Too close for comfort? Psychosemantics and the distal

Dan Ryder
University of British Columbia (Okanagan)

Abstract

Naturalistic theories of intentionality typically fail to explain how our mental representations manage to denote distal things rather than mere disjunctions of proximal stimuli. In this paper, I present a neuroscience-based teleosemantic solution to the distality problem. The key observation is that a certain broad type of distal entity - which includes individuals and kinds - is selectionally relevant to the design of the representational network in the cerebral cortex.
Too close for comfort? Psychosemantics and the distal

What makes a mental representation about what it's about? The majority view among naturalists seems to be that representation has something to do with causation, or information, or correlation, or some other related notion. But such "information-based" views (e.g. Fodor, Prinz, Stalnaker, Usher, Mandik, Tye, and lots of other people who gesture towards this kind of theory\(^1\)) cannot accommodate representation of the distal.

For example, the activity of a "face detector" neuron will be better correlated with some disjunction of proximal stimuli than with faces, since it can be easily "fooled" when a non-face causes facelike proximal stimuli. An information-based view, then, would seem to dictate that such a cell isn't a face detector, but rather a facelike-proximal-stimuli detector - and similarly for most other internal representations. This "distality problem" is a variety of the more general "disjunction problem" (Fodor 1987).

There aren't any well-accepted fixes for this problem. The one that is probably viewed as being the most promising involves adding a pinch of teleology. The hope is that naturally occurring indicators - like the face detector - could have distally directed indicator functions, like a gas gauge does (it has the function of indicating the level of gas in its attached tank, not the current in its lead wire) - except of course, the relevant functions are meant to be evolutionarily endowed.

Fred Dretske offers a teleological theory like this (Dretske 1988). But it isn't widely appreciated that his "solution" still falls victim to the distality problem - though I hasten to add that Dretske himself realizes this (in correspondence). The problem seems to doom any selection-based teleosemantic indicator account, in fact (see fig. 1). In such accounts, what's supposed to endow an indicator I with the function of indicating distal kind K is some selective process. Let's say kind K is horses. If indicator I is selected for indicating horses, that means I's indicating horses has been either 1) causally relevant to the presence of I in this organism or species, or 2) causally relevant to I's recruitment to perform some biologically relevant task. But why suppose that I has been selected to indicate horses rather than a disjunction of local signs ("ls" in figure 1) of horses? If horses have been causally relevant to I's selection, then the local signs of horse that mediate I's ability to indicate horses have also been causally relevant to I's selection. There is no reason to pick the kind horse rather than local signs of horse as I's content.

The bottom line is that information-based accounts, whether teleological or not, are vulnerable to the distality problem.

My proposal for a solution to this problem starts, not from indication, but from a different classical foundation for representation, namely resemblance or isomorphism - the foundation of representation in scale models, for instance. Isomorphism is a relation that obtains between two structures, where a mapping between the elements of the structures preserves a pattern of relations across the mapping (Russell 1927). To get representation, we add a pinch of teleology (figure 2): the small box "models" the big box only if it has the function of mirroring the big box, and element e' "stands in" for e only if it has the function of corresponding to e. Then we can explain, for instance, why model representation is asymmetric. The small box represents the big box because its supposed to be isomorphic to the big box, but not vice versa (even though the isomorphism goes both ways). Similarly, we can explain why your wallpaper isn't a map of Leicester square, even if it happens to be isomorphic to Leicester square, as long as it isn't supposed to be isomorphic to Leicester square.

Figure 1. Dretske's theory is also subject to the distality problem.

Figure 2. Isomorphism and correspondence in the context of an isomorphism
These functions in scale models are normally dependent on the intentions of the model's designer, but it is possible to eliminate this dependence. Model design is often template-based (figure 3), and we can take away the designer's intentions by automating template-based design. The machine in fig. 4 takes a mold of an object and shrinks it, producing a miniature scale-model of its input. The object specific intentions of the designer have been eliminated - the only thing left to determine the representational content of the model produced is the template object, i.e. whatever object it happens upon.

So to know what this kind of model represents, you have to know its history, its origin or template. In addition, you need to know the machine's design principles. The machine in fig. 4 is designed to produce spatial isomorphs from spatial templates, not, for example, density isomorphs from colour templates. This explains two things: 1) why the resultant model only models the spatial structure of the input object, and not its density structure, and 2) why the resultant model's spatial structure is its only aspect that is representational. The model doesn't correctly represent the plane's density structure for the same reason that a black-and-white TV doesn't correctly represent the colour of a zebra. These features of the representations aren't supposed to be similar or isomorphic to anything, even if they happen to be so.

Of course, this "supposed to" doesn't come from nowhere - we still have the machine's designer in play here. That's what we replace with evolution. The idea is that the automatic scale modeler and the brain have two equivalent stages: the first stage is intentional design or evolutionary design, and the second stage is automatic template-based model production, according to some specific design principles.
I think that the brain - or rather the cerebral cortex - is a general-purpose model-building machine. This raises the obvious question: what are the design principles of the cortex? First I'll talk about what it models, then I'll talk about how it models it.

Plausibly, the cerebral cortex is for prediction, broadly construed. If you want a model that facilitates prediction, that model needs to be isomorphic to environmental regularities. One way to do this is to have a dynamic model, like a solar system orrery. The gears in an orrery ensure that the relative planet motions in the model are isomorphic to how they really are. This allows for prediction through the filling in of missing information, of, say, the position of Venus in two months.

In order for the cortex to build dynamic models, we would need something about its structure to be influenced by the co-variational structure of the environment, such that an environmental regularity comes to be mirrored by an internal regularity. Classical associationism fits this description (figure 5), but as we'll see, the cortex appears to be designed to mirror a more complex regularity pattern than pairwise correlation, namely the pattern that characterizes real kinds.

Figure 5. Classical associationism

Boyd, Kornblith, and Millikan (among others)\(^2\), have developed closely related accounts of natural kinds in terms of a particular pattern of regularities. This “unified property cluster” account says that a natural kind is characterized by a cluster of correlated properties, where there's some underlying reason for why they're correlated, and so why they ground reliable inductions; water is the classic example. Water is an example of what I call a "source" or "focus of correlation." Real kinds are a subclass of sources of correlation.

---

Sources of correlation exhibit multiple correlations due to some particular unifying explanation, but the type of explanation can vary from kind to kind. In the case of biological kinds, these correlations aren't due to an underlying chemical structure, but rather due to their common evolutionary history. The clustering of properties of some group of artifacts might be explained by the fact that they all serve some specific function (like the screwdriver), or because they originate from the same plan (like the PowerBook), or because they've been copied for sociological reasons (like a coat of arms and its variants).

My proposal is that the cortex models regularities organized around sources of correlation.

Now we need the design principles of the cortical machine: the SINBAD theory (Favorov & Ryder, 2004; Ryder 2004). The SINBAD theory is primarily a theory of plasticity in the cerebral cortex. It's a proposal for what functional principle underlies the synaptic and dendritic modifications in the principal cells of the cortex (figure 6), the pyramidal cells (or possibly larger multicellular units), as they tune their response properties under the influence of the environment (Favorov & Ryder, 2004; Ryder 2004). In particular, the theory explains why particular principal dendrites settle on particular input-to-output causal profiles. It says that these modifications accord with a very simple rule - that each principal dendrite will make adjustments so that it will tend to contribute the same amount of activity to the cell body as the other principal dendrites on the cell. So if there are 3 principal dendrites, like on the schematic pyramidal cell on the left of figure 6, they will each tend to adjust their connections over time so that they'll consistently contribute 1/3rd of the cell body's activity. I'll call this a tendency towards "dendritic matching".

Figure 6. Pyramidal cell and schematic representation

---

3Millikan (2000) also points out that individuals fall into the same pattern.
The acronym “SINBAD” (which stands for Set of INteracting Backpropagating Dendrites) refers to the mechanism underlying this brute causal tendency - there are a number of possible implementations, which I won't get into now. (The simplest version involves a variant of the gold-standard in neuroscience, Hebbian learning.)

To illustrate, consider a cell that has only two dendrites (figure 7). The SINBAD rule says they're both trying to contribute half, consistently, no matter what inputs the cell gets. In figure 7, the left dendrite receives an input, but it contributed all the cell body's activity, and the other dendrite contributed none - they've failed to match. The SINBAD rule says adjust your connections so that you'll match better next time in this condition - so the "a" connection will be weakened. As long as it gets uncorrelated inputs, the dendrites can't match, and the connections shown are just going to die off.

![Figure 7. A pyramidal cell with two principle dendrites and 6 inputs (3 per principle dendrite). It is currently receiving a single input at "a".](image)

![Figure 8. The cell receives coincident inputs.](image)
In order for the dendrites to learn to match, they have to get correlated inputs. In figure 8, let's say the cell receives a beak input at "a" and a feather input at "e". (That is, input a carries the information [in the Dretskean sense] that a beak is present, and input e carries the information that a feather is present.) Each time there's this coincident activation, those connections will strengthen (or at least "e" will). Meanwhile, the others die off, until you have something like the situation depicted in figure 9. The cell is now detecting the conjunction of beak and feather - they're lifting the cell above threshold, and it's firing. But this correlation isn't a coincidence - beaks and feathers are correlated for some underlying reason. This cell has become a half-decent bird detector - it has tuned to the source of correlation, birds. (I should add that a dendrite can learn to combine variables in particular ways, making use of multiple inputs.4)

Figure 9. The two principal dendrites have achieved consistent matching, by having strengthened correlated inputs "a" and "e", and having dropped inputs "b", "c", "d", and "f".

This is what SINBAD cells do - it naturally comes out of the dendritic matching mechanism that they have a strong tendency to tune to sources of correlation. Next: SINBAD models.

The model develops naturally as SINBAD cells make use of inputs from other cells in the cortical network, in order to get their dendrites to match (figure 10). One example of a cell using other SINBAD cells' outputs is in moving up the cortical hierarchy (figure 11). At the periphery, where cells are closely coupled with sensory inputs, local sources of correlation will be discovered, for example, edges, textures, surfaces, translational motion, and depth. As we move up the hierarchy, these earlier cells' outputs are used by higher areas, and cells will tune to sources of correlation at greater spatio-temporal scales - perhaps 3D shapes - and eventually more abstract things, like kinds.

4Most correlations are of this more complex type - beak and pink aren't well correlated, standing on-one-leg and feathered aren't well correlated... but beak AND one-leg is well correlated with pink and feathered. Again, a cell that finds this correlation will come to tune to its source, a flamingo.
But these relations aren’t restricted to mounting up the hierarchy. Sources of correlation are predictively related, in any direction, both within and across levels of scale and abstraction. Cats are related to fur and to hunting, water is related to taps and salt, and grass is related to greenness and to suburbia. So in tuning to a source of correlation, the dendrites of a particular cell find logical or mathematical functions that relate that source of correlation, not only to sensory inputs, but also to other sources of correlation.

These connections between cells are analogous to the gears in the solar system model; they’re what makes covariation in the network mirror covariation in the environment, allowing for the all-important “filling in” process. 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 thing they’ve tuned to.

This “filling in” is accomplished because a cell that has tuned to a particular thing has multiple sources of information about that thing, from numerous sensory input channels, and also neighbouring cells. If one of those sources of information is blocked off - maybe some sensory inputs - the others will compensate.

To summarize: The multiple dendrites on a SINBAD cell must find functions of their inputs that are correlated. Assuming these correlations aren’t accidental, the cell will tune to their source. In tuning to a source of correlation, a cell will provide neighbouring cells with a useful input, that is, an input that helps their dendrites find correlated functions. So these neighbouring 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. This is obviously a useful thing to have. And it's all driven by individual cells trying to get their dendrites to match. (I should note that this has all been modeled computationally in a large number of input domains.)

The dynamic isomorphism that develops mirrors the deep structure of the environment, with elements that correspond to the individuals and kinds around which environmental regularities are organized. Now, I want to say that, not only is there isomorphism and correspondence to this deep structure, but that this isomorphism and correspondence is teleofunctional or normative. The brain is supposed to mirror the deep structure of the environment, so this is what the brain represents or models. And SINBAD cells are supposed to correspond to real kinds and individuals in the context of this deep isomorphism - so they represent or stand in for, not retinal events, or light, but distal real kinds and individuals.

I think this teleological hypothesis is eminently plausible - and it's here where the neurocomputational details have a real philosophical payoff. It is the fact that SINBAD cells tune specifically to sources of correlation that makes them especially useful (and, so, probably was an important factor in their selection). There are multiple reasons for this.

First, note that it is useless for any mechanism that is designed to mirror the correlational structure of the environment to lock onto accidental correlations. But, one might ask, how can mirroring grounded regularities be selected for, as opposed to just mirroring regularities tout court? In order for a mechanism to be designed to mirror grounded regularities, it must incorporate some way of picking them out. But how could any mechanism do this?

Well, the SINBAD mechanism does this, as compared to its competitors (that is, other broadly “associative” mechanisms, including those actually found in other parts of the brain, like the hippocampus, basal ganglia, amygdala, and cerebellum). It doesn't, of course, incorporate some infallible groundedness detector, but function doesn't require infallibility - it just requires success often enough - often enough for a competitive advantage, for instance.

Multiple correlations are a sign of groundedness. (Related ideas commonly surface in epistemology and metaphysics. In epistemology: corroborating evidence indicates truth. In metaphysics: multiple causal powers that are only contingently co-instantiated indicate a real object.) And SINBAD networks, as opposed to standard associative networks, exhibit a peculiar sensitivity to multiple correlations. So SINBAD
networks, as opposed to other correlation-based mechanisms, have a (fallible) way of winnowing out the useful, grounded correlations. Thus they could have been selected for this ability, and plausibly were - so they plausibly have the specific function of mirroring grounded regularities.

SINBAD networks make very effective use of this ability, as well: this is no spandrel. First, by focusing its associative links on sources of correlation, which are inductively rich, the network acquires enormous, multipotent filling-in capabilities. A cell can “reidentify” (Millikan 2000) a source of correlation in a huge variety of circumstances.

Second, once a SINBAD cell starts to tune to a kind by discovering some of the correlated properties it exhibits, the cell is in a uniquely advantageous position to discover further correlation. So a cell continually adds to its lines of “evidence” for the presence of the source of correlation it's tuning to, continually improving the model's isomorphism.

Third, a cell's inputs largely come from other cells in the network, and since those cells have tuned to sources of correlation, and environmental regularities are often causally determined by interactions among sources of correlation (e.g. real kinds), the network naturally creates a rich mine of useful data in which to discover further correlation.

Relatedly, learning becomes much more efficient when a SINBAD network creates correspondences to sources of correlation. For real-world problems, models of deep structure are generally simpler - having fewer relations involving fewer variables - and that means they're easier to learn. (Compare the "surface" and deep models in fig. 12.) This simplifying effect is a general characteristic of sources of correlation.

The bottom line is that the value of the SINBAD mechanism is entirely dependent on the fact that the network detects and is structured by real kinds, individuals, and other sources of correlation. The richness, productivity, and efficiency of SINBAD model construction is all due to its special fit with sources of correlation. If SINBAD correctly describes the cortex, then it is highly plausible that this special fit explains why it was selected.
Thus I propose that it is part of the function of the cortex to be structured by real kinds, individuals, and the regularities they enter into. It’s a dynamic model-building machine that is designed to produce models of regularities involving sources of correlation. The main design principle of this machine is the dendritic matching rule.

Recall that a specific model represents the thing that structured it in accordance with the design principles of its model making machine. A product of the automatic scale modeler represents the spatial structure that explains its production, according to design. Similarly, a SINBAD network represents the regularity structure that it has come to mirror in accordance with its design principles. This is supposed be a regularity structure involving sources of correlation.

Which sources of correlation? To answer that, we just look at the SINBAD design principles. When things go according to design, a SINBAD cell tunes to a specific source of correlation by capitalizing on it to make its dendrites match. That is, when things go according to design, there will be a single source of correlation - say, the kind horse - that the cell has run into numerous times, and which explains how the cell has managed to achieve consistent dendritic matching. This is analogous to a little bit of an object making its part of a model produced by the automatic scale modeler - that bit of the model then represents the corresponding bit of the object that caused it. (Of course, things might not go according to design, which
can result in an inaccurate model. In the automatic scale modeler, a strong wind or high temperature affects the model. In a SINBAD network, perhaps the model is structured by spurious correlations, or a SINBAD cell might fail to separate two sources of correlation (e.g. jadeite and nephrite) - both contribute to its tuning.)

Returning to the distality problem, we can now see that the various proximal stimuli that mediate the cell’s response do not explain its dendritic matching. Or at least to the minute extent that they do, they aren’t remotely competitive with the relevant kind, e.g. horse, whose inductive richness is the foundation for all of the cell’s dendritic matching. The kind horse explains why the proximal stimuli are correlated, a single proximal stimulus does not. (This isn’t to say that a proximal stimulus, assuming that it is itself a source of correlation, couldn’t be the foundation for dendritic matching in some other cell. But this dendritic matching would have a very different functional profile, and a very different history.) All that is required, then, to solve the distality problem is the following plausible principle:

When two sources of correlation are in competition to explain a SINBAD cell’s dendritic matching, the one that the cell is supposed to correspond to is the source of correlation that best explains its dendritic matching.

This contrasts markedly with an information-based approach, where, without some ad hoc move, one must accept that the referent is what the representation correlates best with - and that’s the disjunction of proximal stimuli. Disjunctive information isn’t “worse” information, but a disjunctive explanation is a worse explanation.

That is my principled solution to the distality problem. The key move is to make distal entities - individuals, kinds, and other sources of correlation - selectionally relevant in the design of SINBAD networks. In Dretske’s case, recall, there was no reason to say that horses, as opposed to local signs of horses, were selectionally relevant to the presence of some indicator, or selectionally relevant to what that indicator is recruited to do. So there was no reason to say that the indicator was designed to detect horses rather than local signs of horses. By contrast, according to the SINBAD hypothesis, the cortex was designed specifically to create correspondences with distal items: kindhood, or more generally "source-of-correlation"-hood, was selectionally relevant to the design of SINBAD networks. So we do have a principled reason to say that a SINBAD cell represents horses rather than local signs of horses.

---

5 I leave it open that Dretske might be able to assimilate my general strategy to his approach, focusing on the selection of the learning mechanisms mediating operant conditioning. If these mechanisms operate by simple pairwise association, such a strategy is unlikely to yield distal contents, but if computational mechanisms are involved, as Gallistel (1990) has argued, the attempt would be more likely to succeed.
References