18

Representations: From Neural Systems to Cognitive Systems

William Bechtel

One of the hallmarks of *cognitive* explanations of behavior is that they appeal to mental representations and operations over them. For example, in explanations of problem-solving behavior, cognitivists posit representations of the current situation, the goal state, and of possible operations that might be performed, and construe problem solving as involving such things as comparisons of the representations of the current state and goal state and alterations in the representations of the current state so as to determine the consequence of various operations. Many of the most vociferous debates in the cognitive science literature have focused on the format of representations – whether they take the form of language-like symbols (or even natural language expressions themselves), pictures or images, or the sorts of distributed representations identified in connectionist models.

Given the propensity of scientists and philosophers to challenge even the most basic assumptions of their inquiry, it should not be surprising that some 30 years into the cognitive revolution some investigators – in particular, those advocating dynamical explanations of cognitive phenomena – would begin to challenge the very need for representations. However, although appeals to internal representations are rampant in cognitive science, they are not limited to that field. Biologists in a variety of fields refer to molecular structures as representations. Of special importance to our purposes, neuroscientists, especially behavioral and cognitive neuroscientists, routinely refer to brain activities as representations. This raises a question: should the attacks on representations that have emerged recently in cognitive science apply equally to appeals to representations in neuroscience?[1] I will argue, rather, that there is a viable notion of representation that is being employed in neuroscience that is not subject to the dynamicist's challenge. This answer, however, raises a further question: is this notion of representation adequate for the purposes for which cognitivists have appealed to representations. In the last section I will sketch how one might build up from the conception of representation employed in neuroscience to one adequate for cognitive accounts.

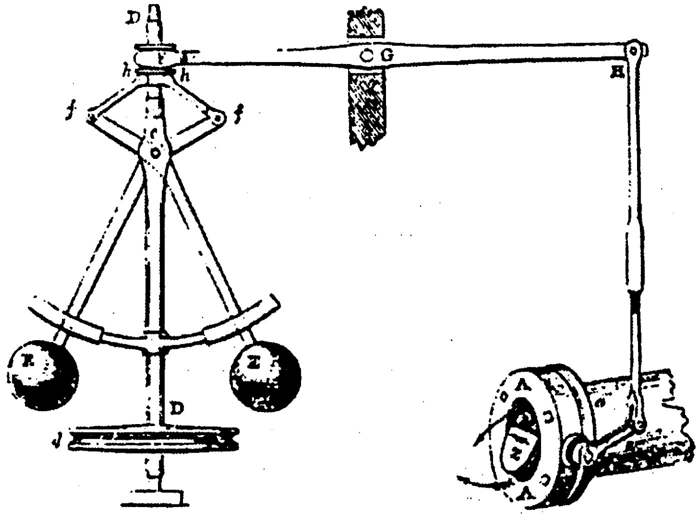

Figure 18.1 Watt's centrifugal governor for a steam engine. Drawing from J. Farley, *A Treatise on the Steam Engine: Historical, Practical, and Descriptive* (London: Longman, Rees, Orme, Brown and Green, 1927).

1 The Dynamicist's Critique of Representations

The basic complaint of the dynamicist critics of representations is that physical systems lacking representations could accomplish many or all of the various tasks for which cognitivists had postulated internal representations (and operations involving them). One strategy the critics used to mount their case was to present examples of systems that accomplish tasks for which we might be tempted to posit representations (e.g. problem-solving tasks) and to demonstrate ways of accomplishing such tasks without internal representations. Van Gelder's (1995) example of the governor James Watt introduced for his steam engine is perhaps the best known.

The task facing Watt was to regulate the output of steam from a steam engine so that the flywheel would rotate at a constant speed regardless of the resistance being generated by the appliances connected to it. Watt's governor was ingeniously simple (figure 18.1). He attached a spindle on a flywheel driven by the steam generated by the steam engine, and attached arms to the spindle which would, as a result of centrifugal force, open out in proportion to the speed at which the flywheel turned. A mechanical linkage between the arms connected the arms to the steam valve so that, when the wheel turned too fast, the valve would close, releasing less steam, thereby slowing the flywheel, but when the flywheel turned too slowly, the valve would open, releasing more steam and speeding up the flywheel. It is in part the simplicity of this device that led van Gelder to reject a representationalist interpretation in which the angle of the arms represented the speed of the flywheel. But, van Gelder argued, the Watt governor solves the very kind of problem for which a cognitivist might be tempted to develop a representationalist solution – one, for example, in which the present and desired speed of the flywheel as well as the operations of opening and closing the steam valve are represented, and rules that apply to the representations are invoked to determine how much to open or close the valve (think of how a computer might be programmed to perform this task). Thus, the Watt governor, for van Gelder, demonstrates how one might perform cognitive tasks that seemingly require representations without them, motivating the program of trying to account for cognitive activities more broadly without positing representations.

One response to van Gelder's strategy is to grant that some simple tasks for which one might be tempted to posit representations do not in fact require them, but to contend that there are other tasks that are "representation-hungry" and do require them (Clark and Toribio, 1994). In this vein, Grush (chapter 19, this volume) argues that we find representations only in contexts where the system does not have immediate access to what is represented, and so representations stand in for what is not physically or temporarily present. In contrast, I have adopted an extreme position and argued that there are representations even in the Watt governor (Bechtel, 1998). Many philosophers have found this to be a mission of madness. They maintain that the notion of representation loses its interest if we can indeed find representations in such simple systems. In this chapter I argue that this is not the case by showing that this is precisely the kind of system for which cognitive neuroscience finds the notion of representation indispensable. When techniques such as single-cell recording and neuroimaging are employed to understand how the brain performs cognitive tasks, implicitly the notion of representation on which they are relying is the same as I invoked in identifying representations in the Watt governor. Thus, rather than being too simple to have representations, the Watt governor provides a simple, illustrative case where we can readily observe the importance of appeals to representations.

2 The Watt Governor as an Exemplar Representational System

Central to demonstrating the presence of representations in the Watt governor is the articulation of an analysis of what it is for something to be a representation. In its most basic sense, a representation is something (an event or process) that stands in for and carries information about what it represents, enabling the system in which it occurs to use that information in directing its behavior. In the simplest case, the system actually acts upon that which is represented (see figure 18.2). For example, we employ maps to stand in for the actual geography, and use those maps to coordinate our behavior in a way that is responsive to that geography – finding a route to a target location or circumventing an obstacle. The angle of the arms of the Watt governor play such a standing-in role: they stand in for the speed of the steam engine

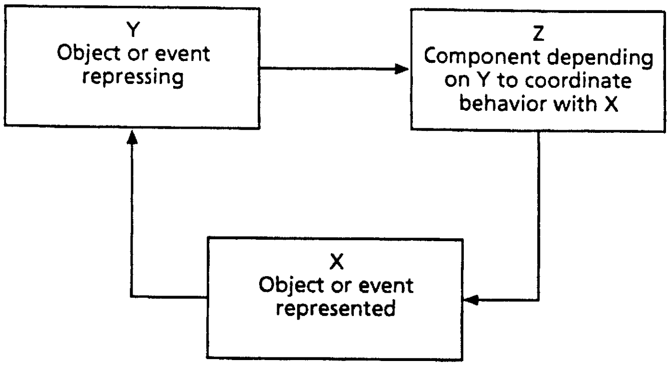

Figure 18.2 Three components in an analysis of a representation: the representation Y carries information about X for Z, which uses Y in order to act or think about X.

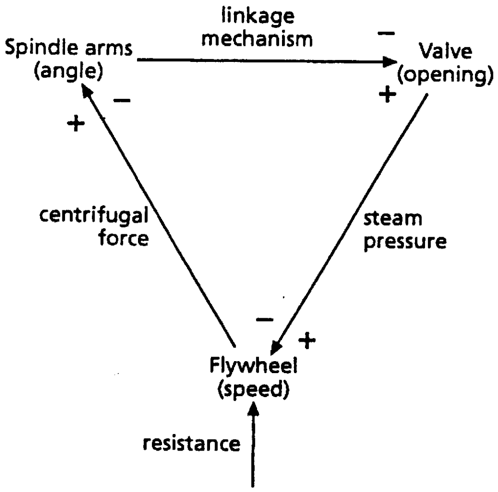

Figure 18.3 Application of this scheme to the Watt governor. The spindle arms carry information about the speed of the flywheel for the valve opening system, which uses the angle to determine the opening, thereby regulating the speed of the flywheel.

in a mechanism that controls the opening and closing of the steam valve and it is because they so stand in that the governor is able to regulate the flow of steam appropriately for the current speed of the flywheel. If someone did not immediately perceive how the governor performed its function, what one would do is explain how the angle of the arms responds to the speed with which the flywheel is turning (i.e. it carries information about the speed), thereby making this information available in a format that can be used by the linkage mechanism to control the flow of steam and thus the speed of the flywheel at a future time (figure 18.3).

This characterization of what it is to be a representation has two components. One is the idea of an information-bearing state. As Dretske (1981), among others, has argued, information can be analyzed causally – an effect of a process carries information about its cause (specifically, that an event of the kind that could cause a particular effect had indeed occurred). Thus, thunder carries information about an electrical discharge. But Millikan (1984) stresses that being a causal effect is not sufficient to turn something into a representation. Something becomes a representation because of how it is used by a system in controlling its behavior. It is because a system is *designed* so as to use a state to inform it about something distal that makes the state a representation.

In this analysis, emphasis is placed on *design*, on how the system is designed to use information-bearing states in coordinating its behavior. One does not just look to how it in fact uses a particular information-bearing state. Otherwise, we end up with an account in which the meaning of a representation depends on whatever ways it is used. An account which appeals to actual use is often referred to a conceptual role semantics, and such accounts have been subject to powerful criticisms (Cummins, 1996). But in my analysis, it is only the designed use that matters. An appeal to the notion of design, of course, requires an account of what it is for something to be designed to do one thing rather than another. In biological systems the usual route is to argue that what something was designed to do was that which its predecessors did more successfully than their competitors and, thus, contributed to their differential reproductive success. It can be relatively easy to tell a story about how a particular trait of an organism contributed to its differential reproductive success (these are often criticized as "just-so stories"), but difficult (though not impossible) to actually demonstrate that the trait really did contribute in this way (Griffiths, 1997). In the case of artifacts that we have designed, we sometimes have privileged access to information about what they were designed to do.

According to this analysis of representations, only some of the things that cause a representation to appear are what it represents. It is in this way that one can account for misrepresentation, and show how states or events can represent things that are not their immediate cause. In biological evolution, a mechanism may be selected in which a state is only rarely actually produced by that which it represents – if it is important that the system respond to the distal state when it occurs, a system may evolve that is subject to many false alarms. In systems designed by humans, we can take this to the extreme – we can design a system in which a particular internal state would occur (or at least be more likely to occur) in response to a distal state, but where the distal state has never occurred and all occurrences of the response have been false alarms (e.g. a system designed to detect a nuclear missile attack).

Behind at least some of the opposition to finding representations in the Watt governor is the fear that if they occur there, representations are everywhere and there is no theoretical bite in calling something a representation (see Haselager and de Groot, forthcoming, for such an argument). Although I will soon argue that there is a point to some promiscuousness about representations, let me first show that under this analysis, representations are not as promiscuous as feared. States or events that carry information are indeed ubiquitous. Any causal effect carries information about what caused it. In emphasizing the consumer, however, we emphasize the kind of system in which the representation occurs. The Watt governor is a particular kind of physical system – a system that is designed to employ states bearing information to control its behavior. While there are many such systems within evolved organisms and human artifacts, not all natural systems employ internal states in such a manner. Representations are found only in those systems in which there is another process (a consumer) designed to use the representation in generating its behavior (in the simplest cases, to coordinate its behavior with respect to what is represented) and where the fact that the representation carried that information was the reason the down-line process was designed to rely on the representation.

Even though this analysis does not make representations as ubiquitous as might be feared, it does entail that representations will appear in many non-cognitive systems. Many physiological systems (e.g. simple biochemical systems such as fermentation where reactions are controlled by feedback mechanisms in which the availability of the product of the reaction determines whether it will continue) will, on this construal, employ representations since processes in them as well as the responses to them were selected because of the process's information-bearing role. But this, I would contend, is as it should be. Representational or intentional vocabulary is in fact regularly used in the sciences dealing with such systems. Without it, scientists would be hard pressed to explain how these systems perform the tasks for which they appear to have evolved. In this context, the Watt governor provides a simple model of the role representational discourse is playing in allowing us to explain such systems.

3 Representational States in Brains

As I noted above, representational talk is widely used in behavioral and cognitive neuroscience. A particularly clear example is Penfield and Rasmussen's (1950) homuncular maps of how the sensory and motor cortex represents different parts of the body (in which they emphasize in terms of the distorted sizes of the homunculus's organs how the maps devote more area to regions from which more sensory input is received or over which there is more motor control – see figure 4.2, above).[2] This is extremely natural insofar as neuroscientists are attempting to analyze how various components of the brain gather and use information in the course of producing behavior. In this section I will draw upon neuroscience research into visual processing in the brain, described more fully in Part III, to show that the appeals of neuroscientists to representations do indeed conform to the analysis in the previous section.

Although there are various choices as to what sorts of neural states or processes serve a representational function, a reasonable place to begin is to assume that it is the firing rate or firing pattern of individual neurons that serves as representations.

If that is the case, what one needs in order to understand the representational system of the brain is means to record the electrical activity of individual neurons and then to determine their representational content. Techniques for implanting electrodes in or next to individual neurons and then amplifying the signal and playing it through an audiometer were developed in the early decades of the twentieth century. The challenge was then to make sense of the firing patterns that could be detected. Although I am arguing that the focus in construing a state or event as a representation is on the consumer of the representation, neuroscientists typically begin by trying to correlate neural activity with external processes that they might represent (as with the method of single-cell recording – see chapter 4, this volume). This is, indeed, quite sensible and does not undercut my claim. An extremely useful first step in determining what the system takes a state or event to represent is to ascertain what information a state or event might carry. Then one can ask the question of how the system was designed to use the information.

Penfield and Rasmussen's homuncular representation of sensory cortex was arrived at through single-cell recording (the homuncular representation of motor cortex resulting from related experiments in which weak electrical stimuli were supplied through electrodes to see what motor organs the stimulated cells would affect). Talbot and Marshall (1941) were the first to apply single-cell recording to cells in the occipital lobe, which was already recognized as figuring in visual processing on the basis of lesion studies. They recorded from cells in Brodmann's area 17 (which later came to be known as V1) in anesthetized cats and monkeys, and correlated activations there with the location of stimuli in the animal's visual field. They showed that each cell responded to a stimulus in a particular part of the visual field (the cell's *receptive field*) and confirmed that cells were so organized in the cortex that they constituted a map that preserved the topography of the visual field (see figure 18.4). To determine what the firing of a cell represented, however, required a further step – determining what kinds of stimuli, when present in a cell's receptive field, caused it to fire. Stephen Kuffler (1953) provided the model for this type of investigation. He recorded from cells in the retina and the lateral geniculate nucleus (LGN) and determined that they were most activated either by a light stimulus in the center of the cell's receptive field surrounded by a dark area (an *on-center* cell), or by a light surround of a dark center (an *off-center* cell; see figures 13.3 and 13.4, above).

As recounted in chapters 10 and 13, Hubel and Wiesel (1962, 1968) extended Kuffler's approach to primary visual cortex, discovering that cells there responded to bars of light. They found that some cells, which they identified as *simple cells* (figure 13.3, C–G), responded to bars with specific orientation only in specific areas of the cell's receptive field, whereas others, which they referred to as *complex cells*, responded to bars with a specific orientation anywhere within the receptive field. The result of Hubel and Wiesel's research was to suggest that firing of cells in V1 represented bars or lines at specific orientations in parts of the visual field.

Chapter 13, above, relates how, in the years after Hubel and Wiesel's pioneering research, discovery of cells in different prestriate, temporal, and parietal lobe areas which responded to specific visual features (color in V4, motion in MT/V5, shapes

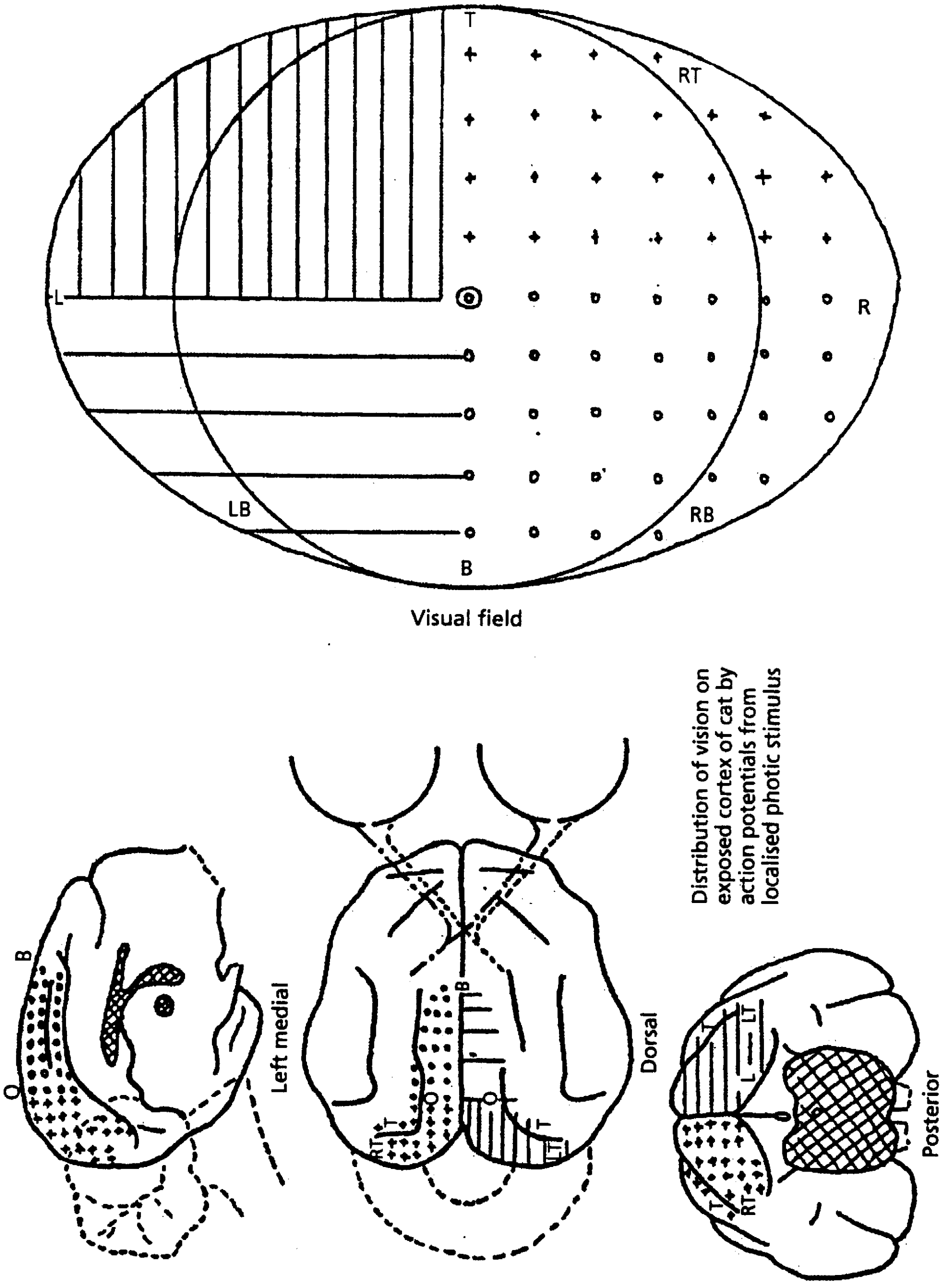

Figure 18.4 Talbot and Marshall's (1941) projection of areas of the visual field on to primary visual cortex in the cat, based on single-cell recording.

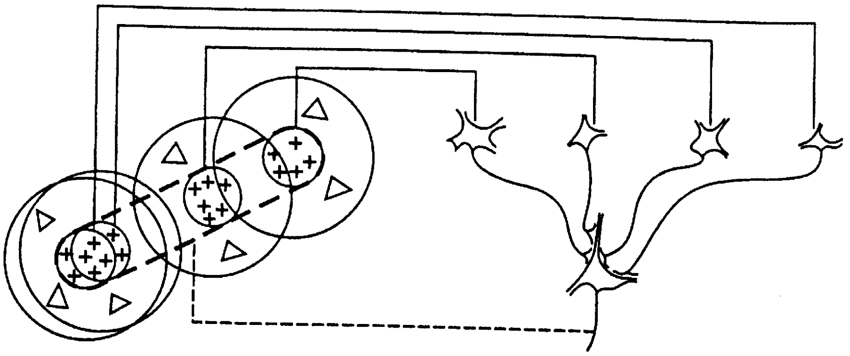

Figure 18.5 Hubel and Wiesel's proposal for a wiring diagram linking with excitatory connections from lateral geniculate cells with *on* centers to a simple cell in primary visual cortex that is excited by an oriented bar of light. From Hubel and Weisel, 1962, p. 142.

in inferior temporal cortex, and location information in posterior parietal cortex) continued with vigor. Van Essen and Gallant (chapter 12, this volume) indicate that 33 brain areas have been identified as principally involved in visual processing. Many of these have been shown to represent different specific types of visual information and currently much effort is being extended to developing accounts of how these representations figure in an overall processing scheme.

As I noted, most of this work has relied on establishing what information might be carried by different neural firings. The account of representation I endorsed, however, placed primary emphasis on the consumer of the representations. But a concern with the consumer of the representation is at least tacit in such research. Investigators would have had little interest in figuring out the information relation between brain activity and distal stimuli unless they assumed the brain was using this information in further processing. To determine how the brain actually consumes this information requires developing processing models which show how activations in later areas in visual pathways utilize the information encoded in earlier stages of processing. The early work of Hubel and Wiesel again provided an exemplar. They proposed simple wiring diagrams of how LGN cells might be wired to simple V1 cells and simple V1 cells in turn to complex V1 cells which were designed to show how downstream cells could compute their representations from more upstream representations (see figure 18.5). By showing that processes at each stage respond to those upstream to arrive at their characteristic response, they show that the upstream processes were representations.[3] Even in this example, the processing is more complex than in the Watt governor, where the consumer employed a simple mechanical linkage. Often complex computational models are required to characterize the processing mechanisms through which representations are used by later stages in processing (this is especially true when it is recognized that the feedback or recurrent processing plays a major role). Much current work, especially in computational neuroscience, is devoted to developing processing models showing how later visual areas can generate their representations from what is represented in earlier visual areas.

Part of the challenge in understanding the consumption of visual representations is that the visual areas are often many steps removed from any behavior, for the performance of which the organism relies on visual information. This is especially true in the ventral pathway from primary visual areas into temporal areas, through which object recognition is assumed to occur (see chapter 11, this volume). Object representations developed in inferior temporal cortex are thought to be employed in higher cognitive processes such as reasoning and problem solving, the detailed neural mechanisms of which have not yet been identified. In the dorsal pathway proceeding to the posterior parietal cortex, on the other hand, it has been easier for researchers to focus on the question of how putative representational states figure in guiding motor action. The basic strategy was developed by Goldman-Rakic (1987) in her research on prefrontal brain areas. She found cells, for example, that were active when a monkey had to temporarily remember a specific direction in which it was directed to move its eyes after a delay interval. She was able to establish that certain cells were active only when the monkey would correctly remember the particular direction of eye movement.

Even when researchers discover sufficient cues to develop an interpretation of what a neural pattern represents, the process of interpretation is quite indirect. This is nicely illustrated in a puzzle Larry Snyder has attempted to address. Focusing on neurons in the parietal cortex which fire when the animal is required to move its arm to a specific location, he asked whether these neurons represented the location to which the animal was *attending* or the location to which the animal *intended* to move. To do this, Snyder developed experimental protocols in which monkeys had to attend to one location while holding an intention to move to another location. This revealed that different areas in the parietal cortex were involved in representing the intention than were involved in the attended representation of a location (Snyder et al., 1997; Batista et al., 1999).

A word of caution is needed. Since the brain is much more complex than the Watt governor and we lack independent access to the design process, hypotheses about what is represented by specific neural activity must be treated as extremely tentative. The project of single-cell recording is limited by the stimuli one thinks to test. It was through serendipity that Hubel and Wiesel thought to test bar stimuli in V1 and that Gross thought to test hand-shaped stimuli for an area in the inferior temporal cortex. It would be easy for a researcher simply to fail to test whether a particular stimulus would drive a cell. In this light it is important to note that Van Essen and Gallant (chapter 12, this volume) found that esoteric stimuli, such as expanding or rotating stimuli, would cause specific MSTd cells, which fired weakly in response to straight line movements, to fire vigorously. Moreover, one should not assume that the cell is only carrying information about the stimulus that causes it to fire more vigorously. As Van Essen and Gallant stress, less than full responses may still carry important information that can be used by downstream consumers. Thus, cells may not be feature detectors, but may be better construed as filters with a representational profile. Finally, as Akins (chapter 20, this volume) emphasizes, neurons do not respond to objective features of the world that we assume they would represent, when we think about how we might design a system from scratch. For example, they may not respond to absolute properties, such as temperature, but rather their response may be relative to the current state of the organism (e.g. whether the stimulus is warmer or colder than background stimulation). (Although Akins rejects a representational analysis, see chapter 16, this volume, for a representational account compatible with these findings.)

In this section I have been focusing on how neuroscientists investigate the representational content of neural firings. A skeptic about representations might question what the scientist gains by identifying particular neural activities as representations. Why not settle for a simple causal model identifying the dynamics of activity in the brain? To answer this question, we need to bear in mind what the goal of inquiry is. Neuroscientists generally assume that the brain is a complex machine whose activity allows the organism to coordinate its behavior with features of its environment. As with the Watt governor, once one has identified components in the system, one wants to know how they facilitate the organism as it extracts information about its environment and deploys that information in determining its behavior. Someone examining Watt's invention might ask: What do the spindle arms do in the governor? The answer would be: They represent information on whether the flywheel is moving too fast or too slow and pass this information to the valve that controls the steam flow. Similarly, one might ask: What does area V5 do in the brain? The kind of answer a neuroscientist would offer is that it represents information about the color of stimuli and provides this information to other areas which use it to determine the identity of the stimulus. Accordingly, construing internal states as representing various aspects of the environment is critical to this endeavor.

4 From Brain Representations to Cognitive Representations

One response to the line of argument in the previous section is to acknowledge that there may be a role for positing representations that fit the above analysis in neuroscience, but to deny that such representations are sufficient for the business of cognitive science. Indeed, the representations that have figured in many cognitive models employ a much more complex format, often one drawn from formal logic and natural languages. Some have argued that language-like or propositional representations are required if we are to account for the cognitive abilities exhibited at least by humans, since humans must encode complex relational information and be able to extract representations of components as needed. Fodor (1975, 1987) has argued that only representations that share critical properties with language, especially that of having a compositional syntax and semantics, are adequate for modeling thought. When compositional rules are invoked, lexical items are put together according to syntactic rules in such a way that the meaning of the composed structure is built up from the meaning of the component lexical items.

The key to Fodor's argument is the observation that human cognition (and perhaps that of other species) exhibits a number of special properties, especially productivity and systematicity. Productivity and systematicity are properties manifest in natural languages, and Fodor argues that they are exhibited in thought as well. Productivity with respect to language refers to the capacity to indefinitely extend the corpus of sentences in a language; applied to thought, it refers to the fact that the range of possible thoughts is not bounded. Systematicity with respect to language refers to the fact that there are relations between the sentences of a language such that if one string is well formed, so is another that results from appropriate substitutions. For example, if *the florist loves Mary* is a sentence of English, so is *Mary loves the florist*. Applied to thought, it designates the fact that a cognitive system that can think one such thought automatically has the capacity to think the other. In a linguistic system in which sentences are composed employing syntactic rules, these properties arise automatically, and would accrue equally to a cognitive system if it employed representations that are language-like in relying on a compositional syntax. Just as he has faulted the representations found in connectionist networks as incapable of accounting for these properties (Fodor and Pylyshyn, 1988), Fodor would find the sort of representation found in the Watt governor or identified in the activities of individual neurons to lack the requisite compositionality and thus to be incapable of exhibiting these properties.

One unfortunate consequence of grounding explanations of cognitive capacities in language-like representations is that it leaves unanswered the question of how such representations might be embodied in the brain. It is clear that the brain is a mechanism that can comprehend and produce linguistic structures, and so must have tools for representing such structures, but it is far less clear that it uses language-like structures for its own internal representations. So there is motivation for starting with representations of the sort discussed in the previous section – ones that seem to figure in the brain itself. The analysis of representations cannot end there, however. Rather, one must show how to build up from the sorts of representations found in the brain to those that exhibit the requisite compositionality.

While filling in the gap between the sort of neural representations I have been discussing and ones that exhibit productivity and systematicity may seem like a tall order, Larry Barsalou's recent work on concepts suggests how it might be done (Barsalou, 1999). Attacking amodal language-like symbols (symbols not tied to a particular sensory modality), Barsalou has argued that "perceptual representations can play *all* of the critical symbolic functions that amodal symbols play in traditional systems, such that amodal symbols become redundant." Barsalou is clear that the perceptual representations he is considering are neural – he describes perceptual symbols as "records of the neural states that underlie perception." (Although much of his discussion focuses on visual perception, he intends his account to include perception in other modalities, including perception of emotion and introspection.)

The attempt to ground cognition in perception goes back at least to the seventeenth-century Empiricists in philosophy, such as Locke. Their program has been much ridiculed, but the target in most attacks is the view that perception gives rise to static pictures or images (images of which we are consciously aware) that are holistic recordings of the input. Perceptual representations for Barsalou, however, are not (despite his reference to them as "records of neural states") pictures or images – they are not recordings. In particular, they are interpreted in such a way that "specific tokens in perception (i.e. individuals) [are bound] to knowledge for general types of things in memory (i.e. concepts)." The key to this move is a proper understanding of neural processing in vision – the brain is not constructing a picture of the world (if it did, it would then need another perceiver to view the picture), but an analysis of the visual input geared to action. This is already suggested by the way the brain decomposes visual processing, with different brain areas analyzing distinct features of a scene as color, shape, or location. Neural activity in different brain areas represents categorization and conceptualization of the visual input – it contains *this* shape, *this* color, or occurs at *this* location.

Barsalou refers to perceptual representations as schematic representations in that only certain features of the perceptual input are represented. He appeals to psychological research on attention to show how a schematic representation is constructed – selective attention isolates and emphasizes pieces of information that is given in perception and facilitates storage of these features in long-term memory. Recent neural research on attention could support the same analysis. Relying on the evidence that different features of stimuli are analyzed in different brain areas, Corbetta et al. (1993) have shown that when subjects are required to differentially attend to different properties of stimuli, brain areas responsible for processing those features are activated, indicating that particular features are being processed. The fact that perceptual symbols are schematic in this manner allows them to be indeterminate in ways that pictures cannot – representing a tiger, for example, as having stripes, but not a determinate number of stripes.

In addition to emphasizing the schematic character of perceptual representations, Barsalou also emphasizes their dynamic character. Different neural records are related temporally in experience, and they give rise to simulations of the way we can attend to different parts of an object over time or the way it itself changes over time. (Like a perceptual representation itself, a simulation is not just a repetition of previous experiences, but a composed structure in which individual components can be put together differently on different occasions. Barsalou refers to the organizing information specifying how different perceptual representations can be related as *frames*, thereby invoking previous cognitive science research on the type of complex information structures that seem to figure in cognition.) For Barsalou, this allows individual perceptual representations to be integrated into what he terms "simulation competences," a capacity that is expanded as humans learn languages which allow them to index and control features in a simulation.

For Barsalou, linguistic representations extend the capacities of the conceptual system built on perceptual representations. He proposes that:

> As people hear or read a text, they use productively formulated sentences to construct a productively formulated representation that constitutes a semantic interpretation.

> Conversely, during language production, the construction of a simulation activates associated words and syntactic patterns, which become candidates for spoken sentences designed to produce a similar simulation in a listener.

But it is clear that while linguistic indexing supplements the cognitive capacities provided by perceptual symbols, it is the perceptual symbols themselves that do the cognitive work for Barsalou. In fact, linguistic symbols are, for him, acquired as simply additional perceptual symbols. Thus, it is important for him to show that they can have the sorts of properties Fodor argued were needed for cognition – productivity and systematicity – without appealing to language-like representations. Barsalou maintains that the very features of perceptual symbols that I have already reviewed provide him the resources to do this.[4] The key is that perceptual symbols and simulations are built up componentially, and thus, just as with linguistic representations, they can be continually put together in new ways, thereby accounting for productivity. They also permit substitutions of different component representations, thereby accounting for systematicity. Barsalou illustrates this potential by employing diagrams much like those used by cognitive linguists (Langacker, 1987). Figure 18.6 is an example. It illustrates how perceptual symbols for object categories (A) and spatial relations (B) can be (C) combined, even (D) recursively, to productively generate new representations. The symbols in this diagram (e.g. the balloon and airplane in A) are not intended as pictures, but to stand for perceptual representations, that is, configurations of neurons that would be activated in representing these objects. The boxes with thin solid lines are intended to represent simulation competences that have developed over many experiences with the object or relation and represent it schematically. The boxes with thick slashed lines then represent particular simulations that might be generated from the simulation competences by combining them, sometimes recursively.

The preceding is only a partial sketch of Barsalou's account of perceptual symbols (he goes on to suggest how even abstract concepts such as *truth* can be constructed from perceptual representations), but it does indicate that there are ways of building up from the sorts of representations found in the brain. The key ingredient in his account is the construal of the kind of analysis the visual system performs, by having different neurons represent such things as shape and color of stimuli, as involving categorization and conceptualization. The separately analyzed features afford composition, thereby providing a resource similar to that Fodor identified for language-like representations. (Perceptual symbols, however, do not thereby become implementations for Fodorian language-like symbols – perceptual symbols are modality-specific and the particular features of the symbols themselves generally specify definite features in what they represent. Unlike amodal language-like symbols, the particular embodiment of the symbols as patterns of neural firing in particular brain regions is important to the information that they carry. One consequence of this, which Barsalou happily endorses, is that different individuals, with different learning histories, are likely to have somewhat different representations.)

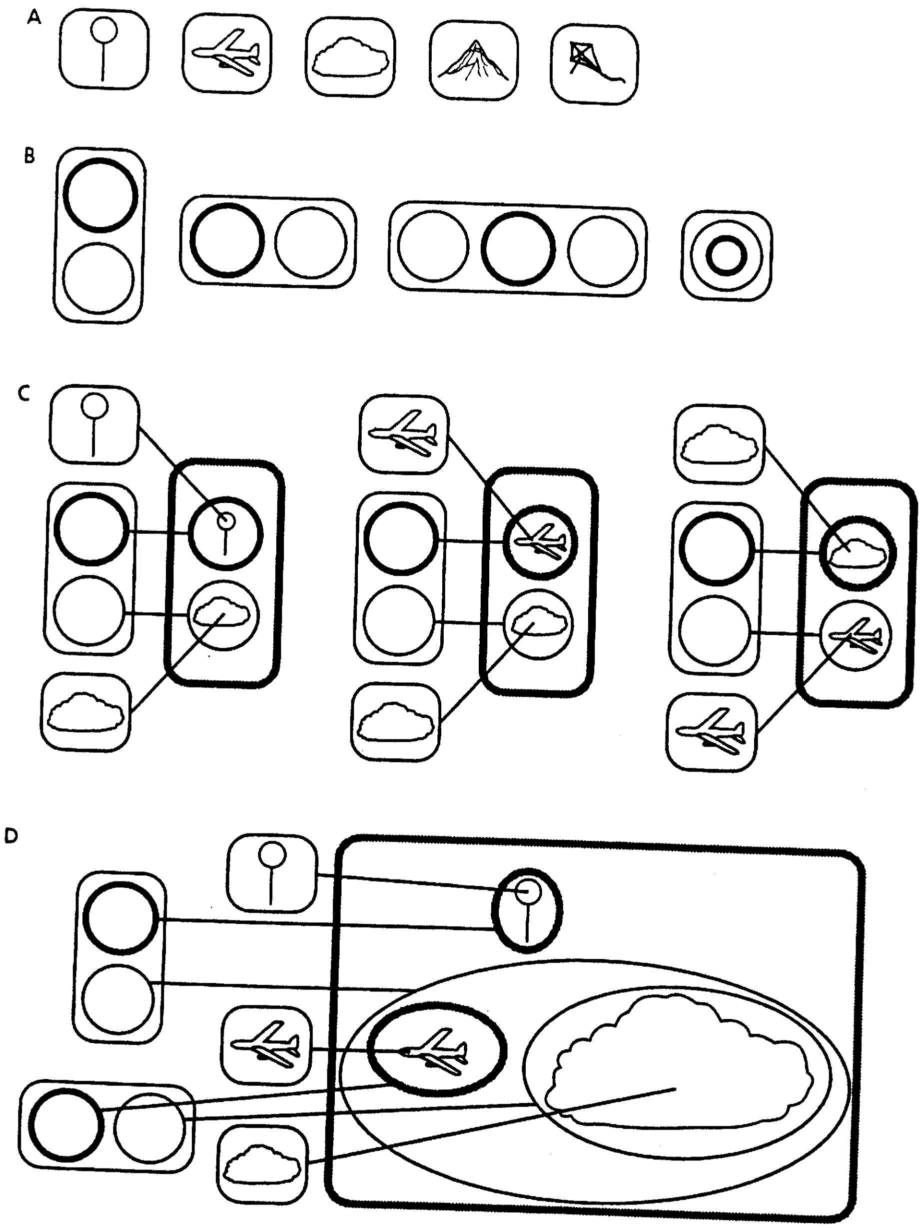

Figure 18.6 Barsalou's representation of how perceptual symbols for object categories (A) and spatial relations (B) implement productivity through combinatorial (C) and recursive (D) processing. Boxes with thin solid lines represent simulation competences; boxes with thick dashed lines represent simulations. From Barsalou, 1999, p. 593.

5 Conclusions

I began this chapter articulating a minimal notion of representation, wherein a representation is an information-bearing state or event which stands in for what it represents and enables the system in which it operates to utilize this information in coordinating its behavior. I argued that representations of this minimal sort are found even in the Watt governor, a simple mechanical device that has figured in some of the opposition to the invocation of representations in cognitive science. I also contended that this minimal notion is what is required for most neuroscientists' references to representations in the brain. In the final section, I tried to show how one might build these neural representations up into representations that exhibit properties such as productivity and systematicity, which have been argued to be characteristics of thought.

Notes

1. For arguments that answer this question in the affirmative, see chapter 21, this volume.
2. Although I am focusing primarily on representations whose content is fixed by what causes them, the accounts can be generalized to representations in the motor system in which the content is specified by what the representations are designed to cause (Mandik, 1999). Although the story gets more complex, one can even conceive of representations for which we need to appeal to both their sensory causes and motor effects in specifying their content.
3. Eventually, researchers hope to discover complete pathways through the system that result in behavioral responses that are appropriate to the information represented at each stage in the system. Until this stage is reached, each imputation of a representation to processes in the system involves taking out a promissory note that can only be repaid by future research.
4. In his own discussion, Barsalou uses the term *productivity* somewhat differently, referring to the ability of subjects to supply instantiations by filling in schemas that were created by filtering out features of the initial perceptual situation. In his treatment of this filling-in Barsalou allows for supplying features that were not part of the initial perception, thus allowing for novelty, including novel representations that violate physical principles. Thus, what he terms productivity is one way of generating new representations, but clearly not the only one present in his account of perceptual symbols.

References

Barsalou, L. 1999: Perceptual symbol systems. *Behavioral and Brain Sciences*, 22.

Batista, A. P., Buneo, C. A., Snyder, L. H., and Andersen, R. A. 1999: Reach plans in eye-centered coordinates. *Science*, 285, 257–60.

Bechtel, W. 1998: Representations and cognitive explanations: Assessing the Dynamicist's challenge in cognitive science. *Cognitive Science*, 22, 295–318.

Clark, A., and Toribio, J. 1994: Doing without representing? *Synthese*, 101, 401–31.

Corbetta, M., Miezin, F. M., Shulman, G. L., and Petersen, S. E. 1993: A PET study of visuospatial attention. *Journal of Neuroscience*, 13 (3), 1202–26.

Cummins, R. 1996: *Representations, Targets, and Attitudes*. Cambridge, MA: MIT Press.

Dretske, F. I. 1981: *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press/Bradford Books.

Fodor, J. A. 1975: *The Language of Thought*. New York: Crowell.

Fodor, J. A. 1987: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

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

Goldman-Rakic, P. S. 1987: Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In J. M. Brookhart, V. B. Mountcastle, and S. R. Geiger (eds), *Handbook of Physiology: The Nervous System*, vol. 5. Bethesda, MD: American Physiological Society, 373–417.

Griffiths, P. E. 1997: *What Emotions Really Are*. Chicago: University of Chicago Press.

Haselager, W. F. G., and de Groot, A. D. (forthcoming): Representationalism versus anti-representationalism: a debate for the sake of appearance?

Hubel, D. H., and Wiesel, T. N. 1962: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160, 106–54.

Hubel, D. H., and Wiesel, T. N. 1968: Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, 195, 215–43.

Kuffler, S. W. 1953: Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16, 37–68.

Langacker, R. 1987: *Foundations Of Cognitive Grammar (Vol. 1)*. Stanford, CA: Stanford University Press.

Mandik, P. 1999: Qualia, space, and control. *Philosophical Psychology*, 12, 47–60.

Millikan, R. G. 1984: *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Penfield, W., and Rasmussen, T. 1950: *The Cerebral Cortex in Man: A Clinical Study of Localization of Function*. New York: Macmillan.

Snyder, L. H., Batista, A. P., and Andersen, R. A. 1997: Coding of intention in the posterior parietal cortex. *Nature*, 386, 167–70.

Talbot, S. A., and Marshall, W. H. 1941: Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophthalmology*, 24, 1255–63.

van Gelder, T. 1995: What might cognition be, if not computation. *Journal of Philosophy*, 92, 345–81.