This is a draft manuscript. Comments are most welcome,
*but please do not quote without permission***.**

Lecture 4
Representations in the Brain: Building to Higher Cognition

One of the central elements in the conception of mechanism which I have been employing is that mechanisms perform activities. It is these activities that guide the decomposition of the mechanism into components and operations and which are to be explained by these components and operations as they are coordinated within the mechanism. I have been focusing on mechanisms within the brain, but have not focused on the brain as a whole. What is its activity? Most generally, the activity of the brain is to control behavior. In the simplest nervous systems, this involves responding to immediate stimuli so as to satisfy immediate needs. In more complex nervous systems, this involves planning and executing behavior over longer durations to achieve goals.

In both the simpler cases and the more complex cases, the challenge for the organism is in part to coordinate its behavior with things external to it. Often the relevant external phenomena are still immediately available to the organism through its senses. Other times they are removed in space and time. In either case, in order to coordinate the organism's responses, the brain must acquire information about the external phenomena. The term *representation*, is widely used to designate an internal state or process which carries this information. Usually the processes of interest are firing patterns of individual neurons or activity in a particular region of the brain, although one might also consider the configurations of synapses that, for instance, enable the brain to recreate previous patterns of activity. (In discussions of representation one often distinguishes the vehicle of representation from the content. In terms of that distinction, these states or processes constitute the vehicle, while that about which they carry information is the content.) The root idea here is that representations *re-present* the external in a format that facilitates the selection of behavior. My first goal in this lecture is to clarify this notion of representation. As I will show, this notion of representation extends beyond brains to control system in general.

The term *representation* is also widely used in the cognitive sciences. For example, in explanations of problem solving behavior, cognitive scientists typically 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 (Miller, Galanter, & Pribram, 1960; Simon, 1996). Unlike neuroscientists, cognitive scientists have typically not had access to states or processes in the brain to identify the vehicles of representation. Instead, them frequently have introduced representations as theoretical posits. Generally one conceptualizes theoretical entities on analogy with entities one is otherwise familiar with, and cognitive scientists have frequently looked to representations used in our culture for their conception of representational vehicles. Since natural languages provide one of the most powerful representational systems humans have devised, it is perhaps not surprising that language-like symbols have frequently been invoked as models for mental representations. But other representational systems humans have devised, such as drawings and pictures, have also been advanced as models. More recently a model drawn from the brain itself has been advanced in connectionist modeling. Many of the most vociferous debates in the cognitive science literature have focused on the requirements on the vehicle for representations in cognitive models.

Now that neuroscience and cognitive science are meeting in cognitive neuroscience, the relation between neuroscience and cognitive science conceptions of representation have taken on greater import. The second objective of this lecture is to sketch a way in which one might envisage building up from the representations neuroscientists invoke in understanding the brain to those thought to be necessary for cognition.

1. Representations in the Brain

The search for representations in the brain begins once one conceives of the brain as acquiring information about the organism's body and its environment from sensory stimuli that can then be used to guide behavior. Thus, one place that one finds appeals to representations is in explaining the operation of visual processing systems. Although the interest in neuroscience is largely in the representations themselves (e.g., whether it is the firing rate of neurons or more specific features of the firing pattern that serves as the representational vehicle) and how they are used in the system, the representations have usually been identified in terms of their ability to carry information. In the simple case in which the question is how the brain acquires information about what is currently present in its environment, the notion of information can be unpacked in strictly causal terms. Any causal effect of something carries information about it (Dretske, 1981). Beginning then with at least a general idea about what the vehicles of representation are, researchers begin by trying to determine which vehicles carry information about which external stimuli.

Although there are various neural states or processes that are candidate vehicles of representation, the action potentials in individual neurons were an initially plausible candidate, once the spiking activity of neurons was confirmed through the research of Edgar Adrian. After he devised means of recording the electrical signal in a single neuron in a suspended frog leg by cutting away all but one tract, he was puzzled by its irregular pattern, but soon arrived at the hypothesis that the neuron conducted electricity in an all-or-none fashion (Adrian & Zotterman, 1926). Adrian not only established the all-or-none character of the action potentials, but concluded that the variation in frequency of action potentials carried differential information: "the frequency of the discharge controls the intensity of the effect which the message produces and is itself controlled by the intensity of excitation" (Adrian & Bronk, 1929). In his 1928 book *The basis of sensation: The action of the sense organs*, Adrian explicitly refers to the nerve fibre as transmitting information. Referring to the research of Francis Gotch and his mentor Keith Lucas that identified the refractory period after an action potential, he says it "gave us for the first time a clear idea of what may be called the functional value of the nervous impulse. They showed what the nerve fibre can do as a means of communication and what it cannot. . . It is of the first importance in the problems of sensation, for it shows what sort of information a sense organ can transmit to the brain and in what form the message must be sent" (Adrian, 1928).

If action potentials carry information, then it is important to determine what information they carry. Since much of his research was focused on action potentials in motor neurons, Adrean viewed the action potentials as specifying actions. (Although this reverses the direction found in sensory systems, it requires only a small change in the model to allow that action potentials can carry information specifying motor activity. See (Mandik, 2002) Adrian also focused on sensory processing systems.[1] Lesion studies in humans and other species in the late 19th and early 20th centuries had established that the postcentral gyrus was involved in sensing cutaneous sensation and Wilder Penfield, following up on studies by Harvey Cushing (Cushing, 1909), mapped out the projection of sensory areas onto Brodmann's areas 1, 2, and 3. These are illustrated dramatically in the sensory homunculus in Penfield and Rasmussen (Penfield & Rasmussen, 1950) famous figure (Figure X). By presenting stimuli to the paws of cats and recording from an area just above the Sylvian fissure (Brodmann's area 43), Adrian (1940) discovered a second area that responded to stimulations of the paw, a result that was soon generalized to other species by Clinton Woolsey (Woolsey, 1943; Woolsey & Fairman, 1946) and others. Determining why there are two representations and how they might differ became a topic for subsequent research.

If neurons divide up the representational task so as to be able to individually carry information about different features of the world external to the organism, one needs a great deal of luck to find the stimulus that will cause a response in an individual neuron. Fortunately, other techniques, such as lesion experiments, provide researchers with clues as to the type of stimuli that might cause responses in any given cell. On the basis of lesions, researchers in the late 19th and early 20th were able to identify a topographical map in primary visual cortex. The first use of single-cell recording in visual cortex was to confirm these maps (Talbot & Marshall, 1941).

Maps such as Talbot and Marshall's suggest a one-to-one mapping of stimuli and responses in the visual cortex, but in fact each cell typically responds to stimuli in a range of the visual field. This is known as the *receptive field* (Hartline, 1938). A crucial step beyond identifying the receptive field of a cell was to determine just what stimuli in the receptive fields would generate large, or even the greatest response. Steven Kuffler (1953) demonstrated that cells in precortical processing areas, the retina and the lateral geniculate nucleus, responded to small light spots surrounded by a dark field, or dark spots surrounded by a light field (these are known as *center-surround* stimuli. Continuing in this endeavor, David Hubel and Thorsten Wiesel began experiments on monkeys and cats trying to stimuli that would cause responses in cells in primary visual cortex. They assumed initially that the cortical cells would respond to the same sort of

[1] One of the things Adrian discovered was that the response of neurons adapts so as to respond less and less to a stimulus that remains constant. One important consequence of this is that whatever content action potential represent, they represent changes in that content.

stimuli and tried, without success, to find cells that would respond. On one occasion, as the slide was dropping into the projector, a bar of light crossed the screen and the cell next to which they had placed the electrode responded vigorously ("over the audiomonitor the cell went off like a machine gun" (Hubel, 1982, p. 439). With that as a clue, they began testing cells with dark and light bars.

Over the first ten years of their collaboration, Hubel and Wiesel probed the striate cortex of both cats (Hubel & Wiesel, 1962) and monkeys (Hubel & Wiesel, 1968) and discovered a rich organization of cells with different response patterns. What they termed *simple cells* had receptive fields with spatially distinct *on* and *off* areas along a line at a particular orientation (most typically, they had a long, narrow *on* areas sandwiched between two more extensive *off* areas). Hubel and Wiesel proposed how several cells with center-surround receptive fields (such as those found in the LGN) might all send excitatory input to a single simple cell. In this regard, it is salient that simple cells predominate in layer 4, which is the input layer to cortex. Whereas simple cells were sensitive to stimuli only at a given retinal location, what Hubel and Wiesel termed *complex cells* were responsive to bars of light at a particular orientation anywhere within their receptive fields. Many complex cells were also sensitive to the direction of movement of bars within their receptive field. Hubel and Wiesel identified these as *complex cells* since their response pattern could be explained if they received input from several simple cells, any of which would be sufficient to cause the complex cell to fire. Complex cells are found primarily in layers 2 and 3 and 5 and 6.[2] In their papers from this period Hubel and Wiesel also distinguished *hypercomplex cells* which responded maximally only to bars extending just the width of their receptive field.

With the research of Kuffler, Hubel, and Wiesel, the idea that the firing of individual cells carried specific information about sensory stimuli and hence represented specific aspects of the stimulus gained credibility. As I discussed in lecture 2, in the three decades after Hubel and Wiesel's initial research, researchers identified a host of additional brain areas in the prestriate parts of the occipital lobe and regions in the temporal, and parietal lobes that figured in visual processing. They have also been able, in many cases, to determine the type of information to which cells in these areas are responsive. Although frequently guided by lesion results, the discoveries generally relied on extensive single-cell recording trials in which many stimuli were presented until ones were found which would elicit activity in the cell. Often the discoveries were fortuitous, as in Gross's discovery of an area in inferior temporal cortex that would respond to the shape of a hand. Two figures by Van Essen and Galant (1994) graphical representation the 33 brain areas that have been identified as principally involved in visual processing and the type of information cells in these areas represent.

Although the task of identifying cells in the brain that would respond to selected stimuli was arduous, it does not suffice for establishing representations in the brain. It only

[2] An important difference between the different layers is that they generally project to different brain areas: layers 2 and 3 to other cortical areas, layer 5 to the superior colliculus, pons, and pulvinar, and layer 6 back to the LGN.

shows that the firing of these cells carries information about that a particular type of stimulus. But carrying information is not sufficient to render something a representation since every event in any causal process carries information about the cause. Why focus on some of these as representations?

Actually, there are two issues involved in differentiating mere information-carrying from representation. First, in order for a cell in a brain area relatively far along a processing pathway (for example a cell in MT or a cell in inferior temporal cortex) to respond to a particular kind of stimulus cells earlier in the pathway must also carry the same information. So why speak of cells in MT as representing motion and cells in the inferior temporal cortex representing hand-shapes when cells in the LGN, V1, and V2 must have carried the same information? There are two relevant differences between, for example, V1 cells and MT cells. First, the receptive fields of MT cells are much larger—they respond to motion in a particular direction in a wide area of the visual field whereas V1 cells have a very small receptive field. This enables MT cells to represent to a particular kind of motion, not just motion in a small area. Second, cells in MT are specifically tuned to respond to motion whereas the cells in earlier areas are tuned to respond to different features such as a bar of light. Accordingly, we can say that the information was only implicit in the activity of cells earlier in the pathway and explicit in MT cells.

The second issue in differentiating representation from information carrying can be recognized by thinking again of representations as re-presenting something. There must be a user to which the stimulus is being re-presented. Although the picture of a homunculus, or little person, sitting inside the brain looking at the representations is rightly rejected, there must be something that responds differentially to the representation in a manner that depends on what it represents. In many cases the question of a user of a particular representation is left implicit. The fact that there are feedforward projections from each of the brain areas identified as representing different things suggests that other brain areas are in fact responding to the earlier areas. Insofar as researchers both develop hypotheses as to what these later areas are responding to and develop models of how they arrive at are able to represent specific features of the visual presentation in virtue of what was represented in the earlier areas, they fill in the account of the user.

There is a further means of ascertaining whether there is a user of the representation and that is to show that what is represented has consequences for behavior. An example I offered in the previous lecture, Newsome's research on perceived motion (Britten, Shadlen, Newsome, & Movshon, 1992), relied on the correlation between the activity in MT and the monkey's behavioral response indicating the direction in which it perceived motion, shows how this can be done in practice. By showing both that the monkey responded to ambiguous stimuli in accord with the activity in MT and that microstimulation of particular MT cells could bias the monkey's response, Newsome brought information about how the representations were used into the assessment of the representations themselves.

Focus on the user of a representation can be particularly informative when there is uncertainty about what the organism is actually represented. For example, areas in parietal cortex are involved in processing information about location. But in a task in which the organism is presented a stimulus to which it responds by either saccading or moving its arm there, does the activation of particular cells indicate *attention* to the location or the *intention* to make a delayed movement to the location? The task is challenging since attention is usually directed towards the target locations of a motion. Larry Snyder and colleagues tried to differentiate the two possibilities. A stimulus was presented at a specific location to which the animal had to make a delayed response, with the color of the stimulus specifying the response. They reasoned that if the activation represented attention to the location, then the specific response the animal was planning (saccading versus reaching) should make no difference. If it was sensitive to the activity to be performed, then it would represent an intention. They found cells in area LIP which responded more when the animal was instructed to saccade to a location in its response field and cells in area PRR (parietal reach region) which showed the reverse pattern, indicating that both areas encoded intention to act (Snyder, Batista, & Andersen, 1997; Snyder, Batista, & Andersen, 2000).

A word of caution is needed. Since we are engaged in reverse engineering the brain and do not have independent access to its design, 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 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 (1994) found that esoteric stimuli, such as expanding stimuli 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.

Another difficulty as Kathleen Akins (1996) emphasizes, neurons do not respond to objective features of the world. In particular, they do not respond to the features that we might think brains would represent if we cast ourselves in the position of a designer of a brain. 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). Akins takes this as a reason to reject a representational analysis, but it seems rather to be a reason to reject a particular account of the content of neural representations and as providing useful advice about what we should look for as contents of representation. As several critics have noted, organisms are not trying to build up complete pictures of the world they inhabit, but acquiring information that is useful in guiding their action (Churchland, Ramachandran, & Sejnowski, 1994). To figure out what about the world an organism represents, then, we need to move beyond a stance of looking at the external world through the lens of our sciences and focus on the needs of the organism (a perspective long advocated in perception by James Gibson (1966), although he opposed a reprenentationalist account of the organism's inner activity).

2. Representations as a General Feature of Control Systems

So far I have focused on the process by which neuroscientists identify what they are willing to call representations in the brain. In this section I turn to the challenge of fleshing out this analysis. Three components emerged from the analysis I developed above—something represented, the representation, and the user of the representation. What the user does with the representation has not been specified in much detail. In the simplest case, the user of the representation uses it to operate directly on what is represented or with respect to what is represented, for example, to move around it. In this case, the loop is closed as in figure X.

This is a model for a simple control system. An exemplar of such a control system is the governor James Watt built for the steam engine. 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 (see Figure X). 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 of 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.

The operation of the Watt governor illustrates the basic scheme I presented in Figure X. The speed of the engine is what is represented, the angle of the arms is the representation, and the linkage mechanism controlling the steam valve is the user of the representation.

There is a bit of perversity in taking the Watt governor as an exemplar of a representational system. It was, after all, Timothy van Gelder's (van Gelder, 1995) example in his argument against treating the mind as a representational system. His contention was that the Watt governor was an example of how a system could coordinate its behavior with things in its environment without representing them. I have responded to his arguments against identifying representations in the Watt governor elsewhere (Bechtel, 1998), and won't rehearse those responses there. But I do need to say something by way of motivating using such a simple system as an exemplar of a representational system besides noting how well it corresponds to the framework that emerged from considering how neuroscientists speak of representations.

The first thing to note is that the Watt governor is a sub-system of a larger system that is designed to determine the behavior of the part of the larger system that is designed to act on the environment (the plant). More specifically, it is the part of the larger system that is to coordinate the activity of the plant with the demands of environment. It is needed by that larger system since otherwise there is no way for the larger system to respond appropriately to the demands placed upon it. In order to produce the right behavior, the larger system must have within it something that stands in for the relevant feature of the environment in a format that can be utilized to control the behavior.

One might object that one can specify the total operation of the Watt governor in the steam engine in purely causal/mechanical terms. That, in itself, should be no objection, since in the context in neuroscience on which I am focusing the goal is to offer a purely causal/mechanical account. But the objection is: once one has described the motion of the flywheel, spindle, angle arms, mechanical linkage, and valve, what else is required? In particular, why add representations to the story. To see why representations are needed, recall that part of identifying a mechanism is to specify what activity it performs. The activity of the Watt governor is to insure that the steam engine runs at a constant speed irrespective of the resistance from the work being performed by the attached appliances. To explain how it performs that activity, we need to explain how it is that the governor opens or shuts the valve *appropriately* to the work being done.[3] It is not sufficient to detail the causal relations occurring in the system over time.

To see this, imagine Watt being confronted with the question: Why did you stick all this elaborate apparatus onto the machine? Couldn't we just leave it off? To answer this, Watt would have to point out that the challenge was to open and shut the valve as needed in response to the current speed of the engine. In order to do this, information about the current actual speed of the engine needed to be made available to the component that would actually open and shut the valve in a format that it could use. That is to say, the information about the current speed of the engine had to be re-presented in a format that could be directly used to control the valve.

The Watt governor is an example of what we now recognize as the simplest form of a control system, one that uses negative feedback control. In this case, information about the activity being performed is fed back from that which is operated on so as to alter in appropriate ways the operation of the plant. The representation directly determines the operation performed on that which is represented. Although there are plenty of examples of this type of control system in biological systems, and such feedback systems were the inspiration for the first generation of cybernetics (see, for example, the classic paper (Rosenblueth, Wiener, & Bigelow, 1943), many control systems are far more elaborate. The activity animals often have to perform with respect to something in their visual field is not to act on it, but to avoid it (e.g., if it is a predator, or even just an obstacle in its path). In this case the control system is controlling behavior but in order to do so it requires information about something else in the environment which may or may not affect the success of the behavior. Moreover, animals often need to coordinate their action with things that are not immediately present, such as a food source that is out of sight. In this case there are no direct causal connections from the thing represented to the

[3] Someone might object that the invocation of appropriateness in this characterization of the system's behavior betrays the fact that we are not dealing with something objective in the world. It might be useful to us to construe a system as representational, but nothing in the world is objectively a representation. But appropriate response is not something merely subjective. It is an objective feature of a system whether it fulfills the demands made upon it in the context of a larger system of which it is a component.

representation. The representation must be maintained in the absence of occurrent causal input. Even more difficult is that the control system may have to take into account changes occurring in what is represented during periods in which it is out of causal contact. This requires the system to represent the dynamic activity of something remote and out of causal contact.

In an illuminating example, Rick Grush (1997) considers the example of an earth-bound controller for a heating plant on a remote space station. By the time feedback information is received by the controller that the temperature has dropped too low and it sends back a command to increase the heat, the temperature has dropped even further. Eventually the feedback system will begin to restore the system to the proper temperature, but by the time that information is received by the controller and it can issue a command to stop increasing the heat, the temperature has will have risen far beyond the target range. To deal with such systems, engineers have devised control systems that maintain a model of the plant—an emulator. Grush envisions, for example, that before the space station was launched, a neural network was trained to emulate the heating plant. Its outputs specify the temperature produced by a certain activation of the heating or cooling system on the space station. After the space station is launched, the emulator stays behind and is used by the local control system to determine what is happening remotely. The local control system then responds to the predictions of what is happening remotely rather than waiting for the signal from the space station and so avoids the disastrous oscillations that would result from utilizing only feedback control.

Grush's main objective is to point out the virtues of relying on such emulator systems and to argue that they constitute the point at which we are properly led into introducing representations. He construes the sort of internal processes within the visual system that I have focused on as *presentations* and differentiates them from representations. While granting that there is a significant evolutionary advance from internal states that are generated from the environment relatively directly from those that are maintained and employed when this connection is broken, and that the latter provide the organism with opportunities which the former do not, I nonetheless maintain that there is a point in maintaining the continuity between the two sorts of processes. In both cases, something is standing in for something else and used to coordinate behavior with that which is represented.

3. Representations in Cognitive Theories

As I noted at the outset, one of the hallmarks of *cognitive* explanations of behavior is that they appeal to mental representations and operations over them. So far I have focused on representations in neural systems and control systems more generally. Are the sorts of things that count as representations in these endeavors adequate for the purposes for which cognitivists have appealed to representations? Interestingly, the account of representation I have advanced here is very similar to that which Newell offers for a symbol system:

> "The most fundamental concept for a symbol system is that which gives symbols their symbolic character, i.e., which lets them stand for some

> entity. We call this concept *designation*, though we might have used any of several other terms, e.g., *reference, denotation, naming, standing for, aboutness*, or even *symbolization* or *meaning*" (Newell, 1980, p. 156).

Newell goes on to offer a definition of designation:

> *Designation*: An entity X designates an entity Y relative to a process P, if, when P takes X as input, its behavior depends on Y.
>
> There are two keys to this definition: First, the concept is grounded in the behavior of a process. Thus, the implications of designation will depend on the nature of this process. Second, there is action at a distance . . . This is the symbolic aspect, that having X (the symbol) is tantamount to having Y (the thing designated) for the purposes of process P (Newell, 1980, p. 156).

Based on this notion of a symbol system, Newell and Simon advanced *The Physical Symbol System Hypothesis*—the hypothesis that "a physical symbol system has the necessary and sufficient means for general intelligent action."

Yet, on further examination, the answer as to whether we could build from neural representations to cognitive representations might seem to be negative. The primary model for representations in cognitive science has been linguistic representations (Fodor, 1975). This is not just because one part of cognitive science, artificial intelligence, involves programs specified artificial languages and uses systems which store representations in the structures posited in these languages in data structures. It is also the case that many of the accounts in which mental representations figure in cognitive science, such as accounts of reasoning and problem solving, seem to require representations with the sort of complexity found in linguistic representations.

The most explicit arguments that language-like representations are required to model cognition are found in the work of Jerry Fodor and his collaborators. Early in his research Fodor introduced the idea of a language of thought (Fodor, 1975) in which a cognitive system could construct and test hypotheses. In response to the advocacy of some cognitive scientists of systems that did not seem to operate over language-like representations but rather to rely on associations of simple representations, Fodor developed arguments attempting to show that an adequate representational system for cognition had to involve a compositional syntax and semantics. Only by positing operations that operated on such representations in virtue of their syntax, he claimed, could one account for important features of cognition, such as its productivity and systematicity (Fodor & Pylyshyn, 1988).

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 & 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 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 –one's that seem to figure in the brain itself. But the analysis of representations cannot end there. 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 17th 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 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 serving to analyze distinct features of a scene as color, shape, or location. Neural activity in different brain areas represents categorization and conceptualization of the visual input—specifying that 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 is 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, Miezin, Shulman, & Petersen (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."

For Barsalou, linguistic representations enable people to index and control features in a simulation, extending 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 underlying their use. 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 X 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.

INSERT FIGURE X ABOUT HERE

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 be ways of building up from the sorts of representations found in the brain. The key ingredient in his construction is to construe 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.)

Conclusion

[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.

The challenge for this lecture was to articulate a notion of representation that both accounted for representations in neuroscience and provided a bridge to the use of representations in cognitive theories. The key idea in the analysis of representation that I developed was that a state in a mechanism was a representation when it carried information about something else that could then be used in determining the behavior of the mechanism. In neuroscience representations in different processing systems are identified in terms of the information they carry and the ways in which they can be used by other parts of the neural system. Barsalou's conception of a perceptual symbol system provides a bridge from such neural representations to cognitive representations. The states normally activated in the context of sensory stimuli can be activated and deployed in mental simulations. Such simulations can be indexed linguistically and utilized in reasoning about the situations they represent. To show that such neurally-based representations can meet all the demand imposed by cognitive processing requires more specific development, but the general strategy is both clear and promising.

References

Adrian, E. D. (1928). *The basis of sensation: The action of the sense organs*. New York: W. W. Norton & Company.

Adrian, E. D. (1940). Double representation of the feet in the sensory cortex of the cat. *Journal of Physiology, 98*, 16P-18P.

Adrian, E. D., & Bronk, D. W. (1929). The discharge of impulses in motor nerve fibres. Part II. The frequency of discharge in reflex and voluntary contractions. *Journal of Physiology, 66*, 119-151.

Adrian, E. D., & Zotterman, Y. (1926). The impulses produced by sensory nerve endings. Part 2: The response of a single end-organ. *Journal of Physiology, 61*, 151-171.

Akins, K. (1996). On sensory systems and the 'aboutness' of mental states. *The Journal of Philosophy, 93*(7), 337-372.

Barsalou, L. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences, 22*, 577-660.

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

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience, 12*(12), 4745-4765.

Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch & J. L. Davis (Eds.), *Parge-scale neuronal theories of the brain*. Cambridge: MIT Press.

Corbetta, M., Miezin, F. M., Shulman, G. L., & Petersen, S. E. (1993). A PET study of visuospatial attention. *The Journal of Neuroscience, 13*(3), 1202-1226.

Cushing, H. (1909). A note upon the faradic stimulationof the post-cenral gyrus in conscious patients. *Brain, 32*, 44-53.

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambrdige, MA: MIT Press/Bradford Books.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

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

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology, 10*, 5-24.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate retina. *American Journal of Physiology, 113*, 59-60.

Hubel, D. H. (1982). Evolution of ideas on the primary visual cortex, 1955-1978: A biased historical account. *Bioscience Reports, 2*, 435-469.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London), 160*, 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London), 195*, 215-243.

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. (2002). Points of view from the brain's eye view: Subjectivity and neural representation. In W. Bechtel & P. Mandik & J. Mundale & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 312-327). Oxford: Blackwell.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex in man: A clinical study of localization of function*. New York: Macmillan.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose, and teleology. *Philosophy of Science, 10*, 18-24.

Simon, H. A. (1996). *The sciences of the artificial* (Third ed.). Cambridge, MA: MIT Press.

Snyder, L. H., Batista, A. B., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature, 386*, 167-170.

Snyder, L. H., Batista, A. P., & Andersen, R. A. (2000). Intention-related activity in the posterior parietal cortex: a review. *Vision Research, 40*, 1433-1441.

Talbot, S. A., & Marshall, W. H. (1941). Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Opthalmology, 24*, 1255-1263.

van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13*, 1-10.

van Gelder, T. (1995). What might cognition be, if not computation. *The Journal of Philosophy, 92*, 345-381.

Woolsey, C. N. (1943). "Second' somatic receiving areas in the cerebral cortex of cat, dog, and monkey. *Federation Proceedings, 2*, 55.

Woolsey, C. N., & Fairman, D. (1946). Contralateral, ipsilateral and bilateral representation of cutaneous receptors in somatic area I and II of the cerebral cortex of pig, sheep and other animals. *Surgery, 19*, 684-702.