This is a preliminary version from P. Thagard, *Mind: Introduction to Cognitive Science*, second edition, to be published by MIT Press in February, 2005. For the references, see http://cogsci.uwaterloo.ca/Bibliographies/cogsci.bib.html.

Chapter 9

Brains

You probably know that your ability to think depends on your brain, but understanding of how processes in the brain contribute to thought is relatively new. Some ancient Greek philosophers believed that the brain is one of the organs of thought, along with the heart and liver; but Aristotle argued that the brain is merely a cooling device for the heart, which he took to be the center of intellectual and perceptual functions (Finger, 1994). This chapter reviews the main methods for investigating the nature of brains and how they contribute to thought, ranging from looking at the effects of brain damage to using machines to scan brain activities. It then discusses how discoveries about brain processes have enriched our understanding of the representations and computations that produce thinking. The chapter then considers the relevance of molecular processes involving neurotransmitters for understanding the relation between thought and brain, and discusses the practical applicability of knowledge about the brain, especially in the treatment of mental illness.

HOW BRAINS ARE STUDIED

Brain Structure and Lesions

The first important method for studying brains was dissection, in which brains are carefully cut apart to reveal their anatomical structure. In the second century A.D., the Roman physician Galen described many brain structures based on dissection of animals such as cows and baboons. The results of dissection of human brains were not reported until the sixteenth century when Vesalius provided detailed anatomical descriptions. He thought that the ventricles, which are open spaces in the brain, are crucial for thought because they produce “animal spirit” which is distributed to the nerves. Studying the anatomy of the brain does not by itself reveal much about its physiology, i.e. how it works.

Insights into the physiology of the brain and its relation to thought came about by the method of *lesions,* which are injuries to specific parts of the body. Lesions of the brain can occur naturally because of tumors or blood clots, or they can be produced by cutting or burning. In the eighteenth century, lesion experiments on dogs revealed that breathing depends on a brain area called the *medulla*: damage to the area causes severe breathing problems. The first recognition that a human cognitive function depends on a specific brain area came in the 1860s when Paul Broca attributed a patient’s inability to use language to a specific part of the brain’s frontal lobes now called Broca’s area. Since then, the contributions of many specific parts of the brain to particular cognitive functions have been discovered, as illustrated in figure 9.1. For example, people with damage to the hippocampus have difficulty forming new memories, and damage to the amygdala can cause inability to feel fear and other emotions.

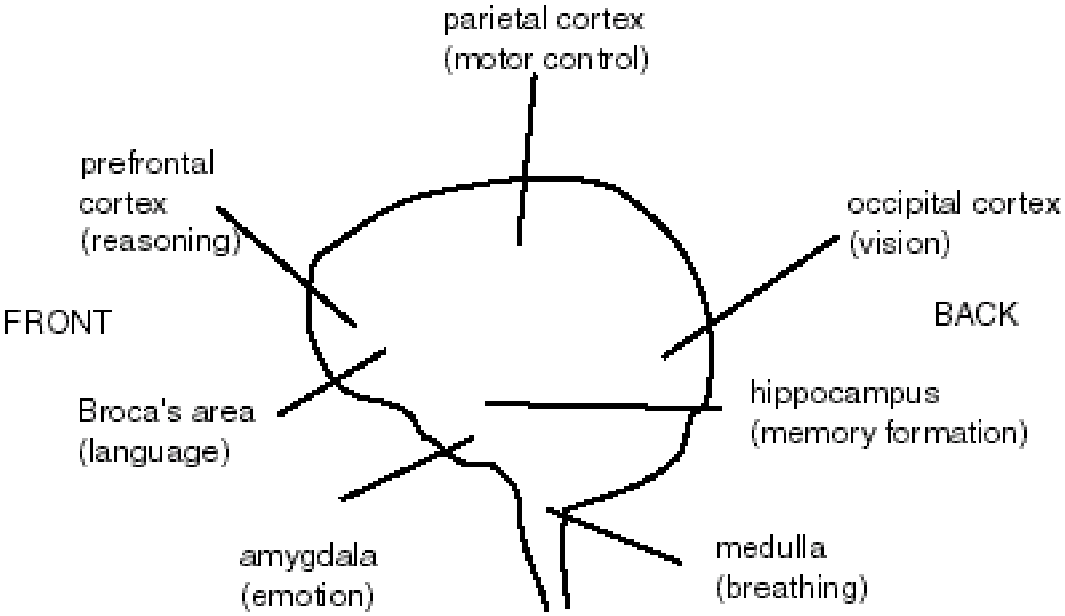

Figure 9.1. Some important brain areas, with their associated functions. For much more detailed maps, see the web sites listed at the end of this chapter.

Electrical Recording and Stimulation

In 1875, Richard Caton reported the existence of electric currents in the brain that vary with different stimuli presented. This made it possible to determine what parts of the brain are active without having to rely on lesions. Today, electrical activity in the brain is recorded by a machine called an *electroencephalograph*, or EEG for short. If you go for an EEG, you will have electrodes attached to your head that transfer to a computer a history of the electrical activity in different parts of your brain. EEG's can be used to diagnose epilepsy, which is brought about by abnormal electrical discharge of brain cells.

EEGs can only identify electrical activity in large regions of the brain, but the electrical activity of particular neurons can be identified by the technique of single cell recording. Electrodes are inserted into the brain to record the firing activity of specific neurons. For example, there are neurons in monkeys that respond to the texture of fat in the mouth, with especially high firing rates when the monkeys are given heavy cream (Rolls, 1999, p. 34). Single cell recording is too invasive to be used routinely on humans. but is sometimes used during brain surgery to identify the contribution of particular neurons to cognitive functions.

Electrical activity in the brain can be stimulated as well as recorded, for example during surgery when current is applied to exposed parts of the brain. Less invasively, transcranial electronic stimulation is performed when electrodes are placed on the head to make current flow through the brain. Alternatively, transcranial magnetic stimulation applies powerful electromagnets to stimulate or disrupt brain activity.

Brain Scans

Whereas EEGs can identify activity in large brain regions and single cell recording applies to individual neurons, neither method reveals much about small brain areas. In contrast, modern brain imaging technologies can identify activity in areas a few millimeters (less than 1/8 inch) across, comprising a few million out of the 100 billion or so neurons in the brain. The two technologies that are currently most useful are positron emission tomography (PET) and magnetic resonance imaging (MRI).

If you go for a PET scan, you will first be injected with a radioactive material that spreads through your bloodstream. The most active parts of your body, including particular brain regions, will require increased blood flow to nourish the cells that are producing the activity. The PET scanner detects increases in radioactivity in blood flow, thereby identifying groups of cells such as neurons that are active. So if you are given a particular mental task to perform, the PET scan may be able to detect what parts of your brain are used to perform it. For example, if you are given a picture to look at, the PET scan will show increased blood flow to the primary visual cortex located at the back of your brain.

PET scanning has a number of limitations, including the use of radioactive material and the inability to localize activity to regions smaller than a cubic centimeter. Hence it is now less commonly used than magnetic resonance imagery, which was originally developed to display the structure of parts of the body. If you go for an MRI, you will be inserted in a large magnet that generates signals from the hydrogen nuclei of water molecules in your body. The MRI machine detects these signals and uses computers to distinguish physical structures based on different signals that they generate. For example, MRIs are often used to diagnose sports injuries by detecting changes to joints and other structures.

To investigate brain processes, specific magnetic pulses can be generated that enable the detection of changes in blood flow; this is called functional MRI, or fMRI. Hence fMRI, like PET scans, can be used to identify brain regions with increased blood flow responding to increased neuronal activity. Images of the brain can be produced in a few seconds with a spatial resolution of a few millimeters. Unfortunately, fMRI does not have the temporal resolution of electrical recording technique such as EEGs, which can detect changes in brain activity that take much less than a second. Nevertheless, fMRI studies have become crucial for helping to identify the specific brain regions involved in various kinds of thinking. For a history and review of techniques for brain mapping and imaging, see Savoy (2001) and Posner and Raichle (1994).

Now that techniques are available for identifying activity of brain regions and even of single neurons, do we still need the computational-representational understanding of mind? Why not explain thinking directly in terms of neuronal activity without talking about rules, concepts, and other representations? Why not focus on the kinds of physical processes found in the brain rather than on computational processes? To answer these questions, we need to examine how brains exhibit representational and computational properties.

HOW BRAINS REPRESENT

Spiking Neurons

A representation is a structure that stands for something by virtue of relations such as similarity, causal history, and connections with other representations. For example, a photograph is a representation of you because it looks like you and because photography causally links it with you. The word “cat” is not similar to cats, but there is a causal link between utterances of this word and the presence of cats, as well as relations between the concept *cat* and other concepts. Let us look at how individual neurons and especially groups of neurons can serve as representations.

The artificial neurons (units) discussed in the last section represent aspects of the world by means of numbers called activations that correspond roughly to the firing rates of real neurons. A typical neuron can fire as much as 100 times per second, and we can think of it as representing a degree of presence or absence of what it represents. For example if a unit represents the concept cat, then its firing many times per second signifies the presence of a cat. However, both natural and most artificial neural networks use distributed representations in which concepts are encoded by a population of neurons: a group of neurons represents a concept by virtue of a pattern of firing rates in all of the neurons. Thus a group of neurons, each with its own firing rate, can encode a large number of aspects of the world.

Focusing on firing rates, however, seriously underestimates the representational capacity of neurons and groups of neurons. The *spike train* of a neuron is its pattern of firing or not firing over a period of time. We can represent a spike train by a sequence of 1s (firing) and 0s (not firing). The spike train 10100 and 00011 both involve a neuron with a firing rate of 2 times out of 5, but they are different patterns. There are far more different spike trains than firing rates: see the Notes at the end of this chapter. Thus a group of neurons with varying spike trains have the capacity to encode an enormously large number of features of the world. See Maass and Bishop (1999) for analysis of the representational and computational capacities of spiking neurons, and Eliasmith and Anderson (2003) for an elegant analysis of neural representation.

We have seen that a single neuron can represent a feature of the world as the result of being tuned to fire more rapidly when that feature is presented. More powerful neural representations arise if the neuron can encode more possibilities by using the temporal properties of different spike trains, and if the neuron is part of a population of neurons that work together to represent many features. In sum, a representation in the brain is a population of neurons whose firing patterns encode information by virtue of having acquired regular responses to particular kinds of input.

Brain Maps

The brain does not try to use all of its billions of neurons to represent everything. Different brain regions represent different kinds of sensory stimuli; for example, the visual cortex at the back of the brain has neurons that respond to different visual inputs. There are neuronal groups whose firing patterns correspond spatially to the structure of the input, for example when a column of neurons fire together to represent the fact that a line is part of the visual stimulus. Thus different parts of the brain have groups of neurons that fire when different kinds of visual, olfactory (smell), taste, auditory, and tactile stimuli are presented. The human brain can do a lot more than represent stimuli presented to it, because groups of neurons can respond to inputs from many groups of neurons, producing a combined representation of what the input neurons represent. For example, there are regions in the frontal cortex of monkeys where the sensory modalities of taste, vision, and smell converge, enabling the representation of fruits and their key properties (Rolls, 1999). It is clear, therefore, that the brain is a superb representational device.

HOW BRAINS COMPUTE

Transformations

But is the brain a computer? The most familiar kinds of computation involve rules for transforming symbols, for example calculating that 2+2=4 and inferring from *p* and *if p then q* to *q*. Such computations are transformations of representations. The brain can also be viewed as performing transformations of representations encoded by the firing patterns of neurons. In general, a physical system is a computational system "when its physical states can be seen as representing states of some other systems, where transitions between its states can be explained as operations on the representations" (Churchland and Sejnowski, 1992, p. 62). To put it in a slogan: No computation without representation. Digital computers and brains are two different kinds of computational system.

Consider, for example, the operations of your visual system. At the back of your eye is the *retina,* millions of cells that are sensitive to light. Retinal cells respond to light reflected from objects into your eye, and send signals through a series of layers of neurons in the visual cortex. Successive layers detect more and more complex aspects of the objects that originally sent light into the eye, as neurons in each layer abstract and transform the firing patterns of neurons in the preceding layer. Thus the visual cortex progressively constructs representations of lines, patterns in two dimensions, and finally three-dimensional colored objects.

The brain transforms neuronal representations into new ones by means of synaptic connections. As we saw in chapter 7, the firing of one neuron can excite or inhibit the firing of another neuron. Hence one group of neurons with its patterns of firing can alter that patterns of firing of another group of neurons to which it is connected by means of synapses between pairs of neurons. There can also be feedback connections from one group of neurons to another, enabling them to influence each other. The brain contains many such feedback influences. In general, computation in the brain consists of interactions between groups of neurons that produce transformations of firing patterns.

Integration

The brain's operation is much more complicated than simply taking sensory input and transforming it. In order to eat a banana, a monkey needs to combine visual, tactile, and other kinds of information about it, and then use this integrated information to guide actions such as ingesting it. Hence much of what the brain does involves operations in central brain areas that combine information from multiple other areas. At the level of individual neurons, we can describe computation in terms of the ideas about activation, excitation, and inhibition presented in chapter 7, but a full understanding of the computational accomplishments of brains requires attention to the higher level operations of transformation and integration just described. Figure 9.2 depicts some of the interconnections of the prefrontal cortex (the front of the front of the cortex) with many other brain areas. Chapter 10 describes a computational model of emotion that describes the brain as making emotional judgments by combining information from the frontal cortex (high level thought), the amygdala (bodily information) , and the hippocampus (memory).

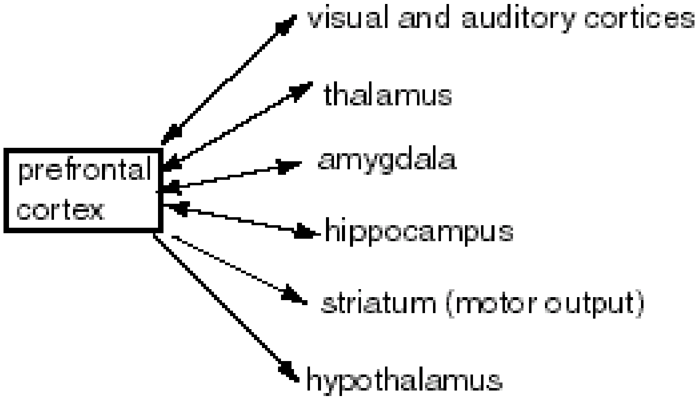

Figure 9.2 Inputs and outputs from prefrontal cortex. Based on Groenewegen and Uylings (2000). Note that connections usually go in both directions. Many connections between the other areas are not shown.

Learning

One of the most impressive computational accomplishments of neural networks is learning, in which changes in the synaptic weights between neurons produces major improvements in the performance of the network. However, networks trained by the backpropagation algorithm discussed in chapter 7 exhibit a problem called *catastrophic interference* (McCloskey and Cohen, 1989). This happens when a network is trained to perform one task, such as forming associations between words, and then trained to perform another similar task. When people undergo such retraining, they usually experience only some loss of the ability to perform the first task, but artificial neural networks can suffer a dramatic drop in performance on the first task when they learn the second.

McClelland, McNaughton, and O'Reilly (1995) argue that the brain's solutions to this problem is to have two complementary learning systems in two different brain regions, the hippocampus and the neocortex (the newer part of the cortex). The hippocampal system permits rapid learning of new items, whereas the neocortex learns slowly by a small adjustments of synaptic strengths through something like the backpropagation algorithm. As illustrated in figure 9.3, initial storage of most information takes place in the hippocampus, and is only gradually consolidated in the neocortical system. Catastrophic interference is avoided because new information coming into the hippocampus has only a small and gradual effect on the neocortex, which retains most of what it already new. Hence learning depends on two different interconnected systems operating with different learning procedures, so that understanding of neural computation involves specification of the operations of mutliple regions.

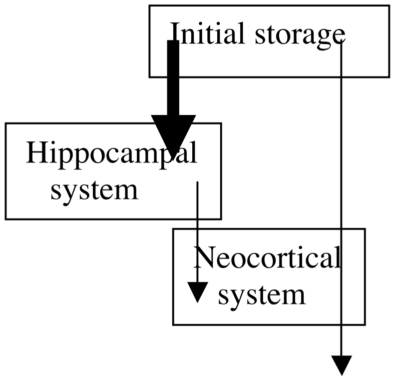

Figure 9.3. A two-system model of memory storage. Adapted from McClelland, McNaughton, and O'Reilly (1995), p. 444.

Thus it is not metaphorical to say that the brain represents and computes, even though its computations are done differently from the kinds most familiar to us in modern digital computers. Understanding of how brains work requires attention to the roles played by particular brain regions such as the hippocampus and prefrontal cortex. Hence we should not think of the brain as one big connectionist network of the sort described in chapter 7, but rather has a highly organized and interconnected system of specialized neural networks. This section has only begun to sketch the computational activities of brains; see Churchland and Sejnowski (1992) and Eliasmith and Anderson (2003) for much more detailed discussions.

HOW MOLECULES MATTER

If you have you ever had a cup of coffee or an alcoholic drink, then you have experienced the effects of chemicals on the brain. Coffee contains caffeine, which blocks the action of the molecule adenosine, which makes people drowsy by inhibiting firing of some neurons. Hence caffeine increases neuronal activity and keeps you awake. In contrast, alcohol can disrupt mental functioning by inhibiting the action of the molecule glutamate, which excites neurons. Both caffeine and alcohol also increase activity of the chemical dopamine, which produces feelings of pleasure. Adenosine, glutamate, and dopamine are all neurotransmitters, molecules that enable one neuron to influence another.

All the models of neurons and brains described earlier in this book are based on electrical activity: neurons fire and provide electrical inputs to other neurons. However, as caffeine and alcohol illustrate, the direct effects of real neurons on each other are chemical rather than electrical, in that molecules are emitted from one neuron and then passed over to another neuron, where they initiate chemical reactions that generate the electrical activity of the stimulated neuron. Figure 9.4 depicts how neurotransmitters are passed from the axon of one neuron to the dendrite of another neuron.

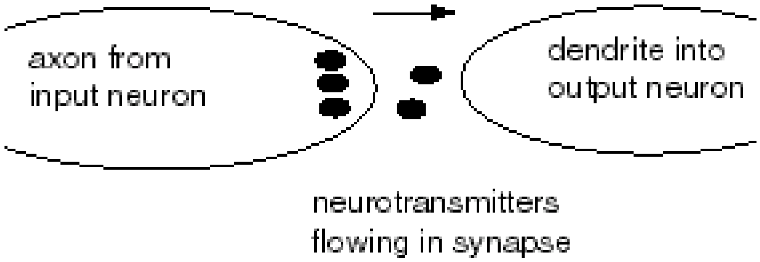

Figure 9.3. Neurotransmitter molecules flowing from one neuron into another.

There are dozens of neurotransmitter operating in the human brain, some with excitatory and others with inhibitory effects. This operation is consistent with the general connectionist ideas described in chapter 7, which assumed excitatory and inhibitory links between neurons. But broader chemical effects on neural computation are produced by hormones such as estrogen and testosterone, which can affect the firing of neurons independent of direct connections. A neuron in one part of the brain such as the hypothalamus may fire and release a hormone that travels to a part of the body such as the adrenal glands, which stimulates the release of other hormones that then travel back to the brain and influence the firing of different neurons. Complex feedback loops can result, involving interactions between the neurotransmitter control of hormone release and the hormonal regulation of neurotransmitter release. These feedback loops can also involve the immune system, because brain cells also have receptors for cytokines, which are protein messengers produced by immune system cells such as macrophages. Thus attention to the neurochemistry of hormones shows an important limitation to connectionist models, in that whether a neuron fires is not just a function of neurons that have synaptic inputs to it (see Thagard, 2002, for further discussion).

PRACTICAL APPLICABILITY

Motivations for studying brains are both theoretical – how does it work? – and practical – how can we help it to work better? Many advances have been made in understanding how mental problems can arise from defects in the functioning of different brain areas and neurotransmitters. For example, children with attention deficit disorder have difficulty concentrating and often fall behind in school. They are generally treated with Ritalin, which stimulates areas of the brain involved in filtering information by increasing the activity of neurotransmitters such as dopamine.

The growth of different brain areas can help to explain variations in behavior as people age. In the past decade, brain scans of children and teenagers have revealed that the brain undergoes remarkable growth at roughly the ages of 1-2 and 11-12 years. The latter growth spurt was a big surprise to researchers, and is generating new explanations of why many teenagers display difficult and risky behavior (Strauch, 2003). Areas of the prefrontal cortex are not fully developed until the early twenties, so that teenagers' decisions are often heavily driven by emotional information in the amygdala, rather than by reasoning about potential risks. They may, for example, take drugs such as ecstasy and cocaine, which provide short-term pleasure by intensely stimulating production of dopamine, but lead to addiction as the result of depletion of dopamine receptors that produces cravings for higher and higher doses.

Neurotransmitters and brain areas are also relevant to explaining many mental illnesses. Schizophrenia, in which people lose touch with reality because of bizarre beliefs and hallucinations, is associated with excess dopamine activity in the prefrontal cortex. Drugs that block dopamine reduce the symptoms of schizophrenia. On the other hand, lack of dopamine can produce problems with motor control found in Parkinson's disease. Depression is often treated with Prozac and other drugs that increase the availability of the excitatory neurotransmitter serotonin by decreasing its reuptake in synapses.

SUMMARY

The early decades of cognitive science, and even the connectionist models of the 1980s, largely ignored how brains produce thinking, But since the 1990s the brain challenge has been increasingly answered by experimental and computational investigations of how brains work. Brain scanning techniques such as PET and fMRI have provided a huge amount of information concerning how different brain regions contribute to various cognitive functions. Computational models of the brain have become biologically richer, both with respect to employing more realistic neurons such as ones that spike, and with respect to simulating the interactions between different areas of the brain such as the hippocampus and the cortex. These models are not strictly an alternative to computational accounts in terms of logic, concepts, rules, images, and connections, but should mesh with them and show how mental functioning can be performed at the neural level. The remarks on neurological plausibility in chapters 2-6 show that such meshing is rapidly progressing.

Moreover, there is increasing understanding of the chemical functioning of brains, in particular how different neurotransmitters and hormones affect neuronal firing. These advances do not require abandonment of the general view of thinking as representation and computation, but they do show the need to expand and supplement earlier cognitive theories.

Understanding of brain mechanisms is invaluable for explaining and treating mental illness. The explanation schema for kinds of mental illness such as schizophrenia is something like this:

Explanation target

Why do people have a particular kind of mental illness?

Explanatory pattern

People usually have normal brain function that involve identifiable brain structures and neurotransmitters.

Defects to these brain structures and chemical processes can disrupt normal functioning.

These defects produce the symptoms of the mental illness.

Many mental illnesses are treated by means of drugs that help to restore the normal functioning of chemical/electrical processes in the brain. The dramatic advances in neuroscience of recent decades do not by themselves solve the mind-body problem, since a skeptic can always maintain that there are aspects of mind such as consciousness that will never succumb to scientific explanation. But we will see in the next two chapters that even questions about the nature of emotions and consciousness are increasingly yielding to neurological investigation.

DISCUSSION QUESTIONS

1. How do different methods reveal different aspects of brain structure and function?
2. What is the relation between brain structures and the kinds of representation discussed in chapters 2-7?
3. Is brain processing really computation?
4. To understand brains, is it necessary to move down to the molecular and chemical level?
5. What aspects of thinking seem to you hardest to explain in terms of brain structures and processes?

FURTHER READING

See Finger (1994) for a readable history of neuroscience. Kandel, Schwartz, and Jessell (2000) is a standard neuroscience textbook. Allman (1999) discusses brain evolution. O'Reilly and Munakata (2000) describe computational models for cognitive neuroscience. Churchland (2002) and Bechtel et al. (2001) provide philosophical discussions of neuroscience.

WEB SITES

Brain and Mind magazine:
    http://www.epub.org.br/cm/
Explore the brain and spinal cord:
    http://faculty.washington.edu/chudler/introb.html
Science Daily brain and mind news:
    http://www.sciencedaily.com/news/mind_summaries.php

The Whole Brain Atlas:
http://www.med.harvard.edu/AANLIB/home.html

NOTES

If a neuron fires 100 times per second, then there are 100 possible firing rates it can have. The average firing rate of the neuron might be, for example, 25 or 50 times per second. A group of 1000 neurons can then represent a huge number of possibilities, 100^{1000}. For spiking neurons, there are 2^{100} different possible spike trains compared to only 100 different firing rates. The number of different possible combinations of spike trains in a group of 1000 neurons is therefore astronomical: $(2^{100})^{1000}$.

Chapter 3 mentioned the debate between the view that many rules and concepts are innate and the alternative view that emphasizes learning. A related debate concerns the extent to which brain structures have been selected by biological evolution for specific functions. Evolutionary psychologists claim that the brain contains a large number of evolved computational devices that are specialized in function, such as a face recognition system, a language acquisition device, navigation specializations, and a routine for detecting cheaters in social situations (Cosmides and Tooby, 1999). In contrast, Quartz and Sejnowski (2002) argue that the brain has evolved to make possible flexible learning: the main representational features of cortex are built from the dynamic interaction between neural growth mechanisms and environmentally derived neural activity.