Philosophy and Phenomenological Research
Vol. LIX, No. 4, December 1999

A Defense of Cartesian Materialism

GERARD O'BRIEN AND JONATHAN OPIE
The University of Adelaide

One of the principal tasks Dennett sets himself in *Consciousness Explained* is to demolish the Cartesian theater model of phenomenal consciousness, which in its contemporary garb takes the form of *Cartesian materialism*: the idea that conscious experience is a *process of presentation* realized in the physical materials of the brain. The now standard response to Dennett is that, in focusing on Cartesian materialism, he attacks an impossibly naive account of consciousness held by no one currently working in cognitive science or the philosophy of mind. Our response is quite different. We believe that, once properly formulated, Cartesian materialism is no straw man. Rather, it is an attractive hypothesis about the relationship between the computational architecture of the brain and phenomenal consciousness, and hence one that is worthy of further exploration. Consequently, our primary aim in this paper is to defend Cartesian materialism from Dennett's assault. We do this by showing that Dennett's argument against this position is founded on an implicit assumption (about the relationship between phenomenal experience and information coding in the brain), which while valid in the context of *classical* cognitive science, is not forced on *connectionism*.

1. Introduction

One of the principal tasks Dennett sets himself in *Consciousness Explained* (1991) is to demolish the Cartesian theater model of phenomenal consciousness.[1] This model treats our phenomenal experience as the result of some kind of *presentation process*, wherein the objects of consciousness are "displayed" for the delectation of the "mind's eye". According to Dennett, while contemporary theorists reject Descartes' dualism, the theater imagery continues to shape their thinking. These days it takes the form of *Cartesian materialism*: the idea that conscious experience is a special process of presentation somehow realized in the physical materials of the brain. This, Dennett thinks, "is the most tenacious bad idea bedeviling our attempts to think about consciousness" (1991, p. 108); one whose hold on the imaginations of philosophers and cognitive scientists alike has prevented any major theoret-

[1] In speaking of 'phenomenal consciousness' our intended target is neither *self-consciousness* nor what has come to be called *access-consciousness* (see Block, 1993, 1995). It is, rather, *phenomenal experience*: the "what it is like" of experience (Nagel, 1974). We will speak variously of 'phenomenal experience', 'phenomenal consciousness', 'conscious experience', or sometimes just plain 'consciousness', but in each case we refer to the same thing.

ical advance in our understanding of this most elusive and puzzling phenomenon. Dennett consequently spends a good deal of time in his book both developing a detailed argument against Cartesian materialism, and constructing the foundations of an alternative model, the *multiple drafts* theory of consciousness, that can stand in its stead.

The now standard response to Dennett's project is that he has picked a fight with a straw man. Cartesian materialism, it is alleged, is an impossibly naive account of phenomenal consciousness held by no one currently working in cognitive science or the philosophy of mind. Consequently, whatever the effectiveness of Dennett's demolition job, it is fundamentally misdirected (see, e.g., Block, 1993, 1995; Shoemaker, 1993; and Tye, 1993).

Our response is quite different. We believe that, once properly formulated, Cartesian materialism is an attractive hypothesis about the relationship between the computational architecture of the brain and phenomenal consciousness, and hence one that is worthy of further exploration. Consequently, our primary aim in this paper is to defend Cartesian materialism from Dennett's assault.

The paper has the following structure. In the next section, after clarifying what Cartesian materialism amounts to, we offer a brief exposition of Dennett's main argument against this position. In section three, we highlight a crucial assumption implicit in Dennett's reasoning, namely, that phenomenal experience cannot be identified with the explicit representation of information in the brain (an assumption we'll call the *dissociation thesis*). This assumption, we argue in section four, is indeed valid in the context of *classical* cognitive science, and hence, if the rest of Dennett's argument goes through, classicists cannot be Cartesian materialists. However, the assumption is not valid in the context of *connectionist* cognitive science, we claim in section five, because connectionists can plausibly adopt an account of conscious experience that aligns the contents of consciousness with the vehicles of explicit information coding in the brain. Connectionists are thus able to embrace a version of Cartesian materialism that is immune to Dennett's attack.

2. Dennett's Demolition of the Cartesian Theater

Dennett initially characterizes Cartesian materialism as "the view you arrive at when you discard Descartes' dualism but fail to discard the imagery of a central (but material) Theater where 'it all comes together' " (1991, p. 107). He continues:

> [It] is the view that there is a crucial finish line or boundary somewhere in the brain, marking a place where the order of arrival equals the order of 'presentation' in experience because what happens there is what you are conscious of. (1991, p. 107)

But one immediately wonders about the way Dennett characterizes his target here. In particular, since it is a commonplace of current neuropsychology that there is no good evidence for a spatially localized seat of consciousness, and since there are so many cognitive scientists who explicitly deny any commitment to a single, central theater, it is puzzling that Dennett makes this *centralism* such a feature of his discussion.

Dennett readily admits that practically no-one working in cognitive science today is likely to acknowledge a centralist view of consciousness (1993, p. 107). His dialectic suggests that he actually has *two* audiences in mind: a naive group, who need to be convinced that the idea of a central theater is no good, by showing them the unhappy consequences of this view; and a better informed group, who explicitly deny belief in a central theater, but whose "hopes and hunches and arguments...betray a sneaking and unrecognized commitment to the view" (1993, p. 920).

Let us suppose, at the risk of being immodest, that we are in well informed company here. What, then, is the basis for the claim that many cognitive scientists, despite their protestations to the contrary, are closet centralists? Clearly the answer we get is likely to vary from case to case, but in the exchange below Tye elicits what would probably be a typical response from Dennett. Tye writes:

> Upon further examination, it appears that Dennett's real motive in attacking Cartesian materialism is that it goes hand in hand (he thinks) with the view that there really is such a thing as phenomenal experience or consciousness, conceived of as distinct from judgment or belief. But the latter view is surely independent of the former. Advocates of the reality of phenomenal seemings and feelings, as apart from judgments and beliefs, are not committed to holding that there is a single place in the brain in which the seemings and feelings occur. (1993, p. 893)

Dennett responds:

> When Tye says that advocates of the reality of phenomenal seemings "are not committed to holding that there is a single place in the brain in which the seemings and feelings occur" this is really just because these advocates have not developed their allegiance to this reality beyond the vaguest of handwavings. Presumably "phenomenal seemings" are seemings-to-me (as opposed to the non-phenomenal seemings that are just seemings-to-my-retina or seemings-to-some-unconscious-control-process), but making this distinction requires that one identify a privileged neural medium – the *Me*dium, you might call it – transduction into which marks the onset of true phenomenal seeming. (1993, p. 920)

This reply is revealing, for it is not insistence on a single, central theater of consciousness that emerges here as the characteristic feature of Cartesian materialism, but rather, belief in a special neural medium with a constitutive role in conscious onset. What is more, Dennett accepts that this medium "might be anatomically spread around" (1993, p. 920). Consequently, Dennett in effect grants Tye's claim that as a phenomenal realist he is not committed to a single place in the brain at which phenomenal seemings occur.

Thus, Dennett's initial characterization of Cartesian materialism looks inadequate, but notice that he has provided a suitable replacement. What Cartesian materialists are committed to is a "privileged neural medium" such that "when information eventually gets transduced into that medium, it enters conscious experience" (1993, pp. 920–21).[2] Whether this medium is spatially localized, or distributed around the entire brain (i.e., whether there be a single theater or multiple minitheaters) is really beside the point. What is essential to the position is that there are points in time, and places in space, at which the onset of consciousness can be precisely located.

Henceforth, we will take the following to properly characterize Cartesian materialism: conscious experience results from a (possibly distributed) process of presentation; a process whereby informational contents enter consciousness as a result of being encoded in a special neural medium (wherever in the brain this medium is realized). So characterized, Cartesian materialism is clearly no straw man. Indeed, it is an initially plausible and attractive empirical hypothesis about the relationship between structures in the brain and phenomenal experience.

Let's now turn to a discussion of Dennett's principal argument against Cartesian materialism. While Dennett develops the argument over a couple of chapters (1991, Chaps. 5–6), its overall structure is very simple and takes the following form:

1. If Cartesian materialism were true (if there were a special medium of consciousness the transduction of information into which marks the onset of phenomenal experience), then in principle it would be possible to determine exactly when and where the onset of phenomenal experience occurs in the brain.

2. It is, in principle, impossible to determine exactly when and where the onset of phenomenal experience occurs in the brain.

Therefore,

3. Cartesian materialism is false.

Dennett takes the second premise of the argument to be its most crucial. In support of this premise, he considers a number of empirical studies concerning the onset of various kinds of phenomenal experience. The *color phi phenomenon* (1991, pp. 119–26) comes in for particular attention. It is well known that reports of apparent motion can be induced in subjects who are

---

2 This way of putting things is due to Mangan (1993a).

exposed to tachistoscopically presented spots of light that are spatially separated yet stationary, if the second spot is illuminated very soon after the first (i.e., with a delay of about 200 msec). In a variation on this experimental paradigm, where the spots of light are of different colors, subjects report that a single moving spot of light changes color mid-movement (e.g., a red spot changes into a green spot somewhere between the end-points). The addition of color makes the phenomenon more striking, but doesn't alter the essential puzzle, which can be summed up like this: How is it that subjects perceive intervening motion, given that a motion signal can't possibly be generated until after the second (stationary) spot has been registered in some way?

Dennett argues that if Cartesian materialism were true, then two quite different explanations of this phenomenon would be possible, and must be distinguished. The first possibility is that the brain receives information about the second (green) spot prior to the transduction of information about the stationary red spot into the medium of consciousness. A story concerning a moving spot that changes from red to green mid-movement is then quickly concocted, which is the only information actually transduced into this medium. On this account the subject has a phenomenal experience of such a moving spot. Dennett calls this a *Stalinesque* revision, as there is a sense in which it involves a "show trial" of something that never really happened. The second possibility is that the brain receives information concerning the green spot after the transduction of information about the stationary red spot into the medium of consciousness, and faithfully transduces this second content, such that the subject has a phenomenal experience of two spatially separated and differently colored spots. But then a post-experiential process of revision very rapidly occurs in the brain (in order to make sense of what has just occurred) and the subject's memory of this incident becomes the story of a moving spot, complete with color change. Since this revision happens extremely quickly (in a fraction of a second), by the time the subject comes to report their experience (even to themselves), they say and believe they only ever saw a moving spot which changed color mid-movement. No trace is left of the initial experience. Dennett calls this second possibility an *Orwellian* revision, since it involves contamination of the true historical record.

The Cartesian materialist, thinks Dennett, is committed to this distinction between Stalinesque and Orwellian revisions. In the case of the phi phenomenon, the Cartesian materialist must be capable of determining (in principle) which of these two kinds of revisions actually occurred (transduction of information into the neural medium of consciousness presumably being a precisely datable event). But – and here's the rub – both models, according to Dennett, can account for *all* the data. "From the inside" the subject cannot distinguish between the two cases, as both accounts agree that "subjects should be unable to tell the difference between misbegotten experiences and immediately misremembered experiences" (1991, p. 125). Moreover, scientists (observing "from the outside") can't determine which revision occurred, because the two models look the same from the computational perspective: both accounts involve, first, content fixations which represent information about stationary spots, followed rapidly by further content-fixations representing moving spots with a color change. Which of these contents is phenomenally experienced? The Orwellian theorist says that they all are, while the Stalinesque theorist believes that only the latter (moving spot) contents enter consciousness. However, in order to answer this question both theories must posit precise moments of conscious onset (points in time, and hence places in space, when contents are transduced into the special neural medium of consciousness) and these are times "whose fine-grained location is nothing that subjects can help them to locate, and whose location is also neutral with respect to all other features of their theories" (1991, p. 125). Where Stalinesque and Orwellian models differ is in the placement of a boundary of consciousness which they are powerless to identify.

The upshot, according to Dennett, is that Cartesian materialism commits us to "a difference that makes no difference" (1991, p. 125). And if there isn't any *real* difference between these two, then it's impossible to determine exactly when and where the onset of phenomenal experience occurs in the brain. And if this is impossible, then Cartesian materialism must be a radically mistaken conception of phenomenal experience – there is no special medium in the brain the transduction of information into which marks the onset of consciousness.

3. The Dissociation Thesis: An Implicit Assumption In Dennett's Argument

In order to defend Cartesian materialism, we're not going to confront Dennett's argument head-on. This is because we don't need to. Instead we will focus on an assumption about the relationship between phenomenal experience and mental representation that Dennett simply takes for granted. In this section we will unearth the assumption. In subsequent sections we will examine its validity.

A crucial step, perhaps *the* crucial step, in Dennett's argument against Cartesian materialism is his assertion that scientists "from the outside" cannot distinguish between Orwellian and Stalinesque revisions. This is a claim not just about the current capacities of our neuroscientists equipped with all the latest neuro-scanning and psychometric technology. It is a claim about what, *in principle,* it is possible to determine by looking into the brain. He asks us to imagine, for example, that in some future golden age of neuroscience we have "truly accurate information (garnered from various brain-scanning technologies) about the exact 'time of arrival' or 'creation' of every representing, every vehicle of content, anywhere in your nervous system" (1991, p. 124). Even with such wonderful neuro-technology in place, he tells us, scientists couldn't distinguish between Orwellian and Stalinesque revisions, because while they could delimit various episodes of content-fixation, they couldn't determine which of these occur in the special medium of consciousness:

> Both [Orwellian and Stalinesque] theorists have *exactly* the same theory of what happens in your brain; they agree about just where and when in the brain the mistaken content [e.g., a moving, color-changing spot] enters the causal pathways; they just disagree about whether that location is to be deemed pre-experiential or post-experiential. (1991, p. 125)

From the neurocomputational perspective (i.e., from a perspective which treats human cognition in terms of the creation and manipulation of content-bearing representational vehicles) the two stories look identical, according to Dennett; first off the visual-processing production line is a vehicle whose content is *stationary-red-spot*, then along comes another vehicle with the content *stationary-green-spot (a-little-to-the-right)*, and finally a third vehicle with the more complex content *moving-red-spot-changing-to-green*. However, the Cartesian materialist is committed to something further: a boundary across which a representational vehicle must pass in order for its content to "enter" consciousness. And this, Dennett argues, is a boundary for which the peering neuroscientist can have no evidence.

But there is a fairly obvious assumption implicitly at work here. The assumption is that we can't entertain an account that *identifies* conscious experience with the explicit representation of information in the brain. Were such an account plausible, then the fact that scientists can (in principle) distinguish between different episodes of content-fixation – and Dennett confirms that "these spatially and temporally distributed content-fixations...are precisely [located] in both space and time" (1991, p. 113) – entails that they would also be capable of isolating the special medium in the brain, transduction into which marks the onset of consciousness. On such an account the very "creation" of a vehicle of content would be a transduction into this special medium. As a consequence the difference between Stalinesque and Orwellian revisions would be readily detectable. In the case of the color phi phenomenon, for example, if the observing neuroscientists discover that the visual cortex, in the course of processing the tachistoscopically presented stimuli, generates a representational vehicle with the content *stationary-green-spot (a-little-to-the-right)* following one with the content *stationary-red-spot* and preceding another whose content is *moving-red-spot-changing-to-green* (in keeping with Dennett's abovementioned account of the neurocomputational dynamics), they would immediately know the revision is Orwellian. On the other hand, if no such representational vehicle is found to intervene between these other two, i.e., if the visual cortex fails to *explicitly* register the existence of a second stationary spot, then the revision is Stalinesque.

We will give Dennett's assumption a name: we'll call it the *dissociation thesis*. This name is apt because the assumption is that phenomenal consciousness and explicit mental representation can come apart, i.e., they are *dissociable*. More specifically, the dissociation thesis is the claim that phenomenal experience cannot be identified with the vehicles that explicitly encode information in the brain, and hence that the mere creation of such a vehicle is not sufficient for its content to be conscious. Consciousness, on this assumption, requires something more. And the whole thrust of Dennett's argument is to close off the theoretical possibility that this something more involves a *further, special* neural medium of representation, into which information must be re-coded in order to be phenomenally experienced.

It is undoubtedly the case that the dissociation thesis is well-entrenched in contemporary cognitive science and philosophy of mind; so well-entrenched that Dennett obviously feels he can take it for granted. It is sheer orthodoxy to hold that our brains explicitly encode and process far, far more information than we ever phenomenally experience at any one moment in time. Indeed, the theoretical focus on the unconscious has become so extreme that Lashley is willing to assert: "*No activity of mind is ever conscious*" (1956, his italics) and, barely less extravagant, Fodor claims that "practically all psychologically interesting cognitive states are unconscious..." (1983, p. 86). However, we must be careful here. Almost all the theorizing about consciousness in recent cognitive science and philosophy of mind has been either explicitly or implicitly conditioned by what is now termed the *classical* computational theory of mind – the doctrine that takes human cognition to be a species of *symbol manipulation*. But as everyone knows, classicism now has an increasingly popular rival in the form of *connectionism*. And what is unclear is whether theorizing about phenomenal consciousness from this non-classical quarter should also be constrained by the dissociation thesis.

In what follows we'll correct this deficiency by considering the relationship between consciousness and explicit representation in both the classical and connectionist contexts. What we'll find is that, unlike classicism, connectionism has a computational ontology rich enough to support the alignment of phenomenal consciousness with the explicit representation of information in the brain. This means that a connectionist need not acquiesce in the dissociation thesis, and so need not abandon Cartesian materialism. We begin with classicism.

4. Classicism and the Dissociation Thesis

The classical computational account of cognition is quite simple to state: human cognitive processes are digital computational processes. What this actually entails about human cognition, however, is a long story, but fortunately one that is now very familiar.[3] The classicist takes the generic computational theory of mind (the claim that cognitive processes are disciplined operations defined over neurally realized representational states), and adds to it a more precise account of both the representational vehicles involved (they are complex symbol structures possessing a combinatorial syntax and semantics) and the nature of the computational processes (they are syntactically-governed transformations of these symbol structures). Our task is to discover what implications this story about human cognition has for the relationship between phenomenal experience and explicit mental representation.

In the classical context the vehicles of explicit mental representation are symbols tokened in the so-called *language of thought* (see Fodor, 1975). A classical theory that denied the dissociation thesis, therefore, would hold that whenever a symbol in the language of thought is tokened, the content of that representation is phenomenally experienced. It would hold, in other words, that contents moving across the unconscious/conscious boundary are distinguished by their transduction into the symbolic medium of the language of thought. And it would hold that whenever information is causally implicated in cognition, yet not consciously experienced, such information is not explicitly encoded in the form of a physical symbol structure.

However, a moment's reflection suffices to show that a classicist can't really contemplate this kind of account of phenomenal experience. 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.[4] In standard explanations of decision making, for example, the classicist makes constant reference to beliefs and goals that have a causal role in the decision procedure. It is also manifest that *most* of the information guiding this process is not phenomenally conscious. That is, many of the conscious steps in a decision process implicate a complex economy of unconscious beliefs interacting according to unconscious rules of inference. While it is possible that all the *rules of inference* are not represented symbolically in the language of thought (as Fodor keeps reminding us, these rules can be "emergents...out of hardware structures" (1987, p. 25)), the mediating train of unconscious beliefs must *interact* to produce their effects,

---

3 The more prominent contemporary philosophers and cognitive scientists who advocate a classical conception of cognition include Chomsky (1980), Field (1978), Fodor (1975, 1981, 1987), Harman (1973), Newell (1980), Pylyshyn (1980, 1984), and Sterelny (1990). For those readers unfamiliar with classicism, a good entry point is provided by the work of John Haugeland (1981, 1985, especially Chaps. 2 and 3).

4 This fact about ourselves has been made abundantly clear by research in the field of artificial intelligence, where practitioners have discovered to their chagrin that getting computer-driven robots to perform even very simple tasks requires not only an enormous knowledge base (the robots must know a lot about the world) but also a capacity to very rapidly access, update and process that information. This becomes particularly acute for AI when it manifests itself as the *frame problem*. See Dennett (1984) for an illuminating discussion.

else we don't have a causal explanation. But the only model of causal interaction available to a classicist involves symbol manipulations. So, either the unconscious includes such symbolic representational vehicles, or there are no plausible classical explanations of higher cognition.

There is a further difficulty for this version of classicism, namely: it provides no account whatever of learning. While we can assume that some of our intelligent behavior comes courtesy of endogenous factors, a large part of our intelligence is the result of a long period of learning. A classicist typically holds that learning (as opposed to development or maturation) consists in the fixation of beliefs via the generation and confirmation of hypotheses. This process must be largely unconscious, since much of our learning proceeds in the absence of conscious hypothesis testing. As above, this picture of learning requires an interacting system of unconscious representational vehicles, and, for a classicist, this means symbols tokened in the language of thought. To reject this picture is to abandon cognitive explanations of learning, because the alternative is to regard learning merely as a process that reconfigures the brain's functional architecture.[5] And any classicist who claims that learning is non-cognitive, is a classicist in no more than name.

The upshot of all of this is that any remotely plausible classical account of human cognition is committed to a vast amount of *unconscious symbol manipulation*. Consequently, classicists can't accept that the only symbolically represented information in the brain is that which is associated with our phenomenal experience, and hence must embrace the dissociation thesis.

Before proceeding to consider whether connectionist cognitive science is in a similar position in this regard, we'll pause to consider in some detail a possible objection to the preceding line of argument. For it might be charged that we have reached this conclusion about classicism by relying on an insufficiently fine-grained taxonomy of classical styles of mental representation. In particular, it might be charged that we have failed to consider the possibility that among the classical vehicles of explicit representation there are distinct *types* (i.e., distinct subspecies of symbol structures) one of which might plausibly be aligned with phenomenal experience. If such a subspecies can be found, then there is room for the classicist to deny the dissociation thesis *with respect to this type of representational vehicle*, and hence provide a bulwark against Dennett's assault.[6]

---

5 The *functional architecture* of a computational device comprises its repertoire of primitive operations as determined by its "hard-wired" physical structure (see Pylyshyn, 1980, 1984). A functional architecture determines, among other things: which symbol structures are primitive; how memory is organized and accessed; how process control is handled; and the extent of information passing that is permissible between modules or routines.

6 We are indebted to an anonymous reviewer of this journal for bringing this objection to our attention.

To begin with, it is important to recognize the tentative character of all current classical conjectures about the representational vehicles involved in human cognition. In particular, the range and nature of the primitive mental symbols, and the kinds of symbol structures built out of them, is still very much up for grabs. What lends weight to this objection, however, is the observation that both conventional computer science and classically-inspired cognitive psychology appear to distinguish between different kinds of representational vehicles. Computer scientists, for example, routinely distinguish between various "data types", ranging from the *simple* (e.g., Boolean, Character, String, Integer, Floating-Point) to the *composite* (e.g., Array, Record, List, Tree) (see, e.g., Scragg 1992); while cognitive psychologists, in their efforts to grapple with human cognition and perception, posit different kinds of representations associated with factual/declarative knowledge (*semantic networks, frames* and *scripts*), perception, (e.g., Marr's *primal sketch* and *2 D sketch* (1982)), imagery (e.g., the *surface representations, literal encodings* and *propositional encodings* posited by Kosslyn (1980)), and the control of automatic skills (*productions*), to name but a few. The suggestion is that the classicist might be able to exploit these distinctions to provide a *vehicle* criterion for distinguishing all those contents that are conscious from those that are not.[7]

Consider, then, some of the data structures in the repertoire of computer science (for a general discussion of different data types, see Von Eckardt 1993, Chap. 5). An *array* is an n-dimensional data structure consisting of an ordered set of elements all of which belong to the same type (simple or composite). An *array type* is fully specified once its base type and number of components have been declared. For example:

```
type Vector: array [1..3] of float ;

type Matrix: array [1..5] of Vector ;
```

[7] There are some obvious initial difficulties that beset this suggestion, at least with respect to the representations developed by cognitive psychologists. First, many of these representational types cut right across the conscious/unconscious boundary. Frames, scripts and semantic networks, for example, are associated both with accounts of long-term declarative memory, and with spreading activation models of (propositional) consciousness (see Stillings et al. 1995, pp. 26–32). A second problem concerns the intended *range* of application of these representations. Cognitive psychologists frequently concern themselves with isolated cognitive domains, e.g., visual perception, language processing, imagery, and so on, and develop special-purpose theories couched in terms of specialized representations and processes. Such accounts do not straightforwardly generalize to other domains, nor is it obvious that the representational types they trade in have any general applicability. This is problematic, given that we are seeking a broad criterion that will distinguish *every* conscious representational vehicle (irrespective of domain) from *each and every* unconscious vehicle. However, despite these problems, there is no reason we can't consider the types of representations proposed in particular models independent of their *intended* applications.

are declarations (in some imaginary programming language) of an array type Vector consisting of a set of three floating-point numbers, and an array type Matrix whose components are themselves arrays. A *record* is also an n-dimensional data structure, but its components need not be of the same type. An array or record is normally stored in a contiguous block of memory, but this need not be the case, since most modern computers have the capacity to *indirectly* address data. This means that the addresses of the array components can be stored in a special set of registers. When the processor attempts to access an array, instead of going directly to the data, it first goes to the address registers, looks up the first address, fetches the first component from that location, then goes back to the address register, looks up the address of the next component, and so on. These address registers are known as *pointers*, because they contain the information required to find pieces of data. Pointers are very important, because they allow for the creation of a different kind of data structure: the *dynamic* data structure. Trees, lists and graphs are of this kind. A *list* is a set of records, each of which consists of two components: a head, which can be of any type; and a tail, which points to another element of the same type. In other words, a list is defined recursively. A *list type* is only fully specified once the type of its head has been specified. Each element in the list is linked to the next via its tail. Because of the way it is defined, a list has no fixed size, i.e., unlike arrays and records generally, the size of a list can vary dynamically during program execution. *Trees* and *graphs* also have this property, but, unlike lists, use pointers to create complex, non-linear structures.

This brief survey by no means exhausts the structured data types of computer science, but does serve to illustrate their general nature. Moreover, it suffices to illustrate the following point: the distinction between structured data types in conventional computer science is actually one based on the *computational role* of these vehicles, not their intrinsic physical properties. An array, for instance, is an indexed set of data representations. Any particular array clearly has determinate intrinsic properties, but what is essential to its being an array (as opposed to, say, a list) is the computational relations into which it enters, and which permit orderly access to its elements (Marcotty and Ledgard 1986, p. 247). In effect, being apt for ordered element selection (and perhaps, being addressable *en bloc*) is definitional for an array, and this capacity is grounded in the relations it bears to the computational system. To take another example: what is central to the identity of a list is the relations between its elements (the fact that they are linked in a certain way) and its open-endedness (the fact that it has the capacity to grow). Both of these features depend on the way list structures are treated by the computational system: the first is reflected in the pointer mechanism that enables the system to gain access to list elements; and the second is reflected in the capacity of the system to build linear data chains. Again we find that what ultimately makes some particular physical structure a list is not its intrinsic properties, but its computational relations.

Similar remarks, we believe, can be made about the kinds of representations posited by cognitive psychologists. Consider, for example, schemas and semantic networks. *Semantic networks* were first introduced to aid in the development of language-parsing programs. In cognitive psychology they are the representation of choice for theories concerning the storage and application of declarative/factual knowledge, i.e., knowledge of particular and general propositions (e.g., *Ben is an aardvark, aardvarks are mammals*) and abstract conceptual knowledge. For this reason they are sometimes referred to as *propositional representations*. A semantic network consists of linked sets of propositional units, each of which is built around a proposition node, with pointers to a relation node and an appropriate number of argument nodes. Argument nodes can refer to concrete particulars (e.g., Ben) or general concepts (e.g., mammal). Conceptual knowledge is often taken to reside in such networks too, and, for each concept, consists in the cluster of general propositions attaching to its node in the network – a structure known as a *schema*. Schemas represent general information about categories, be they objects or events, but leave out specific details (such as the precise color and location of the aardvark that just entered the room). When associated with knowledge of objects, schemas are called *frames*, and when associated with knowledge of events they are known as *scripts*. Semantic networks are assumed to be open-ended, in that the storage of new factual knowledge requires the network to be extended in appropriate ways (Stillings et al., 1995, pp. 26–37).

If we ignore their intended contents, and consider them merely from a structural point of view, it appears that semantic nets (which can also be conceived as collections of frames) are simply instances of graphs. Like graphs they are dynamic structures built out of linked elements, each of which has one or more components carrying fixed values, and one or more pointers to other elements in the network. What distinguishes them from graphs in general is the fact that they are designed to support a particular kind of inferential process known as *property inheritance*. Property inheritance is mediated by special pointers (called IS-A and INSTANCE-OF slots in the context of frames), that indicate the relations of set-inclusion and set-membership (see Rich and Knight, 1991, pp. 251–75). Without going into further details, it is pretty clear that semantic networks and schemas essentially add nothing new to the repertoire of data types provided by computer science. They are merely specialized graph structures over which a particular kind of computational operation is defined.

As a second example, consider the kinds of representations that arise in Kosslyn's well-known account of visual imagery (1980). Kosslyn claims that any cognitive theory must describe the brain in terms of both structures and processes. He develops an account of mental imagery that trades in three distinct kinds of data structures – surface representations, literal encodings and propositional encodings – and numerous kinds of processes that operate on those structures. It is the surface representations that are of particular interest to us, because it is these which Kosslyn takes to support the experience of mental images, and give visual imagery its pictorial flavor. Surface representations consist of regions of activation in a special type of data structure Kosslyn calls the visual buffer.[8] Its components are numerous "loci", which can take on one of two values: active or inactive, and in computer models it is treated as an array. What primarily distinguishes the visual buffer from other array types (apart from the two-valued nature of its components) is the way visual contents map onto it. Kosslyn makes a number of proposals in this regard (1981, pp. 213–15), the most significant of which is that:

> surface images consist of regions of activation in the visual buffer that correspond to regions of depicted objects, with distances among the regions of an object (as seen from a particular point of view) being preserved by distances among the regions used to represent it in the medium (1981, p. 215).

This sounds like a recipe for a pictorial system of representation. For this reason the visual buffer is often described as an analog representational medium. However, this is a mistake, because Kosslyn is quite emphatic that "*distance* in the medium can be defined without reference to actual physical distance but merely in terms of the number of locations [loci] intervening between any two locations" (1981, p. 215.). Thus, while the visual buffer can be regarded as a coordinate space,

> this "space" is not an actual physical one but is rather a functional space defined by the way processes access the structure. The functional relations of the loci in the visual buffer need not be determined by actual physical relations any more than the functional relations of cells in an array in a computer need be determined by the physical relations among the parts of core memory. (1981, pp. 213–14)

The idea seems to be that while a surface representation is not literally pictorial, it is nevertheless treated by the cognitive system, with respect to certain uses, *as if* it were a picture.[9] In other words, there are processes defined over the visual buffer that respect its intended interpretation. The existence of these processes is peculiar to the visual buffer, and is constitutive of its identity as

---

8 Kosslyn actually refers to the visual buffer as a "medium" of representation. He reserves the term "data structure" for the surface representations themselves. Von Eckardt (1993, pp. 172–74) convincingly argues that what Kosslyn calls a "medium" corresponds to the computer science notion of a structured data type, and that what he calls "data structures" are *instances* of particular structured types.

9 The kinds of uses one should have in mind here are those to which a picture can normally be put, e.g., determining co-linearity and adjacency relations among the parts of items it represents. See Tye 1991, pp. 40–41, for further discussion of this aspect of Kosslyn's account.

an array structure. But, just like the other data structures we have examined, this implies that what distinguishes the visual buffer from other representational types are computational relations, not intrinsic characteristics.

These considerations appear to us to seriously undermine any attempt to base a limited denial of the dissociation thesis on distinctions between the data structures standardly encountered in either conventional computer science or classically-inspired cognitive psychology. The distinctions here, while they might appear to pick out distinct kinds of representational vehicles, are actually grounded in differential computational roles – what ultimately distinguishes one data type from another is the repertoire of computational operations in which its instances are implicated. To associate conscious experience with a particular data structure (or restricted range of data structures), therefore, is to align consciousness with a certain kind of computational process, rather than a distinctive type of representational vehicle. Consequently, what this analysis suggests is that on vehicle criteria alone classical cognitive science can only appeal to one kind of representational vehicle for the explicit coding of information: the symbol structure. And since we've already seen that according to classical accounts of human cognition it is not even remotely plausible to suppose that the content of every symbol structure is conscious, classicists have no choice but to embrace the dissociation thesis – a conclusion, we think, that formalizes what most classicists simply take for granted.

What does all this mean for classicists? It suggests that, provided the rest of Dennett's argument goes through, classicists can't be Cartesian materialists, even of the distributed variety. This obviously won't surprise Dennett in the least. What might surprise him, though, is how the whole situation changes once we transpose this discussion into the connectionist context.

5. Connectionism and the Dissociation Thesis

Up until recently most cognitive scientists and philosophers of mind felt constrained to support some kind of classical account of cognition – if only for lack of a viable alternative. Jerry Fodor is famous for arguing that since classicism is the only game in town, if one doesn't play it, then one just doesn't play. But Fodor's quip no longer rings true; over the last ten years we've seen the emergence of an exciting new game. And this new game, the *connectionist* computational theory of cognition, is proving very popular indeed.[10]

---

[10] We are assuming here that connectionism does constitute a *computational* account of human cognition (and is hence a competing paradigm *within* the discipline of cognitive science). Although some have questioned this assumption, we think it accords with the orthodox view (see, e.g., Cummins and Schwarz, 1991; Fodor and Pylyshyn, 1988; Von Eckardt, 1993, Chap. 3).

Whereas classicism is grounded in the computational theory underpinning the operation of conventional digital computers, connectionism relies on a neurally inspired computational framework commonly known as *parallel distributed processing* (or just PDP).[11] PDP directly models some high-level physical properties of the brain. A PDP network consists in a collection of simple processing units, each of which has a continuously variable activation level. These units are physically linked by connection lines, which enable the activation level of one unit to contribute to the input and subsequent activation of other units. These connection lines incorporate modifiable connection weights, which modulate the effect of one unit on another in either an excitatory or inhibitory fashion. Each unit sums the modulated inputs it receives, and then generates a new activation level that is some threshold function of its present activation level and that sum. A PDP network typically performs computational operations by "relaxing" into a stable pattern of activation in response to a stable array of inputs. These operations are mediated by the connection weights, which determine (together with network connectivity) the way that activation is passed from unit to unit.

The PDP computational framework does for connectionism what digital computational theory does for classicism. Human cognitive processes, according to connectionism, are the computational operations of a multitude of PDP networks implemented in the neural hardware in our heads. And the human mind is viewed as a coalition of interconnected, special-purpose, PDP devices whose combined activity is responsible for the rich diversity of our thought and behavior. This is the connectionist computational theory of mind.[12] Our task in this section is to consider whether connectionists, like classicists, are forced to embrace the dissociation thesis.

The vehicles of explicit representation in the connectionist context are the stable patterns of activation generated across neurally realized PDP networks in response to their inputs. Just as in the case of classical symbols tokened in the language of thought, these *activation pattern representations* are causally efficacious, physically structured states, embedded in a system with the capacity to parse them in structure sensitive ways. Being stable enables an activation pattern to contribute to the clamping of inputs to other networks, thus generating further regions of stability (and ultimately contributing to coherent

---

11 The *locus classicus* of PDP is the two volume set by Rumelhart, McClelland, and the PDP Research Group (Rumelhart and McClelland, 1986; McClelland and Rumelhart, 1986). Useful introductions to PDP are Rumelhart and McClelland, 1986, Chaps. 1–3; Rumelhart, 1989; and Bechtel and Abrahamsen, 1991, Chaps. 1–4.

12 Some of the more prominent contemporary philosophers and cognitive scientists who advocate a connectionist conception of cognition include Clark (1989, 1993), Cussins (1990), Horgan and Tienson (1989, 1996), Rumelhart and McClelland (Rumelhart and McClelland, 1986; McClelland and Rumelhart, 1986), Smolensky (1988), and the earlier Van Gelder (1990). For useful introductions to connectionism, see Bechtel and Abrahamsen, 1991; Clark, 1989, Chaps. 5–6; Rumelhart, 1989; and Tienson, 1987.

schemes of action). Moreover, the quality of this effect is *structure sensitive* (ceteris paribus), since it is dependent on the precise profile of the source activation pattern. Finally, while the semantics here is not language-like (lacking a combinatorial syntax and semantics), it typically involves some kind of systematic mapping between locations in activation space and the object domain. A connectionist who wishes to abandon the dissociation thesis must, therefore, identify phenomenal experience with activation pattern representations.[13]

What would be the shape of the resulting connectionist account of consciousness? Such an account would hold that each element of phenomenal experience corresponds with the generation of a stable pattern of activation somewhere in the brain, and conversely, that whenever a stable pattern of activation is generated, the content of that representation is phenomenally experienced. To be more precise, according to this account *each* stable activation pattern is (numerically) identical to a phenomenal experience. In particular, each activation pattern is identical to an experience in which the information content encoded by that vehicle is "manifested" or "displayed" – that is, the "what-it-is-likeness" of each phenomenal experience is *constituted* by the information content that each pattern encodes. In addition, this account would hold that whenever unconscious information is causally implicated in cognition, such information is not explicitly coded in the form of stable patterns of activation.

But is this connectionist version of the story any more plausible than its classical counterpart? We think it is, because there is a crucial representational asymmetry between classicism and connectionism. Whereas classicism is unable to meet the demand for a causally efficacious unconscious without invoking a good deal of unconscious symbol manipulation, connectionism can call upon an *inexplicit* style of mental representation to do the work of

---

13 Theorists involved in laying the foundations of the connectionist approach to cognition recognized a potential role for stable patterns of activation in an account of phenomenal experience. In the very volumes in which connectionism receives its first comprehensive statement (Rumelhart and McClelland 1986; McClelland and Rumelhart 1986), for example, we find the suggestion that:

> ...the contents of consciousness are dominated by the relatively stable states of the [cognitive] system. Thus, since consciousness is on the time scale of sequences of stable states, consciousness consists of a sequence of interpretations – each represented by a stable state of the system. (Rumelhart, Smolensky, McClelland and Hinton 1986, p. 39)

And in another seminal piece, Smolensky makes a similar suggestion:

> The contents of consciousness reflect only the large-scale structure of activity patterns: subpatterns of activity that are extended over spatially large regions of the network and that are stable for relatively long periods of time. (1988, p. 13)

More recently, both Lloyd (1991, 1995, and 1996) and Mangan (1993b) have explored the possibility of aligning conscious experience with these connectionist representational vehicles.

the unconscious, thus leaving the vehicles of explicit representation – stable activation patterns – free to line up with the contents of phenomenal experience. In order to make this point clear, we'll need to say a little more about PDP theory.

Since they are modified whenever a network is exposed to a new input, activation pattern representations merely constitute a transient form of information coding in PDP systems. The long-term representational capacities of these systems rely on the plasticity of the connection weights between the constituent processing units.[14] Through repeated applications of what is known as a *learning rule*, which modifies the connection weights in an incremental fashion, an individual network can be taught to generate a range of stable target patterns in response to a range of inputs.[15] A "trained" network thus encodes the disposition to generate a whole range of activation patterns, in response to cuing inputs. So, in virtue of the configuration of its connection weights and its pattern of connectivity, a network has the ability to *store* appropriate responses to input. Furthermore, such long-term information storage is *superpositional* in nature, since *each* connection weight contributes to the storage of *every* disposition. This has the important consequence that such information is not encoded in a physically discrete manner. The one appropriately configured network encodes a *set* of contents corresponding to the range of explicit tokens it is disposed to generate. For these reasons, connection weight dispositions constitute a distinct, inexplicit style of mental representation. We will refer to them as *connection weight representations*.

Connection weight representations are stored in a PDP network in virtue of its particular pattern of weights and connectivity. However, this pattern of weights and connectivity is also responsible for the manner in which the network responds to input, and hence the manner in which it processes information. This means that the causal substrate driving the computational operations of a PDP network is identical to the supervenience base of the network's stored information. So there is a strong sense in which it is inexplicitly coded information in a network (i.e., the network's "memory") that actually governs its computational operations.[16]

---

14 For good general introductions to the representational properties of PDP systems, see Bechtel and Abrahamsen, 1991, Chap. 2; Churchland, 1995; Churchland and Sejnowski, 1994, Chap. 4; Rumelhart and McClelland, 1986, Chaps. 1–3; and Rumelhart, 1989. More fine-grained discussions of the same can be found in Clark, 1993; and Ramsey, Stich and Rumelhart, 1991, Part II.

15 The most common learning rule employed in simulated PDP networks is *back propagation*, which uses the discrepancy between actual and target output to nudge connection weights in a direction that will reduce the error value – see, e.g., the discussion in Rumelhart, Hinton and Williams, 1986.

16 This is another way of expressing the oft-cited claim that PDP breaks down the conventional code/process distinction—see, e.g., Clark, 1993.

This fact about PDP systems has major consequences for the manner in which connectionists conceptualize cognitive processes. Crucially, information that is superpositionally encoded in the form of connection weight dispositions need not be realized in the form of activation patterns in order to be causally efficacious. There is a real sense in which *all* the information that is superpositionally encoded in a network is causally active *whenever* that network responds to an input.

It's as a direct consequence of this, that connectionism is able to account for our intelligent behavior *without* invoking the unconscious manipulation of explicit representations, and without denying that most of the mind's operations are unconscious. Conscious states are regions of stability in a sea of unconscious activity, causal activity which connectionists can account for entirely in terms of operations that implicate connection weight representations. These take the form of *intra*-network "relaxation" processes, that result in stable activation patterns, and which are determined by the superpositionally encoded information stored therein. Unconscious processes thus *generate* activation pattern representations, which the connectionist is free to align with phenomenal experiences, since none is required to account for the unconscious activity itself. The picture is this: the unconscious *process*, entirely mediated by inexplicitly encoded data, generates a conscious *product*, in the form of explicit representations – stable patterns of activation in neurally realized PDP networks.

What is more, learning can also be accounted for, using these resources. On the connectionist story, learning involves the progressive modification of network connection matrices, in order to encode *further* connection weight representations. Learning, in other words, is a process which actually reconfigures the brain's functional architecture, and hence adjusts the primitive computational operations of the system. But this reconfiguration is fully cognitive, and hence not merely maturational, because these modifications are wholly explicable in terms of the brain's basic style of computation.

What we've established is that connectionists have a computational ontology rich enough to deny the dissociation thesis. Such a denial leads to a connectionist version of Cartesian materialism which holds, like any Cartesian materialist position, that the onset of conscious experience is marked by the transduction of information into a privileged neural medium. The privileged medium, in this case, comprises the vehicles with which information is explicitly encoded in the brain, in the form of stable patterns of activation across its neural networks. And since connectionism holds that from moment to moment, as the brain simultaneously processes parallel streams of input and ongoing streams of internal activity, a large number of such stable activation patterns are generated across hundreds (perhaps even thousands) of neural networks, this version of Cartesian materialism is committed not to a single, central theater of consciousness, but myriad minitheaters scattered right across the cortex.

This suggests that consciousness is both an *additive* and a *distributed* affair. It is *additive* in that our phenomenal experience at any one moment is an amalgam of a number of *phenomenal elements*, each simultaneously contributing to the overall experience. And it is *distributed* in that the processes responsible for experience occur simultaneously in distinct neural structures right across the brain; no one part of the brain is pivotal insofar as the production of phenomenal experience is concerned. Such claims may initially appear counter-intuitive, but recent work in the neurosciences supports this picture.[17] In particular, deficit studies attest to the additive nature of consciousness, even *within* sense modalities. Visual experience, for example, normally appears seamless, because its various components (form, motion, color and so on) are generated simultaneously across the visual cortex. But very localized brain damage can selectively mar this unity. Subjects report losses of color, form and motion, respectively, and strange spatial dissociations of these elements (see, e.g., Smythies, 1994, p. 313). Such dissociations and losses are to be expected in a brain suffering only localized damage, according to this connectionist form of Cartesian materialism, since the elements of conscious experience are generated locally, in specialized neural networks.

Finally, and perhaps most importantly in this context, connectionist Cartesian materialism, precisely because it denies the dissociation thesis, is left intact by Dennett's argumentative assault. To make this point, and in so doing round off our discussion, we will return to the color phi phenomenon and Dennett's distinction between Orwellian and Stalinesque revisions (see Section 2).

A connectionist Cartesian materialist can offer at least two different explanations of the apparent motion exhibited by color phi. Here's the Orwellian story. The subject's brain, when exposed to the tachistoscopically-presented stimuli, first commences processing information regarding the stationary red spot. This implicates multiple neural networks engaged in processes of constraint satisfaction and activation passing, and results in a stabilization that carries the explicit (and therefore conscious) content: *stationary-red-spot.* Sometime during this cognitive episode the subject's brain begins processing input generated by the second stationary (green) spot, and eventually a second stabilization is achieved, this time with the explicit content: *stationary-green-spot (a-little-to-the-right)*, which is consciously experienced. Very rapidly soon after, however, this second stabilization is obliterated and a third repre-

17 For the neuropsychological evidence of the distributed neural basis of consciousness, see, e.g., the papers in Milner and Rugg, 1992. See Bisiach, 1992 for comments on both the distributed and additive nature of consciousness.

sentational vehicle, a third stable pattern of activation, is created (this time with the alternative content *moving-red-spot-changing-to-green*), presumably to suit some phylogenetic imperative that has been hard-wired into the brain. Finally, this third vehicle contrives to interfere with the inexplicit storage of the information that was momentarily represented earlier, such that the latter has no further information processing effects in the system. While the subject did momentarily experience spatially and temporally distinct spots of light, they only remember and hence report seeing a single, moving, spot changing in color from red to green.

A Stalinesque account starts the same way, but deviates from the Orwellian version of events at the point where the visual cortex is processing the second tachistoscopically-presented stimulus. This second processing sequence, a Stalinesque theorist might claim, is not independent of the stabilization responsible for coding the first (conscious) content. Due to their spatial and temporal proximity (deriving from a similar relationship between their stimulus sources) the stabil*ized* networks associated with the first (red) spot interfere with the processing regime in the stabil*izing* networks associated with the second (green) spot, an interaction that is quite in keeping with the connective properties of PDP networks. As a result, the potential content *stationary-green-spot (a-little-to-the-right)* never gets explicitly coded in the visual cortex; instead, the second pattern of activation that ultimately stabilizes carries the fabricated content *moving-red-spot-changing-to-green*, perhaps based on some inbuilt assumptions about object permanence. The subject consequently never experiences a second stationary spot, and their memory faithfully records this fact.

The connectionist Cartesian materialist is not forced to choose between these two explanations; they are free to let future research indicate precisely which content fixations actually issue from the inputs that produce apparent motion. And crucially, by attending to the stable patterns of neural activity in the visual cortex, neuroscientists would be able to tell the difference. If they detect a stabilization with the content *stationary-green-spot (a-little-to-the-right)* immediately following one with the content *stationary-red-spot* and just prior to another whose content is *moving-red-spot-changing-to-green*, the revision is Orwellian; if no such stable pattern is discerned, it is Stalinesque. Contrary to what Dennett claims, this version of Cartesian materialism is committed to a difference that makes a difference.

6. Conclusion

In *Consciousness Explained*, Dennett claims that to be a Cartesian materialist is to be committed to the existence of a boundary across which informational contents must pass in order to be rendered conscious, and that there is nothing in first-person reports, or in the evidence available from the third-person perspective, that could possibly enable us to locate this boundary. However, in putting his case for this claim Dennett tacitly assumes the dissociation thesis, the view that it's not possible to identify conscious experience with the explicit representation of information in the brain.

We've argued that classicists are committed to the dissociation thesis. So if the remainder of Dennett's argument goes through, classicists can't be Cartesian materialists. Connectionists, on the other hand, because of the distinctive computational and representational properties of the PDP framework, are in a position to deny the dissociation thesis. As we've demonstrated, they can plausibly adopt an account of conscious experience which aligns the contents of consciousness with stable patterns of activation in neurally realized PDP networks. What this means is that connectionists, unlike classicists, are in a position to embrace a form of Cartesian materialism that is immune to Dennett's assault. When Dennett claims that the "spatially and temporally distributed content-fixations [in the brain] are precisely locatable in both space and time" (1991, p. 113), connectionists can make the same claim with regard to phenomenal experience. They can conjecture, in other words, that there is indeed a special neural medium in the brain, the transduction of information into which marks the onset phenomenal consciousness. This, we think, represents an attractive hypothesis about the neural basis of our conscious experience, and one that is worthy of further exploration.[18] Whether this connectionist Cartesian materialism ultimately turns out to be a satisfactory account of phenomenal consciousness, however, is something only a great deal of further research will reveal.

References

Bechtel, W. and Abrahamsen, A. (1991). *Connectionism and the Mind.* Blackwell.

Block, N. (1993). Book review of Dennett's *Consciousness Explained. Journal of Philosophy* 90, 181–93.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18, 227–87.

Bisiach, E. (1992). Understanding consciousness: Clues from unilateral neglect and related disorders. In A. Milner and M. Rugg (Eds.), *The Neuropsychology of Consciousness.* Academic Press.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences* 3, 1–62.

Churchland, P. M. (1995). *The Engine of Reason, the Seat of the Soul.* MIT Press.

Churchland, P. S. and Sejnowski, T (1994). *The Computational Brain.* MIT Press.

[18] See O'Brien and Opie (1999) for some explorations along these lines.

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.

Cummins, R. and Schwarz, G. (1991). Connectionism, computation, and cognition. In T. Horgan and J. Tienson (Eds.), *Connectionism and the Philosophy of Mind*. Kluwer.

Cussins, A. (1990). The connectionist construction of concepts. In M. Boden (Ed.), *The Philosophy of Artificial Intelligence*. Oxford University Press

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

Dennett, D. C. (1993). The message is: There is no *me*dium. *Philosophy and Phenomenological Research* 53, 919–31.

Field, H. (1978). Mental representation. *Erkentniss* 13, 9–61.

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

Fodor, J. A. (1981). *Representations*. MIT Press.

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press.

Fodor, J. A. (1987). *Psychosemantics*. MIT Press.

Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* 28, 3–71.

Harman, G. (1973). *Thought*. Princeton University Press

Haugeland, J. (1981). Semantic engines: An introduction to mind design. In J. Haugeland (Ed.), *Mind Design: Philosophy, Psychology, and Artificial Intelligence*. MIT Press.

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

Horgan, T. and Tienson, J. (1989). Representations without rules. *Philosophical Topics* 27, 147–74.

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

Kosslyn, S. M. (1980) *Image and Mind*. Harvard University Press

Kosslyn, S. M. (1981) The medium and the message in mental imagery. In N. Block (Ed.) *Imagery*. MIT Press

Lashley, K. (1956). Cerebral organization and behavior. In S. Cobb and W. Penfield (Eds.), *The Brain and Human Behavior*. Williams and Wilkins Press.

Lloyd, D. (1991) Leaping to conclusions: Connectionism, consciousness and the computational mind. In T. Horgan and J. Tienson (Eds.) *Connectionism and the Philosophy of Mind*. Kluwer.

Lloyd, D. (1995) Consciousness: A connectionist manifesto. *Minds and Machines* 5, 161–85.

Lloyd, D. (1996) Consciousness, connectionism, and cognitive neuroscience: A meeting of the minds. *Philosophical Psychology* 9, 61–79.

Mangan, B. (1993a). Dennett, consciousness, and the sorrows of functionalism. *Consciousness and Cognition* 2, 1–17.

Mangan, B. (1993b) Taking phenomenology seriously: The "fringe" and its implications for cognitive research. *Consciousness and Cognition* 2, 89–108.

Marcotty, M. and Ledgard, H. (1986) *Programming Language Landscape.* Science Research Associates.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* Freeman.

McClelland, J. L. and Rumelhart, D. E., Eds. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Vol. 2: Psychological and Biological Models.* MIT Press.

Milner, A. and Rugg, M., Eds. (1992) *The Neuropsychology of Consciousness.* Academic Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review* 83, 435–50.

Newell, A. (1980). Physical symbol systems. *Cognitive Science* 4, 135–83.

O'Brien, G. and Opie, J. (1999). A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences* 22, 127–96.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences* 3, 111–69.

Pylyshyn, Z. W. (1984). *Computation and Cognition.* MIT Press.

Ramsey, W., Stich, S. and Rumelhart, D. E., Eds. (1991). *Philosophy and Connectionist Theory.* Lawrence Erlbaum.

Rich, E. and Knight, K. (1991) *Artificial Intelligence.* McGraw-Hill

Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. In M. Posner (Ed.), *Foundations of Cognitive Science.* MIT Press.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). Learning internal representations by error propagation. In D. Rumelhart and J. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Vol. 1: Foundations.* MIT Press.

Rumelhart, D. E. and McClelland, J. L., Eds. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Vol. 1: Foundations.* MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L. and Hinton, G. E. (1986) Schemata and sequential thought processes in PDP models. In J. McClelland and D. Rumelhart (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Vol. 2: Psychological and Biological Models.* MIT Press.

Shoemaker, S. (1993). Lovely and suspect ideas. *Philosophy and Phenomenological Research* 53, 905–10.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11, 1–23.

Scragg, G. W. (1992) *Computer Organization: A Top-Down Approach.* McGraw-Hill

Sterelny, K. (1990). *The Representational Theory of Mind.* Blackwell

Stillings, N. A., Weisler, S. E., Chase, C. H., Feinstein, M. H., Garfield, J. L. and Rissland, E. L. (1995) *Cognitive Science: An Introduction.* MIT Press

Smythies, J. R. (1994). Requiem for the identity theory. *Inquiry* 37, 311–29.

Tienson, J. L. (1987). An introduction to connectionism. *The Southern Journal of Philosophy* 26(Supplement), 1–16.

Tye, M. (1991) *The Imagery Debate.* MIT Press

Tye, M. (1993). Reflections on Dennett and consciousness. *Philosophy and Phenomenological Research* 53, 893–98.

Van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science* 14, 355–84.

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