

SINBAD Neurosemantics: A Theory of Mental Representation

DAN RYDER

Abstract: I present an account of mental representation based upon the ‘SINBAD’ theory of the cerebral cortex. If the SINBAD theory is correct, then networks of pyramidal cells in the cerebral cortex are appropriately described as representing, or more specifically, as *modelling* the world. I propose that SINBAD representation reveals the nature of the kind of mental representation found in human and animal minds, since the cortex is heavily implicated in these kinds of minds. Finally, I show how SINBAD neurosemantics can provide accounts of misrepresentation, equivocal representation, twin cases, and Frege cases.

1. Introduction

As physical beings, how is it that we manage to represent the world? We now have very strong evidence that the human mind is identical with or somehow realized in the brain, specifically in the cerebral cortex. It would thus seem reasonable to hope that an understanding of the cortex might help us understand mental representation, if not in general, then at least in our own case. In taking this particular a posteriori approach, it must be admitted that we run the risk of failing to discover any deep facts about how we manage to represent the world. Perhaps we will discover only that, however it is done, it is done with neurons (about three pounds of them). I decided the risk was worth taking, and in this paper, I will try to show that the results are significantly more impressive: the approach yields a naturalistic account of the principles determining reference (and ultimately truth conditions) for our kind of mental representation—in other words, a neurosemantics.

Some will be skeptical of my strategy from the outset, not because they doubt the mind is realized in the brain, but because they doubt the relation between mind and brain is sufficiently transparent that it may be understood by investigating the ‘hardware’. The standard view is that mind and brain may be linked only via multiple levels of composition and explanation. I believe the link is much more

Many thanks to all the people who provided me with valuable comments and discussion, including Dorit Bar-On, Scott Brockmeier, Ruth Campbell, Mark Changizi, Fred Dretske, Chris Eliasmith, Paco Garzón, Chris Hill, Bill Lycan, Mohan Matthen, Ruth Millikan, Stan Munsat, Ram Neta, Jesse Prinz, Zena Ryder, Tim Schroeder, Marius Usher, an anonymous referee for *Mind & Language*, and especially Oleg Favorov.

Address for correspondence: Department of Philosophy and Program in Cognitive Science, Sycamore 026, 1033 East 3rd Street, Indiana University, Bloomington, IN 47405-7005, USA.
Email: dan@danryder.com

direct, but rather than presenting independent arguments against the standard view, I will just let my theory stand forth as a stubborn counterexample.

The 'neurosemantic' theory that I present is derived from the SINBAD model of the cortex (Favorov and Ryder, forthcoming; Ryder and Favorov, 2001). 'SINBAD' stands for 'Set of INteracting BAcKpropagating Dendrites'; it is a computational theory of cortical plasticity based on functional considerations as well as anatomical and physiological evidence. If the theory is correct, networks in the cortex have a powerful tendency to structure themselves isomorphically with regularities in their environment. This tendency is enormously useful, because it allows our brains to yield reliable predictions. Thus if the SINBAD theory is correct, it is reasonable to suppose that it is the biological function of cortical networks to structure themselves isomorphically with their environment. This means that they represent, or more specifically, *model* their environment.

The special features of these representations are what I want to seize upon in order to give an account of *mental* representation. Not only are SINBAD representations to be found in the cortex (the seat of the mind), they represent just the sorts of things that *we* represent (a large class of items that I call 'sources of correlation'). Further, they are eminently suited to play the theoretical role assigned to mental representations by both scientific and folk psychology. In this paper, I will concentrate upon some of the most important aspects of this theoretical role that have been the focus of attention in recent discussions of the naturalization of mental representation, including broad content, the possibility for misrepresentation and equivocal representation, the representation of non-existent things, and cognitively distinct but referentially equivalent representations (i.e. Frege cases). But there is much more about the mental that can be understood with the help of SINBAD neurosemantics, notably the relation between concepts and inferential roles, the explanatory role of content, propositional attitudes, and concept acquisition (Ryder and Favorov, 2001; Ryder, 2002; Favorov *et al.*, 2003).

2. The organization of regularities around sources of correlation

Perception and cognition, the processes that involve mental representations, are valuable because they afford prediction. When an organism recognizes an object as a banana, it can reliably predict that the object is edible. To see how even the very basic act of recognizing food and eating it involves a form of prediction is to see how ubiquitous and all-important prediction truly is. Of course, that an organism predicts and acts accordingly suggests (or even entails) that this prediction involves some kind of *understanding*. But there is a scare-quotes sense of 'prediction' that does not require understanding. The single-celled organism *Didinium* possesses a mechanism that allows it to 'predict' based on sensory information whether the thing contacting it will provide sustenance. If it 'judges' in the affirmative, it will release paralyzing darts from its snout and then ingest its prey, usually another protozoan (Alberts *et al.*, 1989, p. 22). Multicellular organisms have evolved more

elaborate mechanisms to accomplish even better ‘predictive’ tricks. In the mammalian nervous system, ‘prediction’ has reached its apotheosis—or is this real prediction, without scare quotes? I will begin by presenting the SINBAD network as something that performs a ‘predictive’ trick. Then, in sections 5 and 6, I will try to convince you that this mere ‘predictive’ trick constitutes real, intentional prediction by showing that SINBAD cells genuinely represent.

Prediction requires regularity. The fish ‘predicts’ that it will find algae in the calm part of the brook, not because it grasps the laws of ecology, but because it makes use of them in a superficial way. I am primarily concerned with these manifest regularities, especially insofar as they occur in clusters, which pattern permits SINBAD networks to ‘predict’. This clustering pattern has been used as the basis for several closely related, metaphysically innocuous accounts of natural kinds (see especially Boyd, 1991; Kornblith, 1993; Millikan, 1999). We will be going beyond such kinds, but they are a natural place to start.

The ‘unified property cluster’ account says that a natural kind is characterized by a set of correlated properties, where some further principle explains *why* they are correlated, and thus why reliable inductive generalizations can be made over them. For example, water is a substance with multiple correlated properties such as liquidity in certain conditions, clarity, and the ability to dissolve certain other substances, where these ‘surface properties’ are explained by water’s nature or hidden essence, namely its molecular structure. The presence of a group of these correlated properties allows one to infer that a substance is water, which in turn allows one to infer the presence of those properties that the sample of water is not currently exhibiting (e.g. it would freeze at 0° C).

This pattern of regularity organized around a ‘source of correlation’ is not restricted to chemical natural kinds. As J. S. Mill remarked (quoted in Hacking, 1991):

...a hundred generations have not exhausted the common properties of animals or plants. . .nor do we suppose them to be exhaustible, but proceed to new observations and experiments, in the full confidence of discovering new properties which were by no means implied in those we previously knew.

Mill is describing the phenomenon that biological kinds tend to be ‘inductively rich’ (Gelman and Coley, 1991), i.e. they tend to be sources of large amounts of predictively useful correlation. In the case of biological kinds, these correlations are not due to an underlying chemical structure, but rather due to their common evolutionary history.

Millikan (1998; 1999) extends the unified property cluster account beyond natural kinds. Non-natural (but real) kinds also have multiple correlated properties unified by some explanatory reason. Artifacts, for instance, will often have correlated properties because they serve some specific function (e.g. screwdrivers), because they originate from the same plan (e.g. Apple’s iMac), or because they

have been copied for sociological reasons (e.g. a coat of arms and its variants). She also points out that individuals fall into the same pattern. Individuals persist through time, and in doing so tend to retain many of their properties. Each one is thus a source of a large number of multiple correlations among properties. Events and processes, whether particulars or kinds, are also sources of correlation, for instance Halloween, World War II, biological growth, and atomic fusion. Dynamical systems, like homeostatic systems, economic systems, ecosystems, organisms, and many artifacts, have an underlying causal structure that is a source of correlated variation among the systems' global effects.

The point is that this general pattern of regularities organized around sources of correlation is ubiquitous, although different sorts of unifying principles may underlie different instances of multiple correlations. The SINBAD theory says that the principal cells of the cerebral cortex are built to take advantage of the general pattern of regularity characterized by sources of correlation. This general pattern of regularity allows SINBAD networks to become dynamically isomorphic to the environment, and a dynamical isomorphism is a very useful 'predictive' trick.

3. The dispositional profile of SINBAD cells¹

I will be claiming that some neurons have representational powers. The neurons to which I attribute representational powers are *pyramidal cells* (see figure 1), the most common cell type in the cerebral cortex (70% to 80% of the neurons in the cortex fall into this class—see Abeles, 1991; Douglas and Martin, 1998). In this section and the next, I will describe a 'predictive' trick that may be performed by networks of these cells, and which takes advantage of patterns of regularities due to sources of correlation. It is this 'predictive' trick that the SINBAD theory describes. (By itself, the SINBAD theory makes no claims about representation; that is an extra, philosophical step to be presented in sections 5 and 6.)

Like any other neuron, a pyramidal cell receives inputs from other neurons via synaptic connections on its dendrites, elaborate tree-like structures covered in thousands of synapses. Each principal dendrite—an entire tree-like structure attached to the cell body (see figure 1)—produces an activity determined by all of the excitatory and inhibitory inputs that it receives (Spruston *et al.*, 1999). This activity is that dendrite's output, which it passes onto the cell body. The output of the whole cell (delivered elsewhere via the cell's axon) is determined in turn by the outputs of its five to eight principal dendrites (Feldman, 1984).

The input/output profile of a dendrite, and thus its contribution to the whole cell's output, can be modified by adjusting the strengths of both excitatory and inhibitory synaptic connections, and possibly by modifying other properties of the

¹ This section presents a basic introduction to SINBAD networks. For the full story, see Ryder and Favorov (2001), Favorov *et al.* (2003), and Favorov and Ryder (forthcoming).

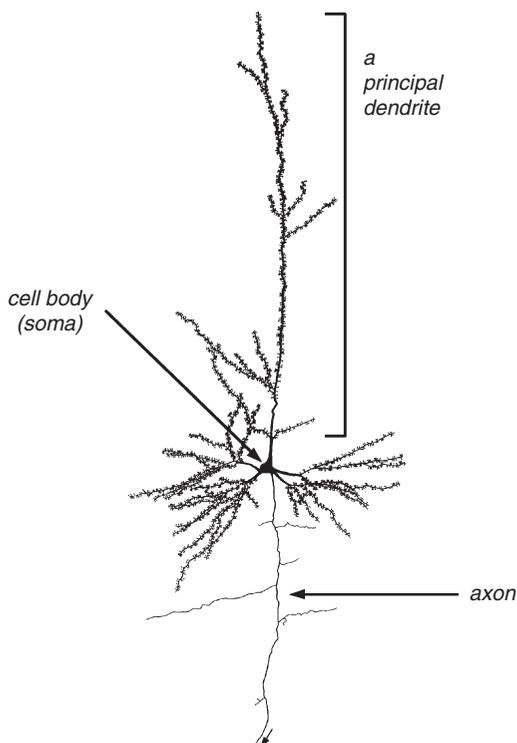


Figure 1 A typical cortical pyramidal cell. The dendrites form the input region of the cell, which transmits its output via the axon. There are a total of five principal dendrites visible on this cell. ('Dendrite' can refer either to a principal dendrite or a sub-branch of a principal dendrite.) Axons from other neurons synapse on one or more of the thousands of tiny spines covering the dendrites; inhibitory synapses may also occur between spines. This cell type's name describes the pyramid-like shape of the cell body, which is due to the arrangement of principal dendrites (one extending upwards, the others radiating horizontally and obliquely downwards).

dendrite as well (e.g. structural properties: see Woolley, 1999; McAllister, 2000). In this way, each dendrite comes to have a causal profile that implements some mathematical function relating synaptic inputs to dendritic output. The cell as a whole also comes to implement some function: in neuroscientific terms, this is what 'learning' consists in for the cell. I will call the process of a dendrite's learning a particular function its 'finding a function' or 'acquiring a function'. (It should be noted that this learning occurs continuously, and at all times the cell is computing *some* function of its input. To 'find a function' is for the cell to reach some stable state, where the function it computes remains relatively constant.)

What principles underlie the adjustments a cell makes to its synaptic connections in order to acquire some function from synaptic input to axonal output? Why do certain connections become highly influential, while others get ignored or even dropped?

And what determines the nature of the influence they come to exert? This influence can be quite nuanced, and it has been shown that dendrites are capable of computing highly complex functions (Mel, 1994; Mel, 1999; Gurney, 2001; Poirazi and Mel, 2001). The SINBAD idea is that learning in cortical pyramidal cells is governed by the following principle: each principal dendrite will adjust its connections so that it will consistently tend to contribute to the cell's output equally with its fellow principal dendrites. That is, if the cell's output = OUT, and the number of principal dendrites = n , each dendrite will tend towards implementing a function whereby it contributes OUT/n whenever the cell is active. So if there are 5 principal dendrites, like on the cell in figure 1, they will each tend to adjust their connections over time so that they will consistently contribute 1/5th of the cell's total output. I'll put this by saying 'they try to match each other's activities'. 'Trying', of course, is just a convenient metaphor. They are not literally trying, it is merely a brute causal tendency that they have.

This brute causal tendency is likely mediated by a physiological signal that is propagated *backwards*, from the cell body to the tips of each principal dendrite. (Thus the SINBAD acronym: the cell as a whole is a Set of INteracting BAcKpropagating Dendrites.) It is hypothesized that this signal carries information about the activity of the cell, information that can be used by the dendrite to modify the function it computes. It is widely accepted that some such signal must exist (for a physiological candidate, see Stuart, 1997). However, it is unclear exactly how much information the signal carries.

At the impoverished end of the scale, the backpropagating signal may just indicate to a dendrite that its host neuron has fired. Such a signal could be the basis for implementing some variety of *Hebbian* learning rule, whereby excitatory synapses that contributed to the cell's output are strengthened (although not beyond some maximum value), while inactive synapses are weakened. Under this type of rule, a principal dendrite would have a tendency to learn to compute a simple function that correlates with the cell's output. Since all of a cell's dendrites are learning to correlate with the same output, this rule will result in the dendrites having some tendency to match.

A more informationally rich type of backpropagating signal would be an error or mismatch signal, one that is proportional to the difference between OUT/n (the 'desired' output) and that principal dendrite's actual output. Such a signal has the following effect, local to each synapse: it causes synaptic and other modifications so as to make the value of future error signals tend toward zero. Each principal dendrite 'tries' to reduce the value of its error signal every time the cell is active, i.e. the brute causal learning mechanism disposes it to produce OUT/n next time. Over time, each principal dendrite will tend toward consistently having an error signal of zero no matter what inputs the cell gets; at this point, each dendrite will consistently contribute the same amount, OUT/n , to the cell's output.² Typically, error

² Those familiar with the neural computation literature will recognize the kinship of this idea with Becker and Hinton's Imax learning procedure (1992), which maximises the mutual information between the outputs of two network modules that receive different but related sensory input. See also Phillips and Singer (1997), who make use of this idea in a Hebbian context.

correction learning raises the spectre of some mysteriously knowledgeable ‘teacher’ telling the learning network(s) all the answers. There is no such worry here; a SINBAD cell’s dendrites teach *each other*. Together they both produce and learn from the cell’s output.

There are various possibilities for how this error signal could be used by a dendrite to modify the function it computes. When the principal dendrite’s activity fails to match OUT/n, modifications are determined by the way ‘blame’ is assigned to the dendrite’s synapses and compartments; different learning algorithms correspond to different ways of assigning this ‘blame’. The complexity of the functions pyramidal cell dendrites can learn will depend upon which particular learning algorithm they implement. The informationally poor Hebbian signal does not assign blame at all; it is not an error signal. This rule would place severe limitations on what functions dendrites could learn (Anderson, 1995; Shouval and Perrone, 1995; Elman *et al.*, 1996, p. 57; Phillips and Singer, 1997; Hanson, 1999). With simple error-correction learning, e.g. the delta rule or the perceptron rule, some of these limitations would be overcome. If the learning algorithm amounts to full-scale error-backpropagation, the dendrites would be able to learn complex non-linear functions.³

Strictly speaking, the basic principles underlying my neurosemantic proposal in sections 5 and 6 are independent of the power of the particular blame-assignment rule implementing the SINBAD idea. However, SINBAD networks are most plausible *qua* models of human cognition if the rule lies towards the more powerful end of the scale. (This is one reason why backprop has been used in SINBAD modelling studies; for further discussion, see Ryder and Favorov, 2001; Favorov *et al.*, 2003; and Favorov and Ryder, forthcoming.) As a result, most of my examples will presuppose some reasonably powerful form of error correction learning.⁴

Let us step back from the specifics of the mechanism and return to the overall dispositional profile of SINBAD cells. Whatever its precise nature, the dendritic matching mechanism turns out to have profound functional consequences. To simplify matters, let us consider a cell that has only two principal

³ Technically speaking, a Hebbian rule would prevent the dendrites from being able to separate non-orthogonal vectors; the delta rule with a linear activation function, while not being limited to orthogonality, does require linear independence. The perceptron rule (which assumes a non-linear activation function) can separate linearly dependent vectors, but can only compute a linearly separable function. Backprop can overcome both linear dependence and linear inseparability.

⁴ It is, in part, the capacity of the backprop version of a SINBAD network to learn complex functions in an unsupervised manner that distinguishes it from other kinds of neural networks. This capacity allows a SINBAD network to mirror the complex, non-linear regularity structure of the environment, going beyond the simple associative structure uncovered by other networks designed to extract regularities (Anderson, 1995).

dendrites. They are trying to contribute an equal amount, 50%, to the cell's output; that is they are trying to match each other's activities. And they are trying to do that consistently, no matter what inputs they happen to get. Suppose the cell's two principal dendrites are connected to the same detector, or sensory receptor. In this situation, it will be very easy for them to match. If they both just pass their input on to the cell body without manipulating it in any way, they will always match. That is, the two dendrites will have to compute the same function of their inputs (or mathematically equivalent functions).

However, dendrites do not get the same inputs, as a rule (Favorov and Kelly, 1996). Thus in the typical situation, the two dendrites' matching task will not be trivial. The learning rule forces each dendrite to try to match the other's outputs by modifying how it treats its inputs. That is, each dendrite must find the function of its inputs that best approximates the output of the other dendrite. They are both trying to do this at the same time, by (hopefully error guided) trial and error, until they manage to match activities *whenever* they get any input.

Let us imagine that the cell is involved in a processing an external event, and treat each dendrite's inputs as carrying information about some environmental condition. Suppose dendrite A receives information about property *a*, while dendrite B receives information about property *b*, i.e. one of A's inputs tends to be active in the presence of *a* and one of B's inputs tends to be active in the presence of *b*. On the basis of these inputs, under what conditions will the dendrites succeed in consistently matching each other's activities? If properties *a* and *b* are unrelated, such that sometimes they occur together and sometimes they do not, dendrite A will not be able to respond to *a* if and only if dendrite B responds to *b*. Suppose, for example, that *a* is the property of being red and rectangular, while *b* is the property of smelling like chocolate. There is no way for either dendrite to 'know' when to respond on the basis of these inputs and when not to, since there is no predictable relation between *a* and *b*. In order for dendrites A and B to learn to match their activities, it is necessary that there be some *correlation* (or anti-correlation⁵) between properties *a* and *b*.⁶ Such a correlation would obtain, for example, if *a* were the property of having a beak, and *b* the property of having feathers. The better the correlation between *a* and *b*, the better the matching that can be achieved. The ideal case: If properties *a* and *b* are perfectly correlated and the dendrites receive their *a* and *b* inputs from perfect detectors, the dendrites can learn to match all the time. Dendrite A can learn to respond when *a* is present and not otherwise by strengthening its *a* input and dropping any other inputs, dendrite B can learn to respond when *b* is present in a similar manner, and A and B will always respond together at the same level of

⁵ I will subsequently suppress the 'anti-correlation' for simplicity.

⁶ Mutual silence would be a solution to their matching problem. In fact, mutual silence would be a solution to *any* matching problem! So the mechanism must incorporate a bias against mutual silence.

activity.⁷ In the process, since A and B pass their activities on to the cell body, the cell will have become a detector of the conjunction of properties *a* and *b*. Of course, the ideal case will rarely happen; correlations and sensory apparatus are rarely perfect. But the cell can still become a pretty good detector of the conjunction of *a* and *b*.

Of course, there are more complex forms of mutual predictability than simple correlation, such as when there is a *multivariate* function of dendrite A's inputs that is correlated with a function of dendrite B's inputs. Call this sort of correlation 'complex correlation'. For example, suppose dendrite A receives information about *c* and *d*, while dendrite B receives information about *e* and *f*. Suppose further that there is no simple correlation among these properties, but rather a complex correlation between *c-xor-d* and *e-xor-f*. The dendrites will then settle on these exclusive-or functions of their inputs so that they can match activities. That is, dendrite A will come to respond (output = 1) only when *c-xor-d* is satisfied, and dendrite B will come to respond only when *e-xor-f* is satisfied; otherwise they will both be inactive (output = 0). Since the functions *c-xor-d* and *e-xor-f* are correlated in the environment, dendrites A and B will now always match their activities, and learning in this particular cell will cease.

You will notice that most reliable regularities are of this complex type. The simple correlation between beaks and feathers is the exception; for example, the property of being yellow and the property of being banana-shaped are not well correlated, since many things besides bananas are yellow, and many bananas are not yellow. Strong correlations usually exist not between pairs of properties, but between pairs of complex functions defined over properties. While being yellow and being banana-shaped are not well correlated, being [*yellow or green*] and *banana-shaped* is well correlated with being *peelable and coming in bunches*, and a range of other similar functions. In order to match consistently, each of the five to eight principal dendrites on a pyramidal cell will have to find one (or possibly several) from this range of multiple correlated functions.

Multiple environmental correlations, whether simple or complex, tend not to be accidental: they have a *source*, something that displays a set of correlated properties for some underlying *reason*. Assuming the correlation between properties *a* and *b* (in our first example) has a source, not only will the cell become a detector of the conjunction of *a* and *b*, as I noted previously, but it will also become a detector of the *source* of the correlation between *a* and *b*. In the case where *a* and *b* are having a beak and having feathers respectively, the source of the correlation between these two properties is a biological kind, *bird*. Similarly, in our second example, there will be a reason behind the correlation between *c-xor-d* and *e-xor-f*, and the cell as a

⁷ As described here, the two dendrites would not *both* tend to increase their outputs to the cell body. Rather, one would tend to increase and the other to decrease, eventually settling on an *average* of their initial contributions—an average which could be very small. To avoid this undesirable result, a term in the learning algorithm encourages a neuron to expand its dynamic range, allowing both dendrites to strengthen their outputs, and causing the neuron to respond strongly in conditions that allow its dendrites to match. See Ryder and Favorov (2001) and Favorov and Ryder (forthcoming).

whole will tend to become a detector of the source of *that* correlation (maybe bananas). In neuroscientific jargon, it will ‘tune’ to that source. In general, SINBAD cells will tend to tune to sources of both simple and complex correlations.

What, then, will they tune to? We saw in the previous section that individuals and all sorts of kinds were sources of correlation, so SINBAD cells, given the right inputs, will tune to individuals and kinds. Thus, for example, a SINBAD cell exposed to information about the surface properties of a natural kind will tend to respond in the presence of the kind’s hidden essence. This is because the cell’s dendrites will discover the correlations among the surface properties for which the essence is responsible. For example, one dendrite might come to respond to: [(clarity *and* solidity below zero) *or* (clarity *and* liquidity above zero)]; another might come to respond to: [tastelessness *and* colourlessness], and another might come to respond to [medium viscosity *and* solvent for table salt]. This is a cartoon example, but you get the idea—water’s essence is the reason behind these complex environmental correlations. When enough of the cell’s dendrites are active (and they will tend to be simultaneously active, indeed, they have learned to be, due to the correlations supported by water’s essence), the cell will reach threshold and fire. In effect, the cell will tune to the kind *water* itself, since only water has all these correlated properties (ignoring ‘twin’ cases, but only for the moment).

4. The development of a dynamic isomorphism

In the previous section, we saw how SINBAD cells capture information about elements of the deep structure of the environment, thus coming to respond to features that explain complex patterns of more manifest properties. This basic operation of a cell tuning to a source of correlation is the building block for the creation of a very general ‘predictive trick.’ When this tuning process takes place over an entire network, the network is transformed so that its flows of activation come to mirror regular variation in its containing organism’s environment. Where the environment has some important variable—a source of correlation—the network will have a cell that has tuned to that source of correlation. And where there is a predictive relation among sources of correlation, the network will be disposed to *mirror* that relation. The activities of the network’s cells will covary in just the way that their correspondents in the environment do. In short, the network becomes *dynamically isomorphic* to the environment.^{8,9}

⁸ The earliest extended physicalist discussion of the dynamic isomorphism idea, and a defense of its relation to the mind, occurs in Kenneth Craik’s *The Nature of Explanation* (1943). See also Cummins (1989) and McGinn (1989, Ch. 3).

⁹ SINBAD models are to be distinguished from self-organizing feature maps, as achieved by Kohonen networks (Ritter, 1995) and competitive learning (Intrator, 1995). These merely produce a map that is topographically organized by stimulus similarity, with more space devoted to more common stimuli. They are not useful for ‘prediction’ since they do not mirror, in their connections, the complex regularity structure of the environment.

The reason that a cortical SINBAD network develops into a dynamic isomorphism is that cells' inputs are not only sensory, but are also (in fact primarily) derived from within the cortical network. A cell's tuning is guided, in part, by these intracortical connections (Phillips and Singer, 1997). Tuning—changing a cell's dispositions to react to the environment—occurs through the modification of a cell's dispositions to react to activity in other cells, mediated by intracortical connections. It is these latter dispositions that come to mirror environmental regularities.

In the feedforward direction, i.e. up the cortical hierarchy from sensory toward higher areas, intracortical connections allow the cells to uncover higher and higher order correlations and their sources. For example, early visual cells (in V1) tune to edges, which are sources of correlation responsible for regularities in raw input data. These cells' outputs will be the following layer's inputs, so cells in this next layer will tune to sources of correlation among edges, e.g. depth (Becker and Hinton, 1992) and types of translational motion. The subsequent layer may begin to tap into movable shapes (which are sources of correlations involving motion and edge depths), and so on (Favorov and Ryder, forthcoming; Favorov *et al.*, 2003). First there will be a tendency to uncover sources of correlation that participate in regularities at progressively larger spatial and temporal scales; then more abstract sources of correlation will be discovered, such as kinds.¹⁰ As this hierarchical tuning occurs, the interlevel connections that mediate it will naturally come to reflect how the edges, depths, shapes, kinds etc. are related.

However, it is not only hierarchical regularities that come to be reflected, since cell tuning is guided not only by inputs from previous layers in the hierarchy. Sitting on top of this hierarchy are complex lateral and feedback projections (Felleman and van Essen, 1991; Lerner *et al.*, 2001). These connections are able to capture regularities within and across levels of abstraction in the same way that regularities come to be mirrored in the feedforward direction. Sources of correlation are related to each other in many different ways. Cats are related to fur and to mice, water is related to taps and salt, and grass is related to greenness and to suburbia. The lion's share of the regularities in which a source of correlation participates is with other sources of correlation, and any number of these regularities may be used by the dendrites of a cell as they improve their ability to match. A cell may start with a tenuous correlational seed,¹¹ but this subtle sign of those correlations' source is enough to put the cell on a path towards discovering the multitude of regularities in which that source participates. As the cell achieves

¹⁰ Early in the processing stream, SINBAD cells will tune to sources of correlation that depend heavily upon the nature of an organism's receptors, but this parochialism can be overcome as one proceeds up the cortical hierarchy.

¹¹ If there is no correlation available, the cell's activity will be low, and it will elaborate its dendrites in search of new inputs until it is able to find a correlation (for a review of dendritic growth, which occurs throughout life, see Quartz and Sejnowski, 1997). If it is still unsuccessful, at some point the cell will 'give up' and degenerate (Edelman, 1987).

more and more robust dendritic matching, the correlational seed ends up producing a complex dendritic tree, which realizes complex functions relating that source of correlation to many others (or rather cells that have tuned to them).

Thus in tuning to a source of correlation, the dendrites of a particular pyramidal cell find functions that relate that source of correlation, not only to sensory inputs, but also to other sources of correlation via intracortical connections. In this way, the connections among cells gain characteristics that dispose flows of activity to mirror regularities involving these sources.¹² This result constitutes a good ‘predictive trick’. When a SINBAD cell is activated, this amounts to the network ‘inferring’ the presence of a particular source of correlation, both directly from sensory input, and indirectly from other cells that are active due to the presence of the source of correlation to which *they* have tuned. A cell that has tuned to a particular variable has *multiple* sources from which it can obtain information about that variable, from numerous sensory input channels and also other SINBAD cells that provide it with input. If one of those sources of information is blocked, e.g. sensory inputs, the others will compensate.¹³ In the context of the networks’ dynamic isomorphism to the environment, a cell that corresponds to the kind *tiger* (because that is what it has tuned to) will light up when all that is seen is a twitching tail, or even a footprint. That is, intracortical connections allow the network to perform the ‘predictive trick’ of ‘filling in missing information.’¹⁴

While it is a perceiver’s individual history of experience that makes a network develop its internal organization, one of the main benefits of a SINBAD network is that, despite the variability in possible histories of experience, and the idiosyncratic design of sensory receptors (see footnote 10), it still develops an isomorphism to objective regularities among objective entities. By capitalizing on a large number of *different* idiosyncratic routes to it, a cell tunes to something objective: a source of correlation, like an individual or kind. Further, the cells in the network covary their activities in accordance with the objective regularities obtaining among these individuals and kinds, again via a large number of different and potentially idiosyncratic routes.

Here then, in brief summary, is how SINBAD networks operate. The multiple dendrites on a SINBAD cell must find functions of their inputs that are correlated. Assuming these correlations are not accidental, the cell will tune to their source. In

¹² If you are worried that there are far too many sources of correlation our brains need to have some cells tune to, consider the fact that in the densely interconnected human cerebral cortex, there are somewhere between 11 and 25 billion pyramidal cells (Pakkenberg and Gundersen, 1997). Compare this to a good adult vocabulary of 50,000 words. (There is also a mechanism to prevent too many cells from tuning to the *same* source of correlation—see Ryder and Favorov, 2001; Favorov and Ryder, forthcoming.)

¹³ Of course, this will create a mismatch between dendrites; if a previously correlated input is *consistently* absent, the dendrite will learn to ignore it in order to achieve a match again with the other dendrites on the cell.

¹⁴ For a computational treatment of a different sort of example—where a SINBAD network makes explicit the hidden structure in a dynamic system (a kitchen sink)—see Favorov and Ryder (forthcoming).

tuning to a source of correlation, a cell will provide other cells with a useful input, i.e. an input that helps *their* dendrites to find correlated functions. Thus these further cells, in turn, tune to sources of correlation, and the process repeats. The end result of this complex multiple participant balancing act is that a SINBAD network comes to be dynamically isomorphic to the environment from which it receives inputs. It is a dynamic isomorphism that mirrors the deep structure of the environment, with elements that correspond not only to sensory features, but also to the individuals and kinds around which environmental regularities are structured. This dynamic isomorphism is a useful thing to have, because it yields reliable ‘predictions’. The fundamental SINBAD mechanism explains *why* networks turn themselves into dynamic isomorphisms. In the end, it is the tendency of single pyramidal cells’ dendrites to match by discovering complex correlations that explains the cells’ (and their containing network’s) ability to yield reliable ‘predictions’.

5. Model representation

You will have noticed that I have not yet said anything about representation. All I have demonstrated thus far is that SINBAD networks develop into structures that are dynamically isomorphic to environmental regularities involving sources of correlation, and that this is a useful ‘predictive trick’. However, you will have no doubt noticed an analogy between SINBAD networks and some familiar representational artifacts that also capitalize on isomorphism, namely *models* (especially dynamic models). One of the main uses of a model is in prediction. Consider a model of the solar system, for example. Each model planet is attached to an arm that extends from the centre hub, the ‘sun’. Inside the hub are a number of gears and pulleys that make the model planets rotate in just the way the real planets do. Suppose we do not know where the planet Mars will be next July. In order to find out, we could rotate the model Earth into the position that corresponds to where *it* will be next July, and read off the position of Mars from the model. The gears in the hub ensure that moving the model Earth into position makes the model Mars move into the correct position. The gears—analogueous to intracortical connections—maintain the dynamic isomorphism between the model and the solar system, allowing us to ‘fill in’ an unknown variable (the position of Mars) from a known variable (the position of the Earth). Just as in the SINBAD network, it is this ‘filling in’ capability of models that underlies their predictive utility.

Such models are dynamic isomorphisms that are representational: a model is a model *of* something. Perhaps this analogy is enough to justify attributing representational powers to a SINBAD network? No it is not. Isomorphism, by itself, is not sufficient for representation. Indeed, a model of the solar system may be isomorphic to many things that it does not thereby represent (perhaps an Indonesian weather system). Similarly, a natural outcropping of rock that happens to be structurally isomorphic to the Empire State Building does not represent it, or

anything else. Neither is isomorphism necessary for model representation; a model may represent something to which it is only very weakly isomorphic—consider a typical child's toy, for example. Finally, isomorphism is a symmetric relation, but representation is asymmetric. The model of the solar system represents the solar system, but not vice versa!

So nothing, including a SINBAD network, represents anything simply in virtue of an isomorphism. What *does* endow a model with its representational powers, then? One might worry that the only answer we are likely to find is a singularly uninformative one: a model represents *x* because its designers and/or users intend that it represent *x*. True, but not very helpful. All is not lost however. One way of understanding the general strategy known as teleosemantics¹⁵ is as a way of explaining *how* the designer or user of an artifact can endow it with representational powers. The feat is accomplished by endowing the artifact with certain special normative properties, particular *functions*, now in the teleological rather than the mathematical sense. And as it turns out, it is possible to have functions in this sense *without* designers or users.

We can take our cue here from Fred Dretske (1988; 1995). Dretske notes that a speedometer, for instance, represents speed. It also indicates, or carries information about speed; what it represents, then, seems to have something to do with what it indicates. However, just as a model may be isomorphic to things other than what it models, a speedometer may indicate features other than speed, e.g. tire size. Why does it not represent these other features? That is because it is not *supposed* to indicate them. It does not have the job or *function* of indicating tire size, it only has the function of indicating speed. It is by giving this *functional* property to the speedometer that its designers and/or users make it truly representational. To represent, for gauges and meters and such things, is to have the function of indicating. (To avoid confusion, I will henceforth disambiguate the term 'function' by using 'teleofunction' for the normative sense, and 'function' for the mathematical sense—though I always mean 'biological function' in the teleological sense. Frequently, I will find it convenient to talk in terms of having a 'purpose' or 'job', and 'being supposed to'; by these I do not intend anything substantially different from teleofunction.)

Let us apply Dretske's form of reasoning to models.¹⁶ In this case, we start with isomorphism rather than indication. We note that modelling seems to have

¹⁵ See Millikan (1984; 1989), Dretske (1988; 1995), Papineau (1987; 1993) and especially McGinn (1989), who explicitly (and independently) endorses this account of representation in models (p. 189).

¹⁶ One might wonder why I do not just go with Dretske's idea and say that SINBAD cells represent because their purpose is to *indicate* sources of correlation. The reason is that the 'predictive' trick of filling in can occur in other functional modes that have nothing to do with indication. A dynamic isomorphism/internal model is useful not only to fill in missing input information, but also to explore hypothetical possibilities, for instance. When imagining the future and choosing the best course, cells perform all sorts of fillings in that are not indications—but they are nonetheless doing their jobs well (Ryder, 2002). Thus my theory has more in common with Millikan's (1984; 1989) than with Dretske's. Millikan's intentional icons are supposed to 'map' onto the environment rather than to indicate.

something to do with isomorphism, but that a model may be isomorphic to things it does not represent. This may be accounted for if we suppose that representation is a teleofunctional property. The reason the model of the solar system represents the solar system and not an Indonesian weather pattern is because the solar system is the thing to which it is *supposed* to be isomorphic. That is (at least part of) its job. That is why the system of balls, arms, and gears is said to *model* the solar system. At the same time, a particular ball in the model is *supposed to correspond*, in the context of the isomorphism, to a particular planet, e.g. the Earth. We sum this up by saying it 'stands in for' the Earth. Models and elements of models both represent: models are said to 'model' structure S in virtue of having the teleofunction of being isomorphic to S, whereas elements of models are said to 'stand in for' x in virtue of having the teleofunction of corresponding to x in the context of an isomorphism (Ryder, 2002).¹⁷

This teleofunctional understanding of representation in models also allows us to understand how a model may *fail* to be isomorphic to what it represents. Our model may represent the solar system, but do so *inaccurately* if the isomorphism it is *supposed* to have is imperfect. There are no obvious limits to how inaccurate a model may be, to how badly it may represent. Finally, the asymmetry of the representational relation also stands explained, since having the teleofunction of being isomorphic to S is an asymmetric relation. (In fact, it is probably not a relation at all, since it is *prima facie* possible to have the job or purpose of being isomorphic to something that does not exist, e.g. Excalibur.)

It is plausible, then, that modelling is normative isomorphism: to represent in the way that models do is to have the job of being isomorphic to something. The upshot is that if we want to show that cortical SINBAD networks *represent* sources of correlation and the regularities they enter into, we must demonstrate, not that they are *isomorphic* to these environmental regularities, but that they are *supposed* to be so isomorphic. This is what I intend to do in the next section.

6. A neurosemantics

6.1 General representational teleofunctions

Supposing that there are SINBAD networks in the cerebral cortex, their teleofunctions are clearly not endowed by any intentional design or use. Cortical networks have *biological* functions, and I follow numerous others in maintaining that biological function is not dependent upon intentionality. In particular, an etiological account of teleofunction cashed out in terms of evolutionary history

¹⁷ Some people hear the expression 'having the (teleo)function of *being*...' as linguistically anomalous, as opposed to 'having the (teleo)function of *doing*...'. If this applies to you, just substitute 'is supposed to be' for the jarring expression, and everything will be fine. What matters is the normativity, not the teleofunctionality per se.

delivers us normative isomorphism, and thus representation, in cortical SINBAD networks.¹⁸

On etiological accounts, the teleofunction of a thing explains its presence in the system of which it is a part. Thus it is the job of a speedometer to indicate speed because that is why it is in the car, i.e. that is why we put it there; we do not install speedometers because of their capacity to indicate tire size. Similarly, it is the (biological) teleofunction of the heart to pump blood because that is the (evolutionary) reason behind its presence. It is because of the capacity of the heart to pump blood that it has proliferated in the biological world, not because of its capacity to make thumping noises.¹⁹

If evolution has endowed the cortex with SINBAD characteristics, it is highly plausible that this is an adaptation whose selectionally relevant feature is the 'predictive' trick it provides—the presence of SINBAD in the brain would then be explained by this feature. But the only reason SINBAD networks yield reliable 'predictions' is that they come to be isomorphic to environmental regularities. Our hypothesis, then, is that SINBAD networks are present in the brain because they come to be isomorphic to environmental regularities; consequently, this tendency would number among such a network's jobs. By our analysis of representation in models, it follows that SINBAD cortical networks would represent. They would represent what they are supposed to be isomorphic to, namely environmental regularities.

In fact, we can be more specific about what SINBAD cortical networks would represent: they would represent environmental regularities involving *sources of correlation*. This is because, plausibly, the evolutionary success of SINBAD networks as 'predictive' tools would depend upon the way they interact *specifically* with sources of correlation. There are at least two reasons for this.²⁰ First, we saw that the clustering of numerous (possibly complex) properties around a source of correlation allows a cell that tunes to that source to have multiple lines of

¹⁸ In order to limit the number of issues I address in this paper, I have adopted without argument my preferred account of teleofunction, namely an etiological one. However, the theory I present is likely to be consistent with alternative accounts, if they prove viable. In particular, some may be attracted to a non-historical account of SINBAD teleofunction (Cummins, 1975; Bigelow and Pargetter, 1987; Schroeder, 1998; McLaughlin, 2001), since this would allow Swampman (Boorse, 1976; Davidson, 1987) to have full-fledged representations. If any teleofunction has a chance of being explicated non-historically, those of SINBAD cells do. This is because a SINBAD network is intuitively a very 'general purpose' device, serving not one *particular* need of an organism, but its needs in *general*.

¹⁹ I am glossing over many subtleties here. In fact, there is no agreed-upon formulation of the etiological idea, originally due to Wright (1973), or of its evolutionary implementation; for discussion see (Millikan, 1984; Allen *et al.*, 1998; Ariew *et al.*, 2002). Since the SINBAD mechanism does not stray near any hotly disputed territory, I think I am safe in assuming that the correct etiological theory will dictate that a SINBAD cortex would have the teleofunctions I attribute to it.

²⁰ These correspond to the related induction and learning teleofunctions that Millikan attributes to 'substance concepts' in her (1998).

'evidence' for its presence. The result is an extremely powerful 'predictive' network, with multipotent capabilities for filling-in. This special feature depends upon the inductive richness of sources of correlation; so if cortical pyramidal cells are SINBAD cells, they are plausibly built (by evolution) to take advantage of sources of correlation.

SINBAD cells take advantage of the inductive richness of sources of correlation not only directly for 'predictive' purposes, but also to facilitate future learning. This is the second reason why the utility of the SINBAD mechanism depends specifically upon its interaction with sources of correlation. Because sources of correlation are inductively rich, once a SINBAD cell *starts* to tune to one by discovering some of the correlated properties it exhibits (the correlational seed), the cell is in a uniquely advantageous position to discover *further* correlation. (This will be the case as long as its dendrites have not come to match their activities perfectly, which they almost never will, due to the presence of noise in the cortical network.) We saw that in this way, a cell continually adds to its lines of 'evidence' for the presence of the source of correlation to which it is tuning, enriching the model's isomorphism.²¹ Again, it seems likely that a cortical SINBAD network would have been selected, not for a capacity to be isomorphic to generic environmental regularities, but to be isomorphic to environmental regularities specifically involving real kinds, individuals, and other sources of correlation. It follows that it is a teleofunction of SINBAD networks to mirror regularities involving sources of correlation, and so these are what SINBAD networks represent or model.

6.2 Specific contents

If SINBAD networks represent, then they must also have the capacity to misrepresent. Along with the representation of the non-existent, this is a hallmark of intentionality, and a crucial test for any theory of mental representation (Dretske, 1986; Fodor, 1987). There are actually several forms of representational error, but the usual place to start is with perceptual misrepresentation, i.e. the perceptual application of a representation to some object or property that is not included in that representation's extension. The question is: how is it possible to misperceive a cat as a dog? It should be pretty clear that we are not yet in a position to formulate conditions for this type of SINBAD misrepresentation. There are two reasons. First, we have not said anything about what it is to *apply* a SINBAD representation to something. Second, although we have seen what SINBAD networks model *in general*, we are not yet in a position to state what determines a *particular* SINBAD

²¹ A pyramidal cell is not necessarily restricted to 5 to 8 lines of 'evidence' from its 5 to 8 principal dendrites. First, multiple functions may overlap on a single dendrite. Second, recent evidence raises the possibility that a cell's separate investigators may not be the principal dendrites, but rather the much larger number of terminal branches (Wei *et al.*, 2001). Third, cells can act in populations, with individual cells in that population tapping into different lines of 'evidence' for the presence of the same source of correlation.

representation's extension. In the present paper, I give an account of particular extensions. I address the matter of perceptual application elsewhere;²² here we can get by with the crude (but usual) assumption that a perceptual representation is applied to the object or event that occasions it (Fodor, 1987; Dretske, 1988).

Models are composed of elements that enter into relations. In the solar system model, the elements are balls, and they enter into spatial and dynamic relations; in a SINBAD network, the elements are cells, and they enter into dynamic relations. A complete analysis of a model's extension must be given in terms of these elements and relations. We have already seen that the SINBAD dynamic relations are supposed to mirror environmental regularities involving sources of correlation. But environmental regularities among *which* sources of correlation, exactly? To answer this question, we must be able to determine what *particular* cells stand in for, i.e. what they are supposed to correspond to in the context of the model's isomorphism.

The difficulty is that for a general learning device such as SINBAD, the etiological account of representational status and the psychosemantics (the specification of content for particular representations) appear to come apart. For a hardwired internal model that has evolved to be isomorphic to a *specific* environmental system, it will be evident which environmental entity a particular cell represents, i.e. is supposed to correspond to.²³ (Though see the remarks in section 7 on indeterminacy.) By contrast, for a SINBAD cell, though we can point to its evolutionary history in order to justify the claim that it is supposed to correspond to and thus represents a source of correlation, there is no *particular* source of correlation its evolutionary history can deliver us as the one it represents. This is where we find a crucial philosophical role for the details of the SINBAD mechanism. These details yield determinate contents for particular SINBAD cells.

Due to its SINBAD mechanism, a cell tends toward achieving an equilibrium state, a state in which its dendrites successfully match one another's activities. The progress towards this state in a SINBAD network can be seen in figure 2, a plot of how well correlated the activities of each cell's dendrites are as a function of time. As you can see, most of the cells achieve near perfect equilibrium. It is when they are in this state that we look to see what the network mirrors, and what the activities of particular cells correspond to.

As we saw above, the dendritic correlations in an equilibrated cell will have an explanation in the environment, some source of correlation. Suppose that, after learning, the explanation for the dendritic correlations in cell SB lies in kind K—K is what I call the 'explanatory source' for this post-learning cell. Now, evolution

²² With reference to Ramsey's vision of beliefs as 'maps by which we steer' (Ramsey, 1931), the subject of perceptual application belongs in an analysis of steering (Ryder, 2002). Here I am giving an account of the maps.

²³ Such hardwired representations will rarely be *mental*, in my view. For example, fixed action patterns, even if they make use of an internal model, are still 'mindless behaviours.'

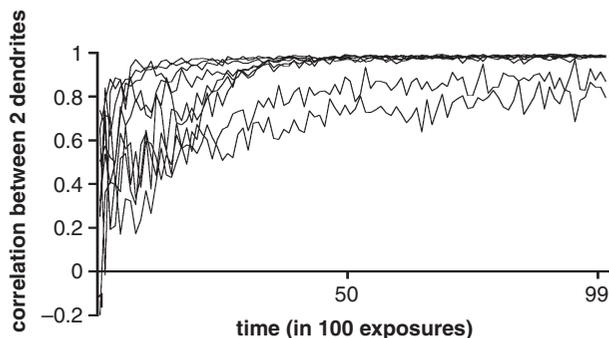


Figure 2 Learning progress in a nine cell network exposed to the surface variables of sinks (see Favorov and Ryder, forthcoming). Each line traces the degree of correlation (by correlation coefficient) between the activities of a single cell's pair of dendrites.

did not design SB to correspond to K. Nevertheless, we are justified in asserting that SB is supposed to correspond to K. This is because the cell, *qua* SINBAD cell, is supposed to yield reliable 'predictions', and its corresponding specifically to kind K is the only reason the cell yields reliable 'predictions'. SB performs its more general job only by corresponding to kind K, so it has the derivative job of corresponding to kind K. (So it represents K.) Compare: the heart fulfills its general job of pumping blood only by rhythmically contracting its ventricles, so it has the derivative job of rhythmically contracting its ventricles.

Let us consider more carefully the way in which a particular source of correlation explains a SINBAD cell's fulfilling its predictive teleofunction. In essence, SINBAD is a learning mechanism, so the way in which a particular source of correlation explains how a cell is able to yield reliable 'predictions' is *historical*. For kind K to be SB's explanatory source is for K to explain how SB was able to *acquire* its 'predictive' abilities. It is the purpose of the cell's dendrites, *qua* SINBAD cell dendrites, to *find* correlated functions of their inputs. These must be *reliably*, i.e. non-accidentally correlated; in other words, there is supposed to be a single source of correlation that causes the dendrites to achieve matching. This is how SB was designed (by evolution) to fulfill its teleofunction of yielding reliable 'predictions'. So in finding the non-accidental correlations whose source was K, SB's dendrites were doing their job. Thus the cell's *coming to* correspond specifically to K was the cell doing what it was designed (by mother nature) to do. Since SB's *coming to* correspond to K was for it to do its job, it is reasonable to conclude that its job is to correspond to K. Since it is supposed to correspond to kind K, the cell represents (stands in for) K. It represents K even though evolution did not design it to correspond to K; it acquired that job through learning.

So we have reason to say that a perfectly equilibrated, post-learning cell represents its explanatory source, i.e. the source of correlation that explains its attainment of equilibrium and 'predictive' value. But imperfectly equilibrated cells (like all of the cells in figure 2) *also* have explanatory sources, whether they are in the middle of a learning trajectory, or have stabilized. Even when a SINBAD cell's equilibrium is not quite perfect, and so its 'predictions' are sometimes inaccurate (in the sense that its containing network is imperfectly isomorphic to environmental regularities), in general there will still be a specific source of correlation that explains the *degree* to which its dendrites were able to acquire matching mathematical functions of their inputs.²⁴ Any SINBAD cell has fulfilled its 'predictive' teleofunction to the extent its dendrites have learned to match through interaction with a particular source of correlation. Thus an imperfectly equilibrated cell—call it SBI—has done its job to the extent that its dendrites have learned to match through interaction with its explanatory source, say kind K, or what amounts to the same thing, to the extent that it has learned to correspond to K. Again, we infer that its job is to correspond to kind K, and that it represents K.

Since SBI's equilibrium is imperfect, there may be many other things it happens to respond to that do not explain its acquisition of the 'predictive' abilities it has. When the cell fires, it does so because all or most of its dendrites are simultaneously highly active. Usually this is to be explained by the fact that the dendrites have learned to respond to some source of correlation. But a particular instance of matching like this can sometimes be *accidental*, i.e. not explained by the source of correlation, kind K, to which the cell has tuned. Perhaps the organism's sensory apparatus fails, or perhaps just enough of the features that characterize kind K happen to be present on the occasion. Then it is just by luck (as opposed to design) that SBI's dendrites predominantly match; there is no sense in which the cell is *supposed* to respond in this situation. It is only supposed to respond (and correspond) to kind K.

Note that this result is consistent with SBI having responded to things other than K while it was learning. These other things do not explain its achievement of its 'predictive' abilities if they are not sources of the environmental correlations that guided the cell towards its current level equilibrium. Something may have caused just enough of the dendrites to respond in order to make the cell fire, but overall there was a failure to match; the SINBAD learning rule tried to *discourage* a response to this sort of thing, and the cell temporarily veered *away* from equilibrium. Further encounters with kind K then got it back on track towards its current level of 'predictive' utility. Thus, if these things that are not kind K but nevertheless caused firing during the learning process explain anything, they explain why the cell has *failed* to achieve perfect equilibrium and 'predictive'

²⁴ There may be various reasons why the cell could not achieve perfect equilibrium: noise in its input channels, variation within the kind it has come to detect (like frogs with three legs), or simply because the cell does not receive inputs carrying the information it needs to make its dendrites correlate perfectly.

reliability. The cell clearly does not have a derivative job to correspond to these other things; it only has a derivative job to correspond to its explanatory source, kind K.²⁵

Thus our account of particular representational contents also furnishes an account of perceptual misrepresentation. A particular SINBAD cell, whether perfectly or imperfectly equilibrated, represents its explanatory source, i.e. the source of correlation that explains how its dendrites have come to match their activities. Since a cell can respond to something *other* than its explanatory source, it can participate in an event of perceptual misrepresentation. As we saw, it might respond to something that has a subset of its explanatory source's correlated features. Suppose there is a cell whose dendritic correlations, and thus predictive utility, are to be explained by its previous exposure to pennies. Suppose this cell responds to a dime in my pocket. (Enough of the penny stand-in's dendrites have high outputs that they make the cell fire in response to the dime. The dime feels like a penny, or perhaps the presence of a penny in my pocket coheres with my internal model of the current situation—I mistakenly remember receiving one in change.) Because it is *pennies*, and not *dimes*, that explain the cell's achievement of (imperfect) equilibrium and its 'predictive' abilities, its job is to correspond to pennies, not dimes. Therefore it represents pennies and not dimes. So when it fires in response to a dime, it (or, properly speaking, the organism) misrepresents the dime as a penny.

I have been assuming that a SINBAD cell always has a single explanatory source. This is an idealization; sometimes we would expect there to be several sources of correlation that are legitimate, partial explanations for a cell's dendrites acquiring matching functions. This would occur, for instance, if two distinct kinds shared all (or most) of the surface properties to which an organism had direct or indirect perceptual access. The organism could not (easily) distinguish between two such kinds, and one of its SINBAD cells might discover multiple correlations and have 'predictive' success due to interaction with *both* kinds. It would then correspond, to some degree, and by design, to both. In this case, we should say that the cell represents *equivocally*. It presents two distinct kinds as one; it is 'confusing' the two, and doing its job less than ideally. (It would not be right to say the cell has a *disjunctive* content, because there is no representational separation of the disjuncts; see also Millikan, 1998.)

So not only are SINBAD cells supposed to correspond to sources of correlation in the context of an isomorphism, the details of the SINBAD mechanism allow us to determine exactly *which* source(s) of correlation a particular SINBAD cell is

²⁵ Note that this avoids a problem that arose for Dretske's (now abandoned) account of representation in *Knowledge and the Flow of Information* (Dretske, 1981). In this book, Dretske identifies the content of an indicator with the information that was instrumental in causing it to develop a particular sensitivity during its 'learning period'. The problem is that Dretske gave no principled way to distinguish between the 'learning' period and the 'use' period (Loewer, 1987).

supposed to correspond to. Since SINBAD cells are the basic elements of a SINBAD network, we can also determine which regularity structure a whole network is supposed to be isomorphic to. We can be confident, then, that SINBAD cells and networks represent; cells stand in for sources of correlation, and networks model regularities structured around sources of correlation. If cortical pyramidal cells are SINBAD cells, then the principal type of cortical representation is SINBAD representation.²⁶ Since the cerebral cortex is the seat of the mind, this gives us some reason to believe that SINBAD representation realizes *mental* representation in us, and other creatures with a cerebral cortex. This conclusion can be bolstered by showing in addition that SINBAD representation exhibits the features characteristic of mental representation, as described by both scientific and folk psychology. We have already seen the most important of these: the capacity to misrepresent. In the next section, I will briefly outline how SINBAD neurosemantics explains a few other features; for more, see Ryder (2002).

7. Further applications

Twin cases: To explain the cell's ability to yield reliable 'predictions', an explanatory source must feature in the causal history of the cell's dendrites acquiring their matching functions. It follows that SINBAD content is broad. A cell—call it 'Sinbad'—that has acquired its 'predictive' abilities through exposure to water will not be equivocal between water (H₂O) and the superficially identical twin-water (XYZ) on twin-Earth (Putnam, 1975), because twin-water is not an explanatory source for it. Similarly, Twinbad (on twin-Earth) refers to twin-water, not water. These representational facts will hold even if Sinbad and Twinbad are inside Oscar and Toscar, who are molecular duplicates. This coheres with the view, dominant in the philosophy of mind,²⁷ that mental content is broad.

Distal reference: According to SINBAD neurosemantics, a cell's representational content explains its tuning. But explanation is a slippery thing. For example, suppose that the kind *horse* explains a cell's tuning in the manner described previously. Horses will have interacted with the cell via a huge number of more proximal causes, like events in the retina. Since these proximal causes also explain the cell's tuning to some degree, should they not also form at least part of the cell's representational content? SINBAD neurosemantics could not then be a theory of

²⁶ One might expect that other kinds of cells in the cortex might implement a different sort of representation, but there is reason to suspect these other, non-pyramidal cells (whose effects are purely local) are ancillary to the basic SINBAD mechanism (see Favorov and Ryder, forthcoming; Favorov *et al.*, 2003).

²⁷ Some advocates include (Burge, 1979; Millikan, 1984; Lycan, 1988; McGinn, 1989; Stalnaker, 1989; Devitt, 1990; Bilgrami, 1992; Dretske, 1993; Baker, 1995; Perry, 2001). For dissent, see (Searle, 1983; Segal, 2000).

mental content. Our thoughts of horses are not somehow contaminated with thoughts of rods and cones.

The first thing to note in response is that the proximal and distal causes are not causally *independent* explanations for a cell's tuning, because tuning to the distal cause occurs *via* the proximal cause. By contrast, in a case of genuine equivocal (sometimes called disjunctive) representation, the two sources of correlation must be *independent* explanations for the cell's tuning, e.g. elms and beeches. Given that horses and retinal patterns are not independent, we must ask which candidate *best* explains how that cell was able to achieve dendritic matching.

Here, there is no contest. Each individual type of retinal pattern²⁸ explains only to a very tiny degree, if at all, the extent to which the cell has achieved dendritic matching. Even adding up all of these explanations does not amount to much. This is both because each pattern was encountered only a very small number of times (perhaps only once?), and because it supports only very weak correlations. In order to tune to and correspond to the kind *horse*, a cell must rely upon correlations among environmental features that are reflected, not in single pattern types, but only indirectly in complex functions of large numbers of different retinal patterns (not to mention all the other sensory modalities). It is only via this more complicated route that horses' 3D shapes, tendency to gallop, and other properties can influence tuning. Each dendrite must discover some complex feature (in the form of a mathematical function) that *is in common* among a multitude of receptor patterns, encountered at different times. The correlation of these complex features is what guides this discovery, and only the kind *horse* grounds *that* correlation. So for a cell that corresponds well to horses, only the kind *horse* can fully explain how it managed to achieve dendritic matching.

The qua problem: Suppose multiple encounters with horses cause a SINBAD cell to acquire matching dendritic functions—is it a horse stand-in or an animal stand-in? The things that explain its achievement of equilibrium are members of both kinds. This is an instance of the qua problem (Devitt, 1981): when one kind subsumes another kind, what determines which kind is represented?

There will normally be a fact of the matter which kind explains how a cell has acquired its predictive abilities. The kind *horse* and the kind *animal* are sources of different sets of multiple correlations that have different underlying (evolutionary) explanations—that is why they are distinct kinds. Horses tend to neigh, are usually domesticated, have a particular shape, particular eating habits, hooves, manes, etc. Animals are characterized by a more abstract set of correlated features with a more ancient evolutionary explanation for their coherence: the capacity for spontaneous motion, a range of sizes, a disjunction of typical methods of

²⁸ Note that the set or collection of receptor patterns is not even a candidate, since it is not a source of correlation. I might add that the only reason a *single* type of pattern could count as a source of correlation is that it may have a consistent cause (e.g. horses), and so may have a recurring syndrome of correlated features due to the influence of that cause.

locomotion, a range of typical colours, and so on. When a cell's representational content is determinate, which will be the usual case in this example, its dendritic correlations will be explained by (a part of) one of those sets of correlated properties rather than the other. It will either be the properties whose correlation typifies horses, or the properties whose correlation typifies animals that will have historically guided the cell to equilibrium by causing synaptic activity. This will be true even if all the individuals that have been encountered are in fact members of both kinds.

Frege cases: As we just saw, which properties were causally responsible for a cell's tuning are relevant to determining that cell's content, because they are relevant to determining which *kind* it was that guided the cell's tuning. However, in order for a cell to represent a specific kind, its dendrites need not have discovered any *particular* set of correlated properties for which that kind is a source. All that is required for two cells to have the same content is for the (possibly differing) correlations they discover to have the same source. It follows that different cells in different organisms having radically different perceptual mechanisms can represent the same thing, as can a single cell across time as it learns more about the source of correlation it is supposed to correspond to. This is a good thing, if SINBAD neurosemantics is ultimately to ground propositional attitude ascription (Fodor, 1998; Millikan, 1998).

However, sometimes we need to draw attention to two different *ways* we have of representing the same source of correlation, e.g. Superman. In a SINBAD network, this would occur when two cells represent Superman, but via different sets of multiple correlations: one characterizing him in his guise as a super-hero, the other in his guise as a newspaper reporter. Two distinct representational vehicles will, in this case, have the same content. Since they are distinct vehicles, however, they may play different psychological (e.g. inferential) roles.²⁹ Connecting up these psychological roles, e.g. by associating these two vehicles, is thus a psychologically significant event. This is how SINBAD explains the cognitive significance of learning (as we often put it) that Clark Kent is Superman.

In SINBAD networks, perfect correlation will encourage strong association. If two cells are so strongly associated that they come to be always active simultaneously, then their activity will cause and be caused by the very same neural firing patterns. Thus for the purposes of psychological explanation, it would be appropriate to consider them to be a single representation, joining together a larger set of correlated features. (This would be a type of 'population code', which I would

²⁹ It may be that the phenomenon of 'seeing as' and what psychologists study under the label of 'categorization' involve these psychological roles. If so, these phenomena are likely to exhibit much individual variation, especially for referents that have a large number of alternative ways in which they can impact an observer (i.e. for most referents beyond those of early sensory representations). Cells in different observers will likely learn to triangulate upon the referent via somewhat different routes.

expect to be the norm, rather than ‘grandmother cells.’) When two cells represent the same thing, their contents will obviously be ‘perfectly correlated’, so the SINBAD mechanism will encourage them to merge into a single representation. This beneficial tendency to merge will usually fail only in cases where circumstances conspire to perceptually isolate two sets of correlated features from each other, so that some members of the two sets are rarely seen to intermingle. Such is the case with Superman/Clark Kent and the Morning Star/Evening Star.

Fictions and other representata non grata: How can a cell refer to something that does not exist? To answer this sort of question, I would normally say that such representations are composed, and my neurosemantics is only meant to apply directly to atomic representations. However, in some cases, e.g. Unicorns and Santa Claus, I think there is a better answer, to be found in the notion of a ‘fictional kind’. This is a kind whose correlated properties (exhibited in pictures, models, movies, novels etc.) are unified by a population’s intentions. A SINBAD cell could represent a fictional kind just as well as any other kind.

The source of correlation schema is also a suitable way to treat other things that are members of mysterious metaphysical categories, like words, Beethoven’s Fifth, and the polka. A SINBAD cell may also tune to geometrical shapes because of their regular pattern of angles and edges, and the resultant position independent symmetries (i.e. correlations) they display. The material a square is made of (sometimes just light, as in a hologram) is irrelevant to the cell’s ability to make its dendrites correlate. It is the *squareness* that explains the cell’s achievement of equilibrium. (This is not a purely causal explanation, but of course not all explanation *is* purely causal.) Combined with an adequate structuralism in mathematics (Benacerraf, 1965; Resnik, 2000), this suggests a way in which the representation of numbers may be accommodated.

Millikan’s (1998; 2000) ‘substances’ are very similar to what I am calling sources of correlation. She also believes that they span many metaphysical categories, though perhaps not quite as many as I have suggested. Her account of substance concepts is, to my knowledge, the closest cousin to my neurosemantics in the current literature. In fact, I am greatly encouraged by the partial resemblance, especially since the convergence arises from opposite directions: she comes from abstract psychological considerations, while I come from the biology. This sort of concordance between top-down and bottom-up is surely a sign of truth.

Teleological indeterminacy: Fodor (1990) has complained that teleosemantic theories are plagued with an insurmountable problem with indeterminacy. The classic example is the frog’s ‘fly detector’, which is a complex of cells in its nervous system that respond to dark moving spots in its visual field, and cause the frog’s tongue to shoot out. A natural interpretation is that this complex of cells has the job of detecting flies, and therefore represents flies. However, Fodor argues that evolution is indifferent to whether the cells detect flies or dark moving spots, as long as, in the frog’s environment, most dark moving spots are flies. So there is

no evolutionary reason to claim that the cells' job is to detect flies rather than dark moving spots, or vice versa.

SINBAD neurosemantics, though it is a teleological theory with an evolutionary component, is not vulnerable to Fodor's complaint of indeterminacy. This is because it relies on evolution only to narrow down the contents of mental representations to sources of correlation; their specific contents are derived, in a very different way, from learning history. If we adapt Fodor's complaint about the frog's fly detector to the aspect of SINBAD neurosemantics that relies on evolution, it rings hollow: 'Evolution is indifferent to whether SINBAD cells correspond (in the context of an isomorphism) to sources of correlation or rather to X, as long as, in the organism's environment, most X's are sources of correlation'. I am not aware of anything that can plausibly be filled in for 'X' that also results in different content assignments for particular cells.

8. Conclusion

Once one understands the underlying SINBAD mechanism, it is relatively simple to understand, in basic outline, the theory of mental representation that emerges from it. SINBAD cells have the purpose, job, or teleofunction of yielding reliable 'predictions', by participating in internal dynamic structures that are isomorphic to the environment. They fulfill this purpose only by tuning to and thereby coming to correspond (in the context of an isomorphism) to sources of correlation. Thus SINBAD cells, as a general class, are supposed to correspond to sources of correlation in the context of an isomorphism, i.e. they represent in the way that the elements of a model do. A particular cell has the derivative job of corresponding to a *particular* source of correlation, namely that source of correlation that historically explains, according to the SINBAD design, how that cell has become able to fulfill its more general job of 'prediction' via filling in. In less than ideal cases, imperfect isomorphism and multiple explanatory sources can lead to misrepresentation and equivocal representation respectively.

Thus we have reason to believe that SINBAD cells have representational content. Since we also have reason to believe that the cerebral cortex is the seat of the mind, if the cortex is indeed composed of SINBAD cells we have reason to believe that *mental* representation, in us, is SINBAD representation. This hypothesis is further supported by the fact that SINBAD cells represent the same sorts of things that we do. They do not represent e.g. disjunctions of receptor activations; they represent sources of correlation, a category that includes kinds, individuals, properties, and other distal items that we represent. Further, they have a number of features, only some of which I have canvassed here, that make them highly suitable to play the theoretical role we expect of mental representations.

I have been concerned to develop a theory of how we and our biological relatives mentally represent. It is possible that there are creatures who possess other ways of constructing internal models of the environment that do not rely

upon the SINBAD mechanism, nor upon sources of correlation. In terms of their representational contents, these creatures' minds would be rather different from our own, and since their mechanism for constructing internal models is different, so would be the principles determining their mental contents. Maybe even more radically different kinds of minds are possible, minds that represent in a way entirely unlike models. On the other hand, perhaps ours is the only physically possible form of intentionality. . . but that is certainly not something we can hope to know *a priori*.

References

- Abeles, M. 1991: *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge: Cambridge University Press.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. 1989: *Molecular Biology of the Cell*. New York: Garland Publishing.
- Allen, C. Bekoff, M., and Lauder, G. (eds.). 1998: *Nature's Purposes: Analyses of Function and Design in Biology*. Cambridge, Mass.: MIT Press.
- Anderson, J.A. 1995: Associative networks. In M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*. pp. 102–107. Cambridge, Mass.: MIT Press.
- Ariew, A., Cummins, R. and Perlman, M. (eds.). 2002: *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press.
- Baker, L.R. 1995: *Explaining Attitudes: A practical approach to the mind*. Cambridge: Cambridge University Press.
- Becker, S., and Hinton, G.E. 1992: Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355, 161–163.
- Benacerraf, P. 1965: What numbers could not be. *Philosophical Review*, 74, 47–73.
- Bigelow, J. and Pargetter, R. 1987: Functions. *Journal of Philosophy*, 84, 181–196.
- Bilgrami, A. 1992: *Belief and Meaning*. Oxford: Blackwell.
- Boorse, C. 1976: Wright on functions. *Philosophical Review*, 85, 70–86.
- Boyd, R.N. 1991: Realism, anti-foundationalism, and the enthusiasm for natural kinds. *Philosophical Studies*, 61, 127–148.
- Burge, T. 1979: Individualism and the mental. *Midwest Studies in Philosophy*, 4, 73–122.
- Craik, K.J. 1943: *The Nature of Explanation*. Cambridge: Cambridge University Press.
- Cummins, R. 1975: Functional analysis. *Journal of Philosophy*, 72, 741–765.
- Cummins, R. 1989: *Meaning and Mental Representation*. Cambridge, Mass.: MIT Press.
- Davidson, D. 1987: Knowing one's own mind. *Proceedings and Addresses of the APA*, 60, 441–458.
- Devitt, M. 1981: *Designation*. New York: Columbia University Press.
- Devitt, M. 1990: Meanings just ain't in the head. In G. Boolos (ed.), *Method, Reason and Language: Essays in Honor of Hilary Putnam*. pp. 79–104. Cambridge: Cambridge University Press.
- Douglas, R. and Martin, K. 1998: Neocortex. In G. M. Shepherd (ed.), *The Synaptic Organization of the Brain*. pp. 459–509. Oxford: Oxford University Press.

- Dretske, F. 1981: *Knowledge and the Flow of Information*. Stanford: CSLI Publications.
- Dretske, F. 1986: Misrepresentation. In R.J. Bogdan (ed.), *Belief: Form, Content and Function*. pp. 17–36. Oxford: Oxford University Press.
- Dretske, F. 1988: *Explaining Behavior*. Cambridge, Mass.: MIT Press.
- Dretske, F. 1993: The nature of thought. *Philosophical Studies*, 70, 185–199.
- Dretske, F. 1995: *Naturalizing the Mind*. Cambridge, Mass.: MIT Press.
- Edelman, G.M. 1987: *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.
- Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. 1996: *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, Mass.: MIT Press.
- Favorov, O.V. and Kelly, D.G. 1996: Stimulus–response diversity in local neuronal populations of the cerebral cortex. *Neuroreport*, 7(14), 2293–2301.
- Favorov, O.V. and Ryder, D. forthcoming: SINBAD: a neocortical mechanism for discovering hidden environmental variables and their relations. *Biological Cybernetics*.
- Favorov, O.V., Ryder, D., Hester, J.T., Kelly, D.G. and Tommerdahl, M. 2003: The cortical pyramidal cell as a set of interacting error backpropagating dendrites: Mechanism for discovering nature’s order. In R. Hecht-Nielsen and T. McKenna (eds.), *Computational Models for Neuroscience: Human Cortical Information Processing*. pp. 25–64. Berlin: Springer-Verlag.
- Feldman, M.L. 1984: Morphology of the neocortical pyramidal neuron. In A. Peters and E.G. Jones (eds.), *Cerebral Cortex*, vol. 1. pp. 123–200. New York: Plenum.
- Felleman, D.J., and van Essen, D.C. 1991: Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1, 1–47.
- Fodor, J. 1987: *Psychosemantics*. Cambridge, Mass.: MIT Press.
- Fodor, J. 1990: *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press.
- Fodor, J. 1998: *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.
- Gelman, S.A. and Coley, J.D. 1991: Language and categorization: The acquisition of natural kind terms. In J.P. Byrnes and S.A. Gelman (eds.), *Perspectives on Language and Thought*. pp. 146–196. Cambridge: Cambridge University Press.
- Gurney, K.N. 2001: Information processing in dendrites – II. Information theoretic. *Neural Computation*, 14(8), 1005–1022.
- Hacking, I. 1991: A tradition of natural kinds. *Philosophical Studies*, 61, 109–126.
- Hanson, S.J. 1999: Connectionist neuroscience: Representational and learning issues for neuroscience. In E. Lepore and Z. Pylyshyn (eds.), *What Is Cognitive Science?*. pp. 401–427. Oxford: Blackwell.
- Intrator, N. 1995: Competitive Learning. In M.A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*. pp. 220–223. Cambridge, Mass.: MIT Press.
- Kornblith, H. 1993: *Inductive Inference and Its Natural Ground*. Cambridge, Mass.: MIT Press.
- Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M. and Malach, R. 2001: A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, 11, 287–297.

- Loewer, B. 1987: From information to intentionality. *Synthese*, 70, 287–317.
- Lycan, W.G. 1988: *Judgement and Justification*. Cambridge: Cambridge University Press.
- McAllister, A.K. 2000: Cellular and molecular mechanisms of dendrite growth. *Cerebral Cortex*, 10, 963–973.
- McGinn, C. 1989: *Mental Content*. Oxford: Blackwell.
- McLaughlin, P. 2001: *What Functions Explain*. Cambridge: Cambridge University Press.
- Mel, B.W. 1994: Information processing in dendritic trees. *Neural Computation*, 6, 1031–1085.
- Mel, B.W. 1999: Why have dendrites? A computational perspective. In G. Stuart, N. Spruston, and M. Häusser (eds.), *Dendrites*. pp. 271–289. Oxford: Oxford University Press.
- Millikan, R. 1984: *Language, Thought, and Other Biological Categories*. Cambridge, Mass.: MIT Press.
- Millikan, R. 1989: Biosemantics. *Journal of Philosophy*, 86(6), 281–297.
- Millikan, R. 1998: A common structure for concepts of individuals, stuffs, and real kinds: more Mama, more milk, and more mouse. *BBS*, 21(1), 55–65.
- Millikan, R. 1999: Historical kinds and the ‘special sciences’. *Philosophical Studies*, 95(1&2), 45–65.
- Millikan, R. 2000: *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.
- Pakkenberg, B., and Gundersen, H.J.G. 1997: Neocortical neuron number in humans: effect of sex and age. *Journal of Comparative Neurology*, 384, 312–320.
- Papineau, D. 1987: *Reality and Representation*. Oxford: Blackwell.
- Papineau, D. 1993: *Philosophical Naturalism*. Oxford: Blackwell.
- Perry, J. 2001: *Knowledge, Possibility and Consciousness*. Cambridge, Mass.: MIT Press.
- Phillips, W.A. and Singer, W. 1997: In search of common foundations for cortical computation. *Behavioral and Brain Sciences*, 20, 657–722.
- Poirazi, P. and Mel, B. 2001: Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron*, 29, 779–796.
- Putnam, H. 1975: The meaning of meaning. In K. Gunderson (ed.) *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.
- Quartz, S.R. and Sejnowski, T.J. 1997: The neural basis of cognitive development: a constructivist manifesto. *Behavioral and Brain Sciences*, 20, 537–596.
- Ramsey, F.P. 1931: *The Foundations of Mathematics and other logical essays*. London: Routledge.
- Resnik, M.D. 2000: *Mathematics as a Science of Patterns*. Oxford: Oxford University Press.
- Ritter, H. 1995: Self-Organizing Feature Maps: Kohonen Maps. In M.A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*. pp. 846–851. Cambridge, Mass.: MIT Press.
- Ryder, D. (2002) *Neurosemantics: A Theory*. Unpublished Dissertation, Chapel Hill: University of North Carolina, Department of Philosophy.
- Ryder, D., and Favorov, O.V. 2001: The new associationism: A neural explanation for the predictive powers of cerebral cortex. *Brain & Mind*, 2(2), 161–194.

- Schroeder, T. (1998) *The Foundations of Mental Representation*. Unpublished Dissertation, Stanford: Stanford University, Department of Philosophy.
- Searle, J. 1983: *Intentionality: an Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Segal, G. 2000: *A Slim Book About Narrow Content*. Cambridge, Mass.: MIT Press.
- Shouval, H.Z., and Perrone, M.P. 1995: Post-Hebbian learning rules. In M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*. pp. 745–748. Cambridge, Mass.: MIT Press.
- Spruston, N., Stuart, G., and Häusser, M. 1999: *Dendritic integration*. In G. Stuart, N. Spruston and M. Häusser (eds.), *Dendrites*. pp. 231–270. Oxford: Oxford University Press.
- Stalnaker, R. 1989: On what's in the head. *Philosophical Perspectives*, 3, 287–316.
- Stuart, G., Spruston, N., Sakmann, B. and Häusser, M. 1997: Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends in Neurosciences*, 20, 125–131.
- Wei, D.-S., Mei, Y.-A., Bagal, A., Kao, J.P.Y., Thompson, S.M. and Tang, C.-M. 2001: Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science*, 293, 2272–2275.
- Woolley, C.S. 1999: Structural plasticity of dendrites. In G. Stuart, N. Spruston, and M. Häusser (eds.), *Dendrites*. pp. 339–364. Oxford: Oxford University Press.
- Wright, L. 1973: Functions. *Philosophical Review*, 82, 139–168.