude towards the intentional/semantic generally tends to be eliminativist, it's unclear just what properties of contentful states his "state space" representations are supposed to preserve. Suffice it that, when he's in a "highly speculative" mode, he contemplates, for example, the possibility of "...a way of representing 'anglophone linguistic hyperspace' so that all grammatical sentences turn out to reside on a proprietary hypersurface within that hyperspace, with the logical relations between them reflected as spatial relations of some kind ... [This would hold out] the possibility of an alternative to, or potential reduction of, the familiar Chomskyan picture" (1989b: 109). This entails that state spaces can represent grammars and such. Like

much else that he says, it certainly sounds as though Churchland has in mind a kind of representation that specifies the contents of neural states in which case he is into intensionality up to his neck. In any event, we propose to read him that way and ask how much of the intuitive notion of content similarity state space representation allows us to reconstruct.

In what follows, we hope to convince you that, for all that's on offer so far, the problem of semantic similarity appears to be intractable.

State Space Representation

"The basic idea . . . is that the brain represents various aspects of reality by a position in a suitable state space; and the brain performs computations on such representations by means of general coordinate transformations from one state space to another" (1989b: 78–9).

For our present purpose, which is semantics rather than the theory of mental processes, only Churchland's account of neural representation need concern us. We commence by trying to make clear how Churchland's state space proposal connects with the more familiar "network" picture of semantics. Churchland's state space proposal is, we suggest, profitably viewed as an attempt to generalize the more familiar "network" picture of semantics and to free it from its specifically empiricist assumptions.

Suppose we start with a roughly Quinean picture of the structure of theories (/languages/belief systems). According to this picture, there are two sorts of ways in which the (nonlogical) symbols belonging to a theory get semantically interpreted. The semantics of the "observation vocabulary" is fixed by conditioning (or other causal) relations between its expressions and observable properties of the distal or proximal environment. The semantics of the rest of the vocabulary is fixed by a network of inferential or (in case the semantics is intended to be naturalistic) causal/associative relations to one another and to the observation terms. The semantic theory of a language thus represents its vocabulary as nodes in a network, the paths of which correspond to semantically-relevant relations among the vocabulary items. Observation terms are at the "periphery" of the network; nonobservational vocabulary is further in. We take it that this geography is familiar.

Recall how the problem of content identity arises on this "network" picture. If the paths to a node are collectively constitutive of the identity of the node (as presumably they will be if, as Quine holds, no analytic/synthetic (a/s) distinction is assumed), then only identical networks can token nodes of the same type. Identity of networks is thus a sufficient condition for identity of content, but this sufficient condition isn't

robust; it will never be satisfied in practice. The long and the short of it is: a network semantics offers no robust account of content identity if it is denied access to an a/s distinction. But maybe a network semantics can nevertheless be made to offer a robust notion of content similarity? The present proposal is to make content similarity do the work that content identity did in semantic theories that endorsed the a/s distinction.

How, then, might a robust metric of content similarity be constructed; one which is defined for nodes belonging to networks that are (perhaps arbitrarily) different from each other? An immediate problem is this: according to the usual understanding, the only fixed points in a network are the nodes that correspond to observation vocabulary. It's only these peripheral nodes that can be identified without specifying the rest of the network that contains them. (We're supposing that we know what it is for two arbitrarily different theories to both have a node that expresses an observable property like red; it's for both to have vocabulary items that are appropriately connected – for example, conditioned – to redness.) It thus appears that, if we are to define a similarity relation over terms in the nonobservation vocabulary, it will have to be by reference to their (direct or indirect) relations to observation terms.

But this picture might well strike one as intolerably empiricistic. It just doesn't seem to be true that the dimensions of content along which words (concepts) can be similar are reducible to the various ways in which they can be connected to observables. There is plausibly something semantically relevant that everything subsumed by the concept uncle has in common with everything subsumed by the concept aunt; but a couple of hundred years of unsuccessful empiricism suggest that what they have in common is not expressible by reference to the observable properties of aunts and uncles. Similarly, mutatis mutandis, for the similarity between the things subsumed by the concept ice and the things subsumed by the concept steam; or between the thing subsumed by the concept the President of the US and the things subsumed by the concept Cleopatra. And so on, endlessly.

Churchland's state space story is best understood in this context. At least since *Scientific Realism and the Plasticity of Mind* (1979), he has been attracted to network semantics and inclined to think that a good semantics must make similarity, rather than identity of content its basic theoretical notion. But he is also suspicious of the sort of empiricism which reduces all semantically relevant relations eventually to relations to observation vocabulary.

> The semantic identity of a term derives from its specific place in the embedding network of the semantically important sentences of the language as a whole. Accordingly, if we wish to speak of sameness of meaning

> across languages, then we must learn to speak of terms occupying analogous places in the relevantly similar networks provided by the respective sets of semantically important sentences of the two languages at issue (1979: 61).

However, he goes on to say a few pages later:

> . . . the aims of translation should include no fundamental interest whatever in preserving observationality . . . languages, and the networks of beliefs that they embody, have an identity that transcends and can remain constant over variations in the particular sensory conduits to which they happen to be tied, and in the particular locations within the language where the sensory connections happen to be made. Accordingly, any conception of translation that ties its adequacy to the preservation of "net empirical content" as conceived by Quine will lead to nothing but confusion (1979: 65–6).

For Churchland, the question is thus how to free the network picture of semantics from its empiricist assumptions, and somehow to generate a robust notion of content similarity in the course of doing so. We read his paper "Some reductive strategies . . ." (1989b) as a failed attempt at this, and we'll argue that a recidivist empiricism is in fact its bottom line.

Churchland's current proposal may now be summarized: A "Quinean" network semantics of the sort we have been discussing can be thought of as describing a space whose dimensions correspond to observable properties, and where each expression of the object language is assigned a position in the space. To say that the concept dog is semantically connected to the properties of barking and tail wagging is thus equivalent to saying that it occupies a position in semantic space that is (partially) identified by its value along the barkingness and tail-waggingness dimensions. Since the empiricism of the standard network proposal resides in the requirement that all the dimensions of the semantic space in which the concepts are located must correspond to observable properties, all you have to do to get rid of the empiricism is abolish this requirement. What's left are semantic state spaces of arbitrary dimensions, each dimension corresponding to a parameter in terms of which the semantic theory taxonomizes object-language expressions, and where similarity of content among the object language expressions is represented by adjacency relations among regions of the space.

Let's now see how Churchland proposes to develop a theory of the semantics of mental representation that accords with this conception. Rather surprisingly, Churchland's analysis starts, not with the paradigms of intensionality (propositional attitudes and concepts) but with sensations. The more or less explicit suggestion is that if we had a treatment

that provided an illuminating semantical account for sensations it might generalize to mental representation at large. Let us, therefore, consider how the state space story is supposed to apply to sensations, bearing in mind that it is Churchland's account of mental representation, rather than his theory of qualitative content, that is our primary concern.

> The qualitative character of our sensations is commonly held to pose an especially intractable problem for any neurobiological reduction of mental states . . . and it is indeed hard to see much room for deductive purchase in the subjectively discriminable but "objectively uncharacterizable" qualia present to consciousness. . . . Even so, a determined attempt to find order rather than mystery in this area uncovers a significant amount of expressible information. . . . Consider . . . the abstract three-dimensional "color cube" proposed by Edwin Land, within which every one of the many hundreds of humanly discriminable colors occupies a unique position or small volume . . . Each axis represents the eye/brain's reconstruction of the objective reflectance of the seen object at one of the three wavelengths to which our cones are selectively responsive. Two colors are closely similar just in case their state-space positions within this cube are close to one another. And two colors are dissimilar just in case their state-space positions are distant.

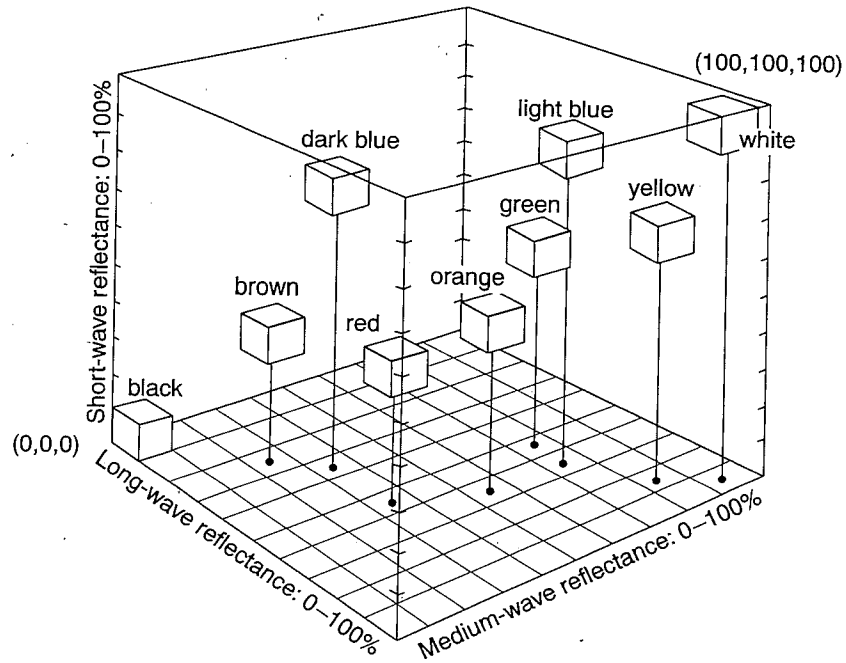

Figure 6.1 Color state space

> We can even speak of the degree of the similarity, and of the dimensions along which it is reckoned (1989b: 102–3).

We emphasize that Churchland views this as an account of the qualitative content of color sensations, not just of the nervous system's capacity to discriminate among colors. As Churchland goes on to say,

> In particular, it suggests an effective means of expressing the inexpressible. The "ineffable" pink of one's current visual sensation may be richly and precisely expressible as a "95 Hz/ 80 Hz/ 80 Hz chord" in the relevant triune cortical system. ... This more penetrating conceptual framework might even displace the common-sense framework as the vehicle of inter-subjective description and spontaneous introspection (1991a: 106).

How plausible is this story for the representation of the qualitative content of sensations? And, how close does it get us to a robust notion of content similarity in general? We'll say just a word or two about the first question, saving most of our attention for the second.

There is, notoriously, a problem of qualitative content that philosophers of mind worry about. It's closely connected with problems about qualia inversion. For example, it seems conceptually possible that the sensation you have when you see things that are grass colored is "just like" the sensation that I have when I see fire engines. If this inversion is systematic, then perhaps there is nothing – in particular, there is no behavioral consequence of our capacity to respond selectively to colors – that would tell this case apart from the normal one in which grass colored things look the same to you as they do to me. The possibility of inverted qualia thus looks to show that behaviorism is false, and an extension of the same considerations suggest that it may show that functionalism is false too.

You might suppose that a theory of the qualitative content of sensations ought to resolve this problem. After all, it's supposed to be precisely qualitative content that gets inverted in inversion examples, and precisely the notion of identity of qualitative content that the examples render equivocal. Churchland's account of qualitative content is, however, of no help at all with these issues. The reason is that if qualia inversion makes any sense at all, it seems conceptually possible that you and I should share the state space pictured in the figure, but that the labels on your cube should be inverted with respect to the labels on mine. Notice that the reason this seems to be conceptually possible is that the dimensions of this state space specify physical properties of visual stimuli rather than parameters of qualitative content *per se*. Since the relation between the property of being a 95 Hz/ 80 Hz/ 80 Hz chord

and being a sensation of ineffable pink would appear to be thoroughly contingent (or, at an any event, thoroughly nonsemantic), it would seem to be conceptually possible that something should have the first property but not the second. This just is the qualia inversion problem; there appears to be no property of a sensation except its qualitative content, upon which its qualitative content is guaranteed to supervene. (In particular, there appears to be no behavioral, or functional, or neurological property upon which it is guaranteed to supervene.) So if you were worried about the qualia problem before you read Churchland, what you should do is keep worrying.

To put this same point in an old-fashioned way, the dimensions of Churchland's state space appear to specify qualia by reference to properties they have nonessentially, and any such specification begs the inversion problem. (Or, if you think that it is "metaphysically necessary" that color sensations have the psychophysical properties that they do, then our point is that this necessity is not engendered by any semantical connection between sensation concepts and psychophysical concepts.) You could, in consequence, know perfectly well that a certain sensation corresponds to a certain "chord in the relevant triune cortical system" and have no idea at all of "what it's like" to have a sensation of that kind, or, indeed, that there is anything that it is like.

The problem so far is that the dimensions in terms of which Churchland proposes to taxonomize qualia don't specify their content. Rather, they appear to taxonomize qualia by psychophysically sufficient conditions for having them. But it might be thought that this is a defect of the example, not a defect of state space semantics as such. Why not stick to the state space notion of mental representation, but add the proviso that the dimensions of the semantic space must really be semantic; they have to taxonomize content bearing states by their contents. Perhaps concepts (like aunt, uncle, steam, ice, the President of the US, and Cleopatra) can be identified with positions in a state space of semantically relevant dimensions, so that similarities among these concepts could be identified with adjacencies in the state space. (The President of the US is close to Cleopatra on the politician dimension, but maybe less close on the nubile dimension.) This, as opposed to the vicissitudes of Churchland's treatment of qualia, is the issue we are really interested in.

In fact, however, we now propose to argue that this suggestion is without substance. The same problems that traditionally arose for theories of content identity also arise for this theory of content similarity, so the appearance of progress is simply an illusion. To begin with the crucial point: the state space story about content similarity

actually presupposes (and therefore begs) a solution to the question of content identity.

Problems of State Space Semantics

The Individuation of Dimensions

What Churchland has on offer is the idea that two concepts are similar insofar as they occupy relatively similar positions in the same state space. The question thus presents itself: when are S1 and S2 the same state space? When, for example, is your semantic space a token of the same semantic space state type as mine? Well, clearly a necessary condition for the identity of state spaces is the identity of their dimensions; specifically, identity of their semantic dimensions, since the current proposal is that concepts be located by reference to a space of semantically relevant properties. We are thus faced with the question when x and y are the same semantic dimensions (for example, when positions along x and y both express degrees of being a politician, or of nubility). But this is surely just the old semantic identity problem back again. If we don't know what it is for two words both to mean nubile, then we also don't know – and for the same reasons – what it is for two spaces both to have a nubility dimension. Perhaps it will be replied that semantic similarity doesn't, after all, require concepts to be adjacent in the very same state space; perhaps occupying corresponding positions in similar state spaces will do. That a regress has now appeared is, we trust, entirely obvious.

It's worth getting clear on what has gone wrong. The old (empiricist) version of network semantics had a story about the identification of the dimensions by reference to which it did its taxonomizing; they were to express observable properties, and an externalist (for example, a causal) theory of some kind was to explicate the relation between observable properties and terms in the observation vocabulary. In particular, that relation was assumed to be specifiable independent of the interpretation of the rest of the vocabulary. However, as we've seen, Churchland's proposal comes down to the idea that the dimensions of semantic state space don't generally correspond to observable properties; they can correspond to whatever properties the brain may represent. This avoids empiricism, all right, but it begs the question how identity of state spaces is itself to be determined. On the one hand, we are assuming that dimensions of semantic state spaces can express whatever properties you like. And, on the other hand, we don't have and can't assume any identity criterion for dimensions that express other than observable

properties. And, on the last hand, to take such a criterion for granted would just be to beg the semantic identity problem.

To repeat: We have a robust notion of semantic similarity only if we have a criterion for the identity of state spaces. We have a criterion for the identity of state spaces only if we have a criterion for identity of dimensions of state spaces. And we have a (nonempiricist) criterion for the identity of dimensions of state spaces only if we have a criterion of "property expressed by a dimension of a state space" that works for arbitrary properties, not just for observable properties. But a criterion for "property expressed" that works for arbitrary properties is just a criterion for identity of meaning. So Churchland's proposal for a robust theory of content similarity fails to avoid the problem of robust content identity. (And, of course, fails to solve it.)

In our book *Holism: A Shopper's Guide*, we offer it as a plausible methodological principle that you can't have a robust notion of content similarity (one that applies across languages, across minds, or across theories) unless you have a correspondingly robust notion of content identity. Churchland's space state semantics provides a graphic illustration of how this principle applies. His explication of an interpersonal notion of content similarity as proximity in semantic state space presupposes an interpersonal notion of identity for the semantic spaces themselves, a notion that Churchland leaves entirely without explication. In consequence, if you're worried about how concepts can be robust, Churchland's state space semantics provides no illumination at all.

We're claiming, in effect, that Churchland has confused himself by taking the labels on the semantic dimensions for granted. The label on a dimension says how positions along the dimension are to be interpreted; for example, it says that they're to be interpreted as expressing degrees of F-ness. To label a dimension as the F-ness dimension is thus invite the question "In virtue of what do the values of this dimension express degrees of F-ness rather than, say, degrees of G-ness?" (Equivalently, for these purposes "What makes it the case that a dimension in your state space expresses the same property F as some dimension in mine does?"). Patently, a semantic theory mustn't beg this sort of question on pain of assuming the very concepts it is supposed to explicate.

Cognitive scientists are forever getting themselves into trouble in this way; it's a fallacy that is particularly endemic among connectionists. Connectionists draw diagrams in which the label on a node tells you what the intentional interpretation of the excitation of the node is supposed to be. But no theory is offered to explain why a node gets the label that it does; it's just semantics by stipulation.

Churchland makes exactly this mistake, only it's the dimension labels