Connectionist Vehicles, Structural Resemblance, and the Phenomenal Mind

Gerard O'Brien and Jon Opie

Department of Philosophy
University of Adelaide
South Australia 5005

gerard.obrien@adelaide.edu.au
http://arts.adelaide.edu.au/Philosophy/gobrien.htm

jon.opie@adelaide.edu.au
http://arts.adelaide.edu.au/Philosophy/jopie.htm

Draft @ May 2001

To appear in "Naturalism and the Phenomenal Mind", a special issue of *Communication and Cognition*, edited by Johan Veldeman.

1. Introduction

We think the best prospect for a naturalistic explanation of phenomenal consciousness is to be found at the confluence of two influential ideas about the mind. The first is the *computational theory of mind*: the theory that treats human cognitive processes as disciplined operations over neurally realised representing vehicles.[1] The second is the *representationalist theory of consciousness*: the theory that takes the phenomenal character of conscious experiences (the "what-it-is-likeness") to be constituted by their representational content.[2] Together these two theories suggest that phenomenal consciousness might be explicable in terms of the representational content of the neurally realised representing vehicles that are generated and manipulated in the course of cognition. The simplest and most elegant hypothesis that one might entertain in this regard is that conscious experiences are identical to (i.e., are one and the same as) the brain's representing vehicles.

There are two profound problems with this hypothesis. First, to identify conscious experiences with the brain's representing vehicles is to assert that *all* of the information encoded in the brain at any moment is phenomenally experienced at that time. Yet it is orthodox in contemporary cognitive science to hold that our brains both represent and process far more information than we experience; that, at any moment, only a very small subset of the brain's representing vehicles contribute their contents to consciousness. This would seem to rule out any straightforward identification of the one with the other (Kihlstrom 1987; Velmans 1991). Second, to identify conscious experiences with the brain's representing vehicles, and additionally, to suppose that the phenomenal character of the former is nothing but the representational content of the latter, is to suggest that representational content is an *intrinsic* property of the brain's representing vehicles. But the prevailing view in the philosophy of mind is that representational content has very little to do with the intrinsic properties of the brain's representing vehicles, and

[1] This description is deliberately generic. Some writers tend to construe the computational theory of mind as the claim that cognitive processes are the rule-governed manipulations of internal symbols. However, we will take this narrower definition to describe just one, admittedly very popular, species of computational theory, viz: the classical computational theory of mind. Our justification for this is the emerging consensus within cognitive science that computation is a broader concept than symbol manipulation. See, e.g., Cummins & Schwarz 1991, p.64; Dietrich 1989; Fodor 1975, p.27; and Von Eckardt 1993, pp.97-116.

[2] See, e.g., Dretske 1993, 1995; Tye 1995a; Tye 1995b.

everything to do with their (e.g., actual, counterfactual or historical) causal relations with the world (Dretske 1981, 1988; Fodor 1987, 1990; Millikan 1984). Representationalism about consciousness, when combined with content externalism, also appears to rule out an identification of conscious experiences with the brain's representing vehicles.[3]

Either of these problems, considered alone, threatens the simple hypothesis bruited above. Taken together they would seem to rule it out definitively. But we hesitate, because the orthodox views in cognitive science and the philosophy of mind that make trouble for this hypothesis have been conditioned by a particular story about the way the brain encodes and processes information. This is the *classical* computational theory of mind—the theory that takes human cognition to be a species of symbol manipulation.[4]

In cognitive science, because it compels us to assume that human cognition involves a great many unconscious symbols, classicism has led to the dominance of what we call *process* theories of consciousness (O'Brien & Opie 1999a). According to such theories, the mere presence of a representing vehicle is not sufficient for consciousness; what matters is that it perform some special computational role, or be subject to specific kinds of computational processes (Baars 1988; Crick 1984; Dennett 1991; Flanagan 1992; Jackendoff 1987; Johnson-Laird 1988). And in the philosophy of mind, because there are no objective relations between the intrinsic properties of symbolic representing vehicles and the things they represent, classicism has contributed to the dominance of *causal*, hence externalist, theories of representational content.

But classicism is no longer the dominant theory of human cognition. It now has a lively competitor in the form of *connectionism*.[5] And when one takes a fresh look at these issues from the connectionist perspective, one finds the terrain considerably altered.

Specifically, connectionism permits the serious exploration of *vehicle* theories of phenomenal consciousness (O'Brien & Opie 1997, 1999a, 1999b). These are theories that explain conscious experience in terms of the intrinsic nature of the representing vehicles deployed by the brain—in terms of what these vehicles *are* rather than what they *do*. And connectionism also licenses a different approach to representational content, one couched not in terms of causal relations between the brain and the environment, but in terms of resemblance relations between intrinsic properties of the brain's representing vehicles and aspects of the world. Indeed, it is our contention that a *resemblance* theory of representational content is actually mandatory for connectionism (O'Brien 1999; O'Brien & Opie 1999c, Forthcoming).

Our hypothesis is that conscious experiences are identical to the brain's representing vehicles, and that the phenomenal character of the former is nothing but the representational content of the latter. The task of this paper is to demonstrate how a connectionist approach to the encoding and processing of information in the brain disarms the two problems with which this hypothesis appears to be burdened. In the next section we establish that connectionism has the resources to hazard a plausible *vehicle theory of consciousness*. Then, in section three, we sketch the outlines of a *structural resemblance theory of representational content*: one that explains representational content in terms of the intrinsic properties of the brain's representing vehicles, thereby avoiding content externalism. We finish with some observations about what the

---

[3] Indeed, the representationalist theory of consciousness and content externalism together entail *consciousness externalism*: the claim that "phenomenology ain't in the head" (Tye 1995b, p.151). This is something that some philosophers are prepared to live with (e.g., Dretske 1995; Tye 1995b). Others consider it a reductio of the representationalist theory of consciousness, and hence insist on distinguishing the *phenomenal* content of conscious experiences from their *representational* content (e.g., Revonsuo 2000).

[4] For those readers unfamiliar with classicism, a good entry point is provided by the work of Haugeland (1981; 1985 especially Chps.2 and 3).

[5] For useful introductions to connectionism, see Bechtel & Abrahamsen 1991; Clark 1989, Chps.5-6; Tienson 1987.

connectionist vehicle theory of consciousness and the resemblance theory of representational content together entail about the nature of the phenomenal mind.

2. A Connectionist Vehicle Theory of Consciousness

Identifying conscious experiences with the brain's representing vehicles is problematic because it is hard to deny that our brains represent and process far more information than we experience. The existence of declarative memory, for example, makes it plain that our brains unconsciously store a huge amount of information. However, given that theorists commonly distinguish between different styles of information coding in computational devices (Cummins 1986; Dennett 1982; Pylyshyn 1984), it is not unreasonable to suppose that the brain deploys more than one style of representation. Consequently, there is an obvious amendment to our original proposal that is in keeping with its general spirit. Rather than hold that consciousness is identical to the brain's representation of information *simpliciter*, we might conjecture that it is identical to a particular style of information coding in the brain. We explore this idea below by developing a connectionist account of the brain's representing vehicles.

2.1 Connectionist Styles of Mental Representation

Human cognitive processes, according to connectionism, are the computational operations of a multitude of connectionist networks implemented in the neural hardware in our heads.[6] A connectionist network is a collection of interconnected processing units, each of which has an *activation level* that is communicated to the rest of the network via modifiable, weighted connection lines. From moment to moment, each unit sums the weighted activation it receives, and generates a new activation level that is some threshold function of its current activity and that sum. A connectionist network typically performs computational operations by "relaxing" into a stable *pattern of activation* across its constituent units, in response to the input it receives. This relaxation process is mediated by the connection weights, because they determine how, and to what extent, activation is passed from unit to unit.

The representational capacities of connectionist networks rely on the plasticity of the connection weights between the constituent processing units.[7] By altering these connection weights, one alters the activation patterns the network produces in response to its inputs. As a consequence, an individual network can be taught to generate a range of stable target patterns in response to a range of inputs. These stable patterns of activation, because they are generated rapidly in response to the flux of input impinging on individual networks, constitute a transient form of information coding, which we will refer to as *activation pattern representation*.

While activation patterns are a transient feature of connectionist networks, a "trained" network has a relatively long-term capacity to generate a set of distinct activation patterns, in response to cueing inputs. So a network, in virtue of its connection weights, can be said to *store* appropriate responses to input. This form of information coding, which is sometimes referred to as *connection weight representation,* is the basis of long-term memory in connectionist systems. Such long-term storage of information is superpositional in nature, since each connection weight contributes to the storage of every stable activation pattern that the network is capable of generating. Consequently, the information that is stored in a network is not encoded in a physically discrete manner. The one appropriately configured network encodes a *set* of contents corresponding to the set of activation patterns it is capable of generating.

---

[6] In this context connectionist networks are to be understood as *idealised* models of real neural networks, which, although unrealistic in certain respects, capture the computationally significant properties of neural networks (see, e.g., Churchland & Sejnowski 1992, ch.3; O'Brien 1998; and Opie 1998).

[7] For good general introductions to the representational properties of connectionist systems, see Bechtel & Abrahamsen 1991, Chp.2; Churchland 1995; Churchland & Sejnowski 1992, Chp.4; and Rumelhart & McClelland 1986, Chps.1-3.

Connectionism thus presents us with two quite distinct styles of information coding that might be deployed in the brain: activation pattern representation, and connection weight representation. This immediately suggests how to amend our original proposal about phenomenal consciousness. Since stable activation patterns are relatively short-term, causally potent responses to current input, whereas connection weights are the basis of long-term information storage, we might seek to identify conscious experiences with stable patterns of activation in neurally realised connectionist networks. This is the basis of the *connectionist vehicle theory of phenomenal consciousness*.

It is possible to give a more general formulation of this theory. In connectionist theorising, activation patterns are the entities that receive an interpretation, such that each pattern of activation across a network has a distinct semantic value (often specified in terms of a semantic metric). In this respect activation pattern representations are akin to the tokens on the tape of a Turing machine. An individual pattern, just like a symbol on the tape, is an element in a system of physically structured objects for which there is a semantics (a mapping between individual representing vehicles and some represented domain), and a "parser" mechanism that is capable of recognising and responding to semantically significant variations in physical structure.[8] In the case of a Turing machine the parser is the read/write head through which the tape passes. An activation pattern is "parsed" by virtue of having effects on other networks.

Given the above, we believe it is warranted to apply to connectionism the now standard terminology and say that stable activation patterns represent information in an *explicit* fashion.[9] Formulated more generally, a vehicle theory of phenomenal consciousness thus identifies conscious experiences with the vehicles of *explicit* representation in the brain. What is specific to a connectionist vehicle theory is the further claim that these vehicles are stable patterns of activation in neurally realised connectionist networks (Atkinson et al. 2000; O'Brien & Opie 1997, 1999a).

Notice that this identity claim solves our problem regarding unconscious information storage in the brain. As already remarked, long-term memory in a connectionist network does not involve activation pattern representations, but depends on the network's configuration of connection weights. There are no physically discrete structures corresponding to the individual stored contents, because they are encoded in a superpositional manner. Such contents are therefore not explicit; they are merely *potentially explicit* (Dennett 1982, p.216-7). Information only becomes explicit when rendered as a stable pattern of activation by an appropriate cueing input. One can therefore consistently maintain a vehicle theory of phenomenal consciousness (as defined above) while holding that a great deal of information is stored *inexplicitly*, and hence unconsciously, in the brain.

2.2 Why Classicism Can't and Connectionism Can

Now that we have formulated the vehicle theory of phenomenal consciousness in terms of the distinction between explicit and inexplicit information coding, it is natural to wonder whether classicism has access to this kind of theory. A classicist can certainly allow that much of the information stored in the brain is merely potentially explicit. Might a classicist therefore identify conscious experiences with mental symbols generated from moment to moment in the brain, and relegate the unconscious to the merely potentially explicit? Here we briefly argue that such a

[8] This formulation is adapted from Dennett 1982, p.216. One important respect in which activation pattern representations differ from classical symbols is that their semantics is not language-like. Symbol structures, unlike activation pattern representations, have a (concatenative) combinatorial syntax and semantics. The precise nature of the internal structure of connectionist representations is a matter of some debate (see, e.g., Fodor & Pylyshyn 1988; Smolensky 1987; and van Gelder 1990).

[9] For a detailed argument to this effect see O'Brien & Opie 1999a, pp.133-7; for discussion see Clapin & O'Brien 1998.

vehicle theory is not available to classicists, and identify the crucial computational difference between classicism and connectionism in this regard.

Classicism is committed to a digital conception of the computational processes involved in cognition, and a symbolic conception of the vehicles over which those processes are defined. The computational capacities of a digital device are embodied in the rules that regulate the behaviour of its explicit representing vehicles. Thus, a Turing machine (the abstract model for all digital computers) has its computational powers in virtue of the rules it embodies; both those explicitly written down on the machine's tape, and those that take the form of primitive instructions built into the machine's read/write head. Likewise, says the classicist, the computational powers of the brain depend upon the rules it embodies; both those explicitly tokened in the form of neural symbol structures, and those that are part of the brain's "functional architecture" (Pylyshyn 1984).

To defend a classical vehicle theory of phenomenal consciousness one must claim that the only symbols generated in the course of cognition are those whose contents we consciously experience. Now, although one is occasionally conscious of rules that govern one's thinking (as for example when doing mental arithmetic), by and large we are aware of the *products* of cognitive processes rather than the (rule-governed) processes themselves. This is particularly evident in the case of visual processing, which appears to be dependent on a set of very complex, and entirely unconscious rules for the construction of visual representations (Hoffman 1998; Palmer 1999). For this reason a classical vehicle theorist is forced to assume that most of the rules that govern cognition and perception are part of the functional architecture of the brain, rather than being among its explicit symbol structures. The trouble with this suggestion is that it is implausible to suppose that classicism can delegate *all* the cognitive work of the unconscious to functional architecture.

Whenever we act in the world, whenever we perform even very simple tasks, it is evident that our actions are guided by a wealth of knowledge concerning the domain in question.[10] To give one simple example, suppose you read in a paper:

> Police have revealed that the victim was stabbed to death in a cinema. Unfortunately, the chief suspect is now known to have been on an express train to Edinburgh at the time of the murder.[11]

On the basis of this information you will probably conclude that the police have the wrong man. However, notice that in reaching this conclusion you have relied on quite a number of beliefs which likely never entered consciousness, such as: a person can't be in two places at once, to stab someone you must normally be in close proximity to your victim, there are no cinemas on express trains to Edinburgh, and so on. To explain this kind of reasoning one seems forced to suppose that there is some medium within which unconscious beliefs are represented, in order that they may causally interact to produce rational outcomes. While it is possible that reasoning is partly dependent on inexplicit inference rules, functional architecture is normally assumed to be insensitive to changes in background assumptions or beliefs.[12] Our network of beliefs, on the other hand, is highly flexible and highly sensitive to new input. It is therefore not reasonable to ascribe the causal powers of unconscious beliefs to pieces of functional architecture.

To summarise: cognition appears to be actively shaped by a great deal of unconscious, domain specific knowledge; a classical vehicle theorist cannot suppose that such knowledge takes the form of explicit symbols, either in the form of data or rules, on pain of inconsistency;

[10] This fact about ourselves has been made abundantly clear by research in the field of artificial intelligence, where it gives rise to the so called "frame problem". See Dennett 1984 for an illuminating discussion.

[11] This example is adapted from Johnson Laird 1988, pp.219-20.

[12] In Pylyshyn's terms the functional architecture is "cognitively impenetrable" (1984).

nor is it reasonable to suppose that such knowledge is part of the brain's functional architecture, as we argued above.

But doesn't this leave open the possibility that unconscious beliefs are merely potentially explicit? In the classical context this won't wash. Consider again the operation of a Turing machine. Such a device is typically capable of rendering explicit a good deal of information beyond that written on its tape; information that is potentially explicit in virtue of the symbols currently on the tape, and the mechanisms resident in the machine's read/write head. But such information can only influence the behaviour of the system when it is rendered explicit. In order for potentially explicit information to throw its weight around it must first be physically embodied as symbols written on the machine's tape. Only then, once these symbols come under the gaze of the machine's read/write head, can the information they encode causally influence the computational activities of the system. Consequently, although a classicist can certainly allow for the storage of information in potentially explicit form, information so encoded is never causally active.

Since we have exhausted all other possibilities, it seems that the only way unconscious knowledge can causally influence cognition on the classical account is if it takes the form of explicit symbol structures. As Fodor puts it: "No Intentional Causation without Explicit Representation" (1987, p.25). We conclude that classicism doesn't have the resources required to develop a plausible vehicle theory of phenomenal consciousness. Any classicist who seeks a computational theory of consciousness has no alternative but to embrace a *process theory*—a theory that distinguishes between conscious and unconscious information not on the basis of the way it is encoded (there are both conscious and unconscious symbols), but in terms of the kinds of processes mental symbols are subject to.

This analysis of the prospects for a classical vehicle theory of phenomenal consciousness highlights the crucial difference between classicism and connectionism. Whereas potentially explicit information is causally impotent in the classical framework (it must be rendered explicit before it can have any effects), the same is not true of connectionism. This makes all the difference. In particular, whereas classicism, using only its inexplicit representational resources, is unable to meet all the causal demands cognition places on the unconscious (and is thus committed to a good deal of unconscious symbol manipulation), connectionism holds out the possibility that it can (thus leaving stable activation patterns free to line up with the contents of consciousness).

Potentially explicit information is encoded in a connectionist network in virtue of its relatively long-term capacity to generate a range of explicit representations (stable activation patterns) in response to cueing inputs. This capacity is governed by a network's configuration of connection weights. However, we saw earlier that a network's connection weights are also responsible for the manner in which it responds to input (by generating activation pattern representations), and hence the manner in which it processes information. This means that the mechanism driving the computational operations of a connectionist network is identical to the mechanism responsible for its long-term storage of information. So there is a strong sense in which it is the potentially explicit information encoded in a network (the network's "memory") that actually governs its computational operations.

This fact has major consequences for the connectionist take on cognitive processes. Crucially, information that is merely potentially explicit in connectionist networks need not be rendered explicit in order to be causally efficacious. The information that is encoded in a network in a potentially explicit fashion is causally active whenever that network responds to input. Consequently, most of the work that a classicist must assign to unconscious symbol manipulations, can in connectionism be credited to operations involving inexplicitly represented information. Explicit representations, on this alternative conception, are the *products* of unconscious processes, and thus a connectionist can feel encouraged in the possibility of aligning phenomenal experience with these representing vehicles.

In summary, the connectionist vehicle theory holds that phenomenal consciousness is identical to the explicit representation of information in neurally realised connectionist networks. Unconscious processing is treated as activation passing, both within and between networks, that results in stable patterns of activation. Such processing is entirely governed by potentially explicit information. Unconscious processes thus generate activation pattern representations, which the connectionist is free to identify with conscious experiences, since none is required to account for the unconscious activity itself.

3. A Structural Resemblance Theory of Representational Content

The hypothesis we are considering is that conscious experiences are identical to the brain's vehicles of explicit representation, and that the phenomenal character of the former is nothing but the representational content of the latter. In the previous section, we sought to show that this hypothesis can be made plausible by invoking connectionist resources, despite the fact that human cognition involves a great deal of unconscious information. But this still leaves the following problem: whereas our hypothesis implies that representational content is an *intrinsic* property of the brain's vehicles of explicit representation, the orthodox view in the philosophy of mind is that representational content has little to do with the intrinsic properties of the brain's representing vehicles. In what follows we address this issue by developing a new approach to the content of the brain's representing vehicles. We argue that this alternative understanding of representational content is actually mandatory for connectionism, and hence for the connectionist vehicle theory of consciousness in particular.

3.1 Computational Architecture and Representational Content

Philosophers of mind have long assumed that the problem of explaining how the brain's representing vehicles acquire their representational content is orthogonal to debates about the brain's computational architecture. In a recent book on connectionism, for example, Horgan and Tienson put it this way:

> [T]hroughout this book we have taken representation for granted. In doing so, we simply follow connectionist (and classicist) practice…. It is assumed that natural cognitive systems have intentional states with content that is not derived from the interpreting activity of other intentional agents. But it is not the business of either classical cognitive science or connectionist cognitive science to say where underived intentionality "comes from". (Horgan & Tienson 1996, p.13)

It is our contention, however, that this assumption is quite mistaken. Classicism and connectionism each imposes significant restrictions on the kind of theory of representational content that is permissible. In this subsection we will explain why.

The task of a theory of representational content is to explain how nervous systems can be in the representing business in the first place—how brain states can be about aspects of the world. It is a commonplace in the philosophy of mind that a theory of representational content must be *naturalistic*, in the sense that it cannot appeal to properties that are either non-physical or antecendently representational.[13] Given this constraint, it has frequently been observed that there would seem to be just two different objective relations that the brain's representing vehicles are capable of bearing to the world, and hence which might form the basis of a

---

[13] See, e.g., Cummins 1989, pp.127-29; 1996, pp.3-4; Dretske 1981, p.xi; Field 1978, p.78; Fodor 1987, pp.97-8; Lloyd 1989, pp.19-20; Millikan 1984, p.87; and Von Eckardt 1993, pp.234-9.

naturalistic explanation of representational content. These are *causation* and *resemblance*.[14] But which of these relations is the most appropriate, in our view, is determined by the brain's computational architecture.

Classicism operates with a *digital* conception of neural computation, and a *symbolic* conception of the brain's representing vehicles. The computational capacities of a digital device are embodied in the rules that regulate the behaviour of its explicit representing vehicles, rather than in the structural properties of the vehicles themselves (see, e.g., Fodor, 1987). Classicism has thus fostered a climate (in both cognitive science and the philosophy of mind) in which the theoretical focus is directed mainly at the computational/causal *relations* that representing vehicles enter into, and not at their *intrinsic* properties. This theoretical focus has had a significant impact on the development of theories of representational content. On the one hand, it has completely inhibited the development of resemblance approaches to representational content, since, as Cummins observes, "[classical] computationalists must dismiss similarity theories of representation out of hand; nothing is more obvious than that [symbolic] data structures don't resemble what they represent" (Cummins 1989, pp.30-1). And on the other, it has encouraged the development of causal theories of content, since causation would appear to be the one objective relation that symbols are capable of bearing to the world.

The computational capacities of connectionist networks, by contrast, are not inherited from rules that are distinct from the intrinsic properties of its representing vehicles. Indeed, as we saw in the previous section, connectionism dispenses with the classical distinction between representing vehicles and the processes that act on them (the so-called code/process divide—see Clark 1993). The very substrate that stores, in potentially explicit form, everything that a network "knows" (i.e., the network's configuration of weighted connections) is the mechanism that governs its computational operations. Connectionist devices achieve their computational competences not by applying rules to the representing vehicles they generate, but by deploying learning procedures which gradually shape these vehicles so that they come to resemble aspects of the task domains over which they operate (O'Brien 1999).

Consider, as an example, NETtalk (Sejnowski & Rosenberg 1987). NETtalk transforms English graphemes into contextually appropriate phonemes. This task domain is quite abstract, comprising the letter-to-sound correspondences permitted in the English language. Back-propagation is used to shape NETtalk's activation landscape—which comprises all the potential patterns of activity across its 80 hidden units—until the network performs accurately. Once it is trained up in this fashion, there is a systematic relationship between the network's activity and the target domain, such that variations in activation patterns systematically mirror variations in letter-to-sound correspondences. It is this resemblance relation that is revealed in the cluster analysis which Sejnowski and Rosenberg applied to NETtalk. And it is this resemblance relation that makes it right and proper to talk, as everyone does, of NETtalk's having a *semantic metric*, such that its activation landscape becomes a *representational* landscape.

When such a resemblance relation exists between a network's representing vehicles and its task domain, there is no need to apply rules to those representing vehicles in order to govern their processing. Instead, the computational processes of the network are governed by the model of the task domain that it embodies. Thus, when NETtalk is exposed to an array of graphemes, the resemblance relation embodied in its connection weights automatically produces the contextually appropriate phonemic output.

The upshot of all of this is that the computational capacities of a connectionist network are embodied in the intrinsic properties of its representing vehicles (see section 3.3 below). As a consequence, the almost exclusive focus in contemporary philosophy of mind on the *causal*

---

14 See, e.g. Von Eckardt 1993, pp.149-52. Some philosophers think that convention is a third possibility, but this is controversial, since it is not clear that convention is consistent with the naturalism constraint.

relations into which the brain's representing vehicles enter is no longer wholly appropriate. Connectionism brings with it an additional focus on the *intrinsic properties* of the representing vehicles themselves. Indeed, connectionism compels us to explain representational content in terms of resemblance relations between the intrinsic properties of the brain's representing vehicles and their target domains.[15]

But exactly what kind of resemblance relations are required to ground the representational content of connectionist representing vehicles? We take up this question in section 3.3 below. Before doing so, however, we need to consider resemblance more generally.

3.2 Varieties of Resemblance

Resemblance is a fairly unconstrained relationship, because objects or systems of objects can resemble each other in a huge variety of ways, and to various different degrees. However, one might hope to make some progress by starting with simple cases of resemblance, examining their possible significance for connectionist representing vehicles, and then turning to more complex cases. Let us begin, then, with resemblance between concrete objects. The most straightforward kind of resemblance in this case involves the sharing of one or more physical properties. Thus, two objects might be of the same colour, or mass, have the same length, the same density, the same electric charge, or they might be equal along a number of physical dimensions simultaneously. We will call this kind of relationship *physical* or *first-order resemblance*.[16] A representing vehicle and its represented object resemble each other at first order if they share physical properties, that is, if they are equal in some respects. For example, a colour chip—a small piece of card coated with coloured ink—is useful to interior designers precisely because it has the same colour as paint that might be used to decorate a room.

First-order resemblance, while relevant to certain kinds of public representation, is clearly unsuitable for connectionist representing vehicles, since it is incompatible with what we know about the brain's neural networks. Nothing is more obvious than the fact that our minds are capable of representing features of the world that are not replicable in patterns of neural activity. Moreover, even where the actual properties of neural networks are concerned, it is unlikely that these very often play a role in representing those self-same properties in the world.

There is, however, a more abstract kind of resemblance available. Consider colour chips again. Interior designers typically use *sets* of chips or colour *charts* to assist them in making design decisions. In other words, they employ a *system* of representations which depends on a mapping of paints onto chips according to their shared colour (their first-order resemblance). A useful side effect of having such a system is that when one wants to compare paints (eg., 2-place comparisons such is "this one is bolder than that one", or 3-place comparisons such as "this one harmonises better with this one than with that one") one can do so by comparing the cards. This is because the system of chips embodies the *same pattern of colour-relations* as the paints. Whenever pairs or triples of paints satisfy particular colour relationships, their ink-coated proxies fall under mathematically identical relations.

Similar remarks apply to surface maps. What makes a map useful is the fact that it preserves various kinds of topographic and metrical information. The way this is accomplished is by so arranging the points on the map that when location A is *closer to* location B than location C, then their proxies (points A, B and C on the map) also stand in these metrical relations; and when location A is *between* locations B and C, then points A, B and C stand in the same (3-place)

---

[15] This general approach to mental content has a venerable history in philosophy, but up until recently any kind of resemblance theory was thought to suffer from a number of fatal flaws (see, e.g., Cummins 1989, chp.3). Over the last few years a couple of philosophers have started to take this approach seriously again, especially in the form of second-order resemblance relations (see, e.g., Blachowicz 1997; Cummins 1996; and Gardenfors 1996).

[16] We are here adapting some terminology developed by Shepard & Chipman 1970.

topographic relation; and so on. The utility of a map thus depends on the existence of a resemblance relation that assigns points on the map to locations in the world in such a way that the spatial relations among the locations is preserved in the spatial relations among the points.

We will speak here of *second-order resemblance*.[17] In second-order resemblance, the requirement that representing vehicles share physical properties with their represented objects can be relaxed in favour of one in which the *relations* among a system of representing vehicles mirror the *relations* among their objects. Of course, the second-order resemblance between colour charts and paints is a consequence of the first-order resemblance between individual chips and their referents. And in the case of surface maps, space is used to represent space. But one can typically imagine any number of ways of preserving the pattern of relations of a given system *without* employing first-order resemblance. For example, the height of a column of liquid in a mercury thermometer is used to represent the temperature of any object placed in close contact with it. Here, variations in height correspond to variations in temperature.

Weather maps provide a more compelling example. On such a map, regions in the earth's atmosphere (at some specified elevation) are represented by points, and contiguous regions of equal atmospheric pressure are represented by lines known as "isobars"(see Figure 1). More significantly, the *spacing* of isobars corresponds to atmospheric pressure *gradients*, knowledge of which can be used to predict wind velocity, the movement of fronts, and so on. The representation of pressure gradients by isobar spacing is second-order, because for any two points on the map the *relative* spacing and orientation of isobars in the vicinity of those points corresponds to the *relative* size and direction of pressure gradients at the represented regions. Moreover, since *geometric* relations among lines and points on the map are being used to represent *pressure* relations this is a case of "pure" second-order resemblance—it doesn't depend on an underlying first-order resemblance.

The significance of second-order resemblance for explaining the representational content of the brain's representing vehicles is this. While it is extremely unlikely that first-order resemblance is applicable to mental representation (given what we know about the brain) the same does not apply to second-order resemblance. Two systems can share a pattern of relations *without* sharing the physical properties upon which those relations depend. Second-order resemblance is actually a very abstract relationship. Essentially nothing about the physical form of the relations defined over a system of representing vehicles is implied by the fact that it resembles a system of represented objects at second-order; second-order resemblance is a formal relationship, not a substantial or physical one.[18]

It is a little acknowledged fact that one general approach to representational content in the philosophy of mind exploits second-order resemblance. We have in mind the group of theories that go variously by the names *causal*, *conceptual*, or *functional role semantics*.[19] These *functional role* theories (as we shall call them) share a focus on the *causal* relations that a system of mental representing vehicles enter into; where they differ is in the class of causal relations they take to be significant for mental representation. What informs this causal focus is the idea that a system of vehicles represents a domain of objects when the former *functionally resembles* the latter. A functional resemblance obtains when the pattern of *causal* relations among a set of representing vehicles preserves at least some of the relations among a set of represented objects.

---

[17] See Palmer 1978; Shepard & Chipman 1970, and Shepard & Metzler 1971. Blachowicz 1997; Cummins 1996; Gardenfors 1996; Johnson-Laird 1983; O'Brien 1999; and Swoyer 1991 have all recently applied the concept of second-order resemblance to the problem of explaining representational content.

[18] A consequence of this is that a system of *mental* vehicles (which by assumption is a set of brain states) is not only capable of standing in a relationship of second-order resemblance to concrete or natural systems, but also to abstract systems such as logical formalisms and theories. This is presumably a welcome outcome.

[19] See, e.g., Block 1986, 1987; Cummins 1989, pp.87-113; Field 1977, 1978; Harman 1982; Loar 1982; McGinn 1982; and Schiffer 1987.

Nonetheless, while it is not always made clear (even by their proponents!) that these functional role theories rely on second-order resemblance,[20] it is clear that they are not appropriate for unpacking the representational content of connectionist representing vehicles. The moral of the previous subsection, remember, was that connectionism compels us to explain representational content in terms of resemblance relations between the *intrinsic* properties of the brain's representing vehicles and their target domains. Functional role theories don't do this;

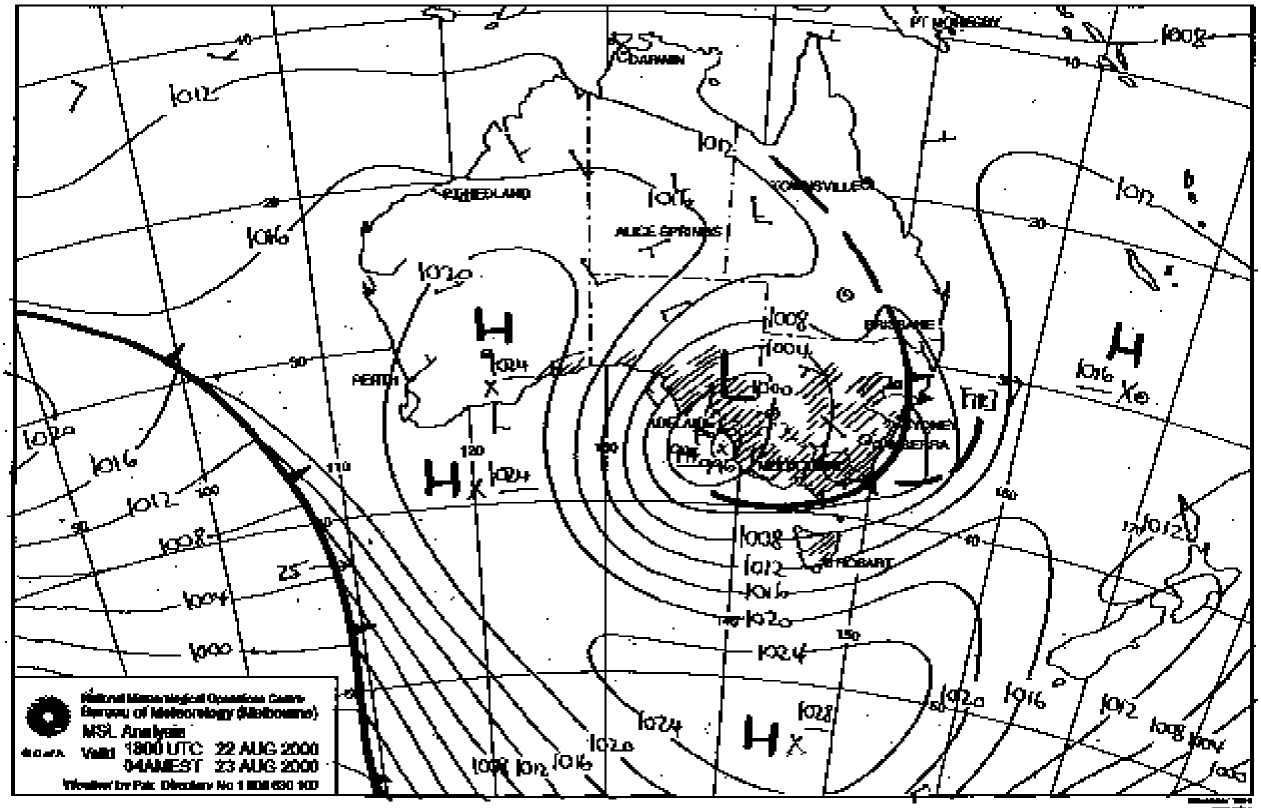

Figure 1. A weather map showing a low pressure cell over South Australia. The isobars around the low are closely spaced, indicating a steep pressure gradient.

they focus instead on the causal, and hence *extrinsic*, relations among a set of representing vehicles. So what kind second-order resemblance relation *is* appropriate for connectionist representing vehicles? We are finally in a position to provide an answer to this question.

3.3 *Structural* Resemblance and Connectionist Vehicles

There is another variety of second-order resemblance: one based on the *physical* relations among a set of representing vehicles. We will say that one system *structurally resembles* another when the physical relations among the objects that comprise the first preserve some aspects of the relational organisation of the objects that comprise the second. Structural resemblance is quite different from functional resemblance. What determines the functional/structural distinction is the way relations in the second system are preserved by the first: by *causal* relations in the case of functional resemblance, by *physical* relations in the case of structural resemblance. In neither case is there any restriction on the kinds of relations allowed in the second system—they can be relations among objects, properties or relations; they can be physical, causal or conceptual.[21]

[20] Cummins (1989, pp.114-25) and Von Eckardt (1993, pp.209-14) are important exceptions.

[21] Note that it follows from this that both functional and structural resemblance can be *asymmetric* relations: one system can functionally/structurally resemble a second, without the converse obtaining.

Structural resemblance grounds all the various examples of representation discussed in the last subsection. A surface map preserves spatial relations in the world via spatial relations among map points. Since spatial relations are a species of *physical* relations this clearly qualifies as an instance of representation grounded in structural resemblance. Likewise, the representing power of a mercury thermometer relies on a correspondence between one physical variable (the height of the column of mercury) and another (the temperature of bodies in contact with the thermometer). And in weather maps the relative spacing of isobars is employed to represent relative pressure gradients. In each of these cases we can identify a system of vehicles whose physical relations ground a (non-mental) representing relationship.

Most importantly in the context of this paper, structural resemblance would seem to be the right second-order resemblance relation for explaining the representational content of connectionist representing vehicles.[22] As an example consider Cottrell's face-recognition network (see Churchland 1995, pp.38-55, for discussion). This network has a three layer feed-forward architecture: a 64x64 input array, fully connected to a hidden layer of 80 units, which in turn is fully connected to an output layer comprising 8 units. Each unit in the input layer can take on one of 256 distinct activation values, so it is ideal for encoding discretised grey-scale images of faces and other objects. After squashing through the hidden layer these input patterns trigger three units in the output layer that code for face/non-face status and gender of subject, and five which encode arbitrary 5-bit names for each of 11 different individuals. Cottrell got good performance out of the network after training it on a corpus of 64 images of 11 different faces, plus 13 images of non-face scenes. He found that the network was: i) 100% accurate on the training set with respect to faceness, gender and identity (name); ii) 98% accurate in the identification of *novel* photos of people featured in the training set; and iii) when presented with entirely novel scenes and faces, 100% correct on whether or not it was confronting a human face, and around 80% correct on gender.

What is significant about the face-recognition network, for our purposes, is the way it codes faces at the hidden layer. Cluster analysis reveals that the network partitions its hidden unit activation space into face/non-face regions; within the face region into male/female regions; and then into smaller sub-regions corresponding to the cluster of patterns associated with each subject (see Figure 2). Within the face region each point is an abstract (because compressed) representation of a face. Faces that are similar are represented by points that are close together in the space, whereas dissimilar faces are coded by points that are correspondingly further apart. So the relations among faces which give rise to our judgments concerning similarity, gender, etc., are preserved in the distance relations in activation space.

All of this suggests that the face-recognition network supports a *structural* resemblance between activation patterns and the domain of human faces. Hidden unit activation space is a *mathematical* space used by theorists to portray the set of activation patterns a network is capable of producing over its hidden layer. Activation patterns themselves are physical objects (patterns of neural firing if realised in a brain), so distance relations in activation space actually codify *physical* relations among activation states. Consequently, the set of activation patterns generated across any implementation of Cottrell's face-recognition network constitutes a system of representing vehicles whose physical relations capture relations among human faces.

The foregoing would be moot if the structural resemblance embodied in Cottrell's network were causally inert. But it's this structural resemblance that, in effect, powers both the computational and the behavioural capacities of the network. Structural resemblance thus appears to ground representational content in any connectionist computational system. Accordingly, if the brain is a network of connectionist networks, as connectionism maintains,

[22] In this context, as in our earlier discussion of the connectionist vehicle theory of consciousness, connectionist networks are to be understood as *idealised* models of real neural networks, which, although unrealistic in certain respects, capture what may well be the key structural features whereby the brain represents its world.

then the representational content of its representing vehicles is determined by their physical relations and hence, ultimately, their intrinsic properties. A structural resemblance theory thus has the power to deliver exactly what the connectionist vehicle theory of phenomenal consciousness is committed to: an account of representational content which takes it to be an intrinsic property of the brain's explicit representing vehicles.

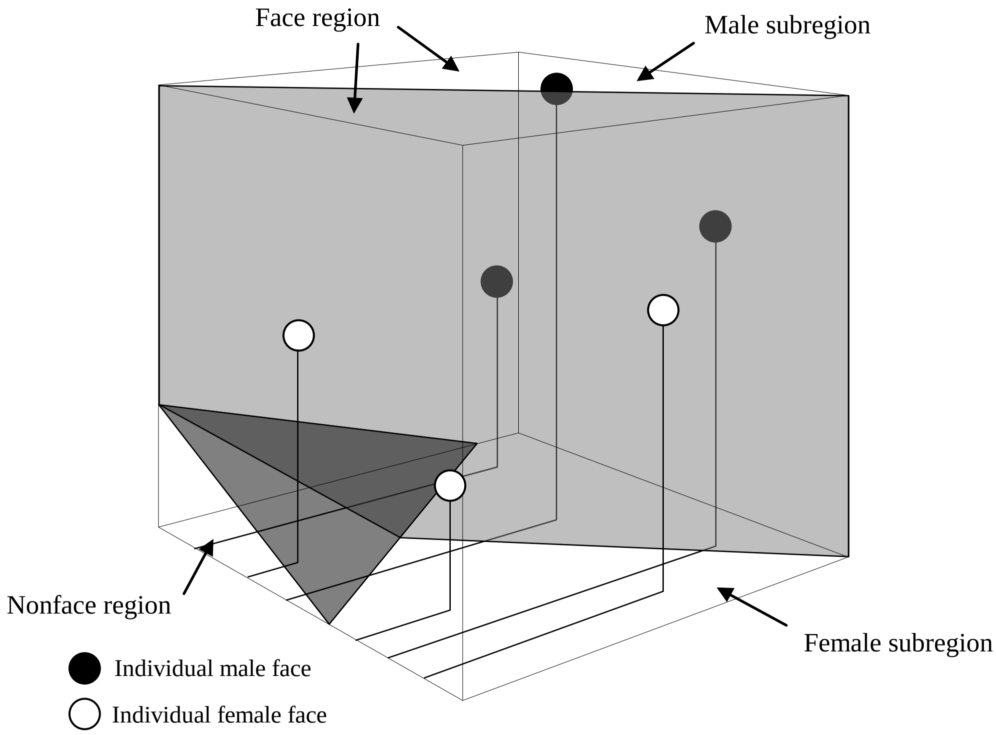

Figure 2. The hierarchy of learned partitions across the hidden unit activation space of Cottrell's face recognition network (after Churchland 1995, p.49).

4. The Phenomenal Mind

To identify conscious experience with the vehicles of explicit representation in the brain is to equate consciousness with intrinsic properties of certain (network) states of the brain. To further claim, as we do, that the phenomenal character of experience is none other than the representational content of these same vehicles is to claim that representational content is likewise intrinsic to these brain states. We've argued that connectionism allows us to seriously consider this double-barrelled conjecture, and that a structural resemblance theory of representational content, far from being merely consistent with a connectionist vehicle theory of consciousness, is actually mandatory for connectionism. But so far we haven't considered the implications of this hypothesis for the character and role of the phenomenal mind—the mind as given in first-person experience.

In the previous section we noted that the task of a theory of representational content is to explain how nervous systems can be in the representing business in the first place—how brain states can be about aspects of the world. We also noted that there appear to be just two different objective relations that the brain's representing vehicles are capable of bearing to the world, and which are therefore capable of forming the basis of a naturalistic explanation of representational content—namely, causation and resemblance. What we didn't say, however, is that the former carries a heavier existential burden than the latter: in order for a representing vehicle to bear a causal relation to some represented object, there must be a causal mechanism that connects them; whether a vehicle (or system of vehicles, in the second-order case) resembles an object, on the other hand, depends solely on the intrinsic properties of the vehicle (or system of vehicles).[23]

As a consequence of these different existential commitments, causal and resemblance theories yield different conceptions of the "aboutness" of representational content.[24] Aboutness, from the perspective of a causal theory, is a matter of the brain's representing vehicles being plugged into the world in an appropriate fashion. To represent the world is to be causally connected to it. Resemblance theories, by contrast, ground representational content in properties that are intrinsic to the brain. To represent the world is to construct an internal model of it.

This last point has some important implications for our understanding of the phenomenal mind. The proposal being scouted in this paper is that the phenomenal character of consciousness is identical to the representational content of the brain's explicit representing vehicles. However, representational content, according to a structural resemblance theory, is not a matter of causal connection; it is a matter of resemblance grounded in structural properties of the brain's representing vehicles. Thus, although our conscious experiences seem to place us in direct contact with the world outside our brains, they actually do no such thing. Conscious experiences are really internal models of the world.

From this perspective, the world of experience—our rich and highly structured phenomenology of body and physical environment—is a "virtual reality" constructed by the brain and "projected outwards" (Revonsuo 1995, 1997, 2000). To the extent that this virtual reality (the world of *appearance*) mirrors the actual world (the world of *reality*), the behaviour it drives will be appropriate. But where the appearances are seriously out of step with reality (as, for example, in dreams and hallucinations), the behaviour that results will be correspondingly anomalous.[25]

What the structural resemblance theory of representational content thus demonstrates is how the phenomenal mind can be causally relevant to cognition. In summary form the story goes like this. On a structural resemblance account, representing vehicles come by their contents by virtue of their local physical/structural properties. Since it is such local properties that ultimately determine the course of cognition, the structural resemblance account satisfies what we have elsewhere called the *causal constraint* on a theory of mental representation (O'Brien & Opie Forthcoming): the requirement that representational contents actually causally drive behaviour. But if the phenomenal character of conscious experience is constituted by the contents of the brain's explicit representing vehicles, as we conjecture, then we get the following significant implication: that the complex phenomenal character of experience turns out to be none other than the content bearing structure of the physical vehicles that actually shape cognition. On this picture, the phenomenal mind amounts to a powerful *representational medium:* a complex of internal analogues which, by virtue of their physical structure, actively shape thought processes and effector responses, and thereby our behaviour.[26]

---

[23] This point was highlighted by Charles Sanders Peirce (as reported in Von Eckardt 1993, pp.150-1). Moreover, the essential difference between causation and resemblance in this regard survives intact even when it is acknowledged that, since the brain's vehicles are capable of representing the non-existent, it must be possible for the former to bear causal/resemblance relations to the latter. The standard move here is to invoke counterfactuals: a representing vehicle can be about a non-existent object so long as were the latter to exist, the former would bear the appropriate causal/resemblance relation to it (see, e.g., Fodor 1990, pp.100-1)

[24] As Cummins points out, exactly what representational content consists in can't be stipulated in advance of a mature theory of mental representation, since the theory must be allowed to speak for itself (1989, p.12).

[25] An obvious example is the behaviour of schizophrenics when they are subject to hallucinations and delusions. Another is the behaviour of people who suffer from REM sleep behaviour disorder (Revonsuo 2000, p.66).

[26] Our hypothesis therefore exorcises the spectre of epiphenomenalism, which haunts most recent accounts of representational content and consciousness.

What may we surmise about the detailed character of the phenomenal mind on the basis of these considerations? Simply that it must be intricate enough and flexible enough to support a wide range of (dynamic) resemblance relations with behaviourally salient states and processes of our bodies and the environment. This suggests that empirical phenomenology should not only look inwards, to the world of appearance, but also outwards, to the scientifically apprehended structure of reality, in order to gather clues as to the nature of experience. It is important to remember, however, that the virtual reality which guides us from moment to moment, for all that it resembles the actual world in many respects, clearly has properties of its own—it is full of idealisations, approximations, and errors. For this reason our primary method in exploring the phenomenal mind must remain the careful study of its neural correlates and the behaviour it shapes (including introspective reports), under controlled stimulus conditions. Our ultimate ambition should be to produce a detailed map of experience that can be related at all levels to the structural properties of the brain.

References

Atkinson, A.P., Thomas, M.S.C. & Cleeremans, A. (2000). Consciousness: Mapping the Theoretical Landscape, *Trends in Cognitive Sciences*, 4(10): 372-382.

Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

Bechtel, W. & Abrahamsen, A.A. (1991). *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. B. Blackwell.

Blachowicz, J. (1997). Analog Representation Beyond Mental Imagery, *Journal of Philosophy*, 94(2): 55-84.

Block, N. (1986). Advertisement for a Semantics for Psychology. In French, P.A. (ed.), *Midwest Studies in Philosophy*. Univ Minnesota Pr.

Block, N. (1987). Functional Role and Truth Conditions, *Proceedings of the Aristotelian Society*, 61: 157-181.

Churchland, P.M. (1995). *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*. MIT Press.

Churchland, P.S. & Sejnowski, T.J. (1992). *The Computational Brain*. MIT Press.

Clapin, H. & O'Brien, G. (1998). A Conversation About Superposition and Distributed Representation, *Noetica: Open Forum*, 3(10).

Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. MIT Press.

Clark, A. (1993). *Associative Engines: Connectionism, Concepts, and Representational Change*. MIT Press.

Crick, F. (1984). Function of the Thalamic Reticular Complex: The Searchlight Hypothesis, *Proc Natl Acad Sci U S A*, 81(14): 4586-90.

Cummins, R. (1986). Inexplicit Representation. In Brand, M. & Harnish, R.M. (eds.), *The Representation of Knowledge and Belief.* University of Arizona Press.

Cummins, R. (1989). *Meaning and Mental Representation*. MIT Press.

Cummins, R. (1996). *Representations, Targets, and Attitudes*. MIT Press.

Cummins, R. & Schwarz, G. (1991). Connectionism, Computation and Cognition. In Horgan, T. & Tienson, J. (eds.), *Connectionism and the Philosophy of Mind*. Kluwer.

Dennett, D. (1982). Styles of Mental Representation, *Proceedings of the Aristotelian Society, New Series*, 83: 213-226.

Dennett, D.C. (1984). Cognitive Wheels: The Frame Problem of AI. In Hookway, C. (ed.), *Minds, Machines and Evolution*. Cambridge Univ Pr.

Dennett, D.C. (1991). *Consciousness Explained*. Little, Brown.

Dietrich, E. (1989). Semantics and the Computational Paradigm in Cognitive Psychology, *Synthese*, 79: 119-141.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Blackwell.

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

Dretske, F. (1993). Conscious Experience, *Mind*, 102(406): 263-283.

Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.

Field, H. (1977). Logic, Meaning, and Conceptual Role, *Journal of Philosophy*, 74: 379-409.

Field, H. (1978). Mental Representation, *Erkenntnis*, 13: 9-61.

Flanagan, O. (1992). *Consciousness Reconsidered*. MIT Press.

Fodor, J. (1975). *The Language of Thought*. Harvester Press.

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.

Fodor, J. (1990). *A Theory of Content and Other Essays*. MIT Press.

Fodor, J. & Pylyshyn, Z.W. (1988). Connectionism and Cognitive Architecture: A Critical Analysis, *Cognition*, 28: 3-71.

Gardenfors, P. (1996). Mental Representation, Conceptual Spaces and Metaphors, *Synthese*, 106(1): 21-47.

Harman, G. (1982). Conceptual Role Semantics, *Notre Dame Journal of Formal Logic*, 23: 242-256.

Haugeland, J. (1981). Semantic Engines: An Introduction to Mind Design. In HAUGELAND, J. (ed.), *Mind Design*. MIT Press.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. MIT Press.

Hoffman, D.D. (1998). *Visual Intelligence: How We Create What We See*. Norton.

Horgan, T. & Tienson, J. (1996). *Connectionism and the Philosophy of Psychology*. MIT Press.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. MIT Press.

Johnson Laird, P.N. (1988). A Computational Analysis of Consciousness. In Marcel, A.J. & Bisiach, E. (eds.), *Consciousness in Contemporary Science*. Clarendon Press/Oxford University Press.

Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press.

Johnson-Laird, P.N. (1988). *The Computer and the Mind: An Introduction to Cognitive Science*. Fontana.

Kihlstrom, J.F. (1987). The Cognitive Unconscious, *Science*, 237: 1445-1452.

Lloyd, D. (1989). *Simple Minds*. MIT Press.

Loar, B. (1982). Conceptual Role and Truth-Conditions, *Notre Dame Journal of Formal Logic*, 23: 272-283.

McGinn, C. (1982). The Structure of Content. In Woodfield, A. (ed.), *Thought and Object: Essays on Intentionality*. Clarendon Pr.

Millikan, R.G. (1984). *Language, Thought and Other Biological Categories.* MIT Press.

O'Brien, G. (1998). The Role of Implementation in Connectionist Explanation, *Psycoloquy*, 9(6).

O'Brien, G. (1999). Connectionism, Analogicity and Mental Content, *Acta Analytica*: 111-131.

O'Brien, G. & Opie, J. (1997). Cognitive Science and Phenomenal Consciousness: A Dilemma, and How to Avoid It, *Philosophical Psychology*, 10(3): 269-286.

O'Brien, G. & Opie, J. (1999a). A Connectionist Theory of Phenomenal Experience, *Behavioral and Brain Sciences*, 22(1): 127-148.

O'Brien, G. & Opie, J. (1999b). Finding a Place for Experience in the Physical-Relational Structure of the Brain, *Behavioral and Brain Sciences*, 22(6): 966-7.

O'Brien, G. & Opie, J. (1999c). Putting Content into a Vehicle Theory of Consciousness, *Behavioral and Brain Sciences*, 22(1): 175-196.

O'Brien, G. & Opie, J. (Forthcoming). Notes toward a Structuralist Theory of Mental Representation. In Clapin, H., Staines, P. & Slezak, P. (eds.), *Representation in Mind: New Approaches to Mental Representation*. Greenwood.

Opie, J. (1998.). Connectionist Modelling Strategies, *Psycoloquy*, 9(30).

Palmer, S. (1978). Fundamental Aspects of Cognitive Representation. In E.Rosch & B.Lloyd (eds.), *Cognition and Categorization*. Lawrence Erlbaum.

Palmer, S.E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press.

Pylyshyn, Z.W. (1984). *Computation and Cognition : Toward a Foundation for Cognitive Science*. MIT Press.

Revonsuo, A. (1995). Consciousness, Dreams, and Virtual Realities, *Philosophical Psychology*, 8: 35-58.

Revonsuo, A. (1997). How to Take Consciousness Seriously in Cognitive Neuroscience, *Communication and Cognition*, 30: 185-206.

Revonsuo, A. (2000). Prospects for a Scientific Research Program on Consciousness. In Metzinger, T. (ed.), *Neural Correlates of Consciousness*. MIT Press.

Rumelhart, D.E. & McClelland, J.L. (1986). *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*. MIT Press.

Schiffer, S.R. (1987). *Remnants of Meaning*. MIT Press.

Sejnowski, T. & Rosenberg, C. (1987). Parallel Networks That Learn to Pronounce English Text, *Complex Systems*, 1: 145-68.

Shepard, R. & Chipman, S. (1970). Second-Order Isomorphism of Internal Representations: Shapes of States, *Cognitive Psychology*, 1: 1-17.

Shepard, R.N. & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects, *Science*, 171(972): 701-3.

Smolensky, P. (1987). The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn, *Southern Journal of Philosophy*, 26: 137-161.

Swoyer, C. (1991). Structural Representation and Surrogative Reasoning, *Synthese*: 449-508.

Tienson, J. (1987). Introduction to Connectionism, *Southern Journal of Philosophy*, 26: 1-16.

Tye, M. (1995a). A Representational Theory of Pains and Their Phenomenal Character. In Tomberlin, J.E. (ed.), *AI, Connectionism and Philosophical Psychology*. Ridgeview.

Tye, M. (1995b). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.

van Gelder, T. (1990). Compositionality: A Connectionist Variation on a Classical Theme, *Cognitive Science*, 14(3): 355-384.

Velmans, M. (1991). Is Human Information Processing Conscious?, *Behavioral and Brain Sciences*, 14: 651-726.

Von Eckardt, B. (1993). *What Is Cognitive Science?* MIT Press.