

2

Neural Networks: Not Just Clever Computers

Many years before the “Chunnel” was built across the English Channel between Britain and France, there was a joke about a businessman who had won the low bid for building that tunnel. The businessman’s friends congratulated him for winning the bid and asked what his building plan was going to be. He replied, “Simple! We’ll start one tunnel from England, start another from France, and hope they meet in the middle*.”¹

If the biology of the brain is the “England” of that joke, and the psychology of behavior, emotion, and cognition is the “France,” then the growing interdisciplinary field of neural networks is one of the tunnels between them. The businessman’s wry description is a pretty good metaphor for the creative process of building model networks. We who construct these models sometimes work “top down” from observed human or animal behavior. At other times we work “bottom up” from the physiology of *neurons* (nerve cells) comprising the brain. Like the imaginative tunnel builder, we start at one or another end (either with the psychological results or with what we know about physiology and anatomy) and then refine our model to make it fit better with the other end. In the process we go back and forth until we have reached a certain stage of understanding.

Running a model on a computer involves precise mathematical calculation, but *constructing* models is much less precise: it’s more an art than a science.² So the criteria for deciding when a model is an adequate one for the data we wish to understand are also imprecise ones: they are intuitive rather than logical. Typical models are published in the form of a network diagram. In technical articles, unlike this

* The real Chunnel was in fact built from both the English and French ends in the early 1990s. By that time, however, laser technology enabled the meeting between the two tunnels to be much more precise than the joke implied.

COMMON SENSE AND COMMON NONSENSE

book, such diagrams may be accompanied by a set of equations and/or some computer code that describe the network's operations. The goal is to achieve with computer simulations some results that can be interpreted as analogous to some set of behavioral or neural or psychological data. One example of such data is *Pavlovian conditioning* (the type of learning that occurs when dogs are trained to salivate at the sound of the bell by repeated presentation of the bell followed by meat powder). Another is a card sorting test used by clinical neuropsychologists to detect frontal lobe damage.

What do we mean, though, by simulating the learning of, say, a bell-to-food association using mathematical equations solved on a computer? It may sound to some readers like comparing two things that are entirely different — “comparing oranges with fork lifts.”³ But mathematics has been used for centuries as a descriptive language for everything in nature. Most people are by now used to using mathematics to describe physical variables — numbers of atoms, positions of objects, electrical charges, and so forth. Psychological variables, such as strength of a hunger drive or memory of a bell, don't seem terribly “mathematical.” But they are part of nature, so we expect that scientists will eventually find precisely describable biological effects that approximate these drives or memories, even if we can never fit them exactly. In the meantime, we can use our networks and equations not as exact fits to data, but as metaphors⁴ for psychological effects that are partly measurable.

At the current stage of knowledge, neural network models are crude approximations of brain parts. So if a modeler succeeds in simulating on a computer some set of behavioral data, there will usually be other behavioral data that the current model can't quite reproduce. Then, in future work, she or he or other colleagues try to extend and refine the model so it can also account for the other sets of data. By this process, the models gradually become more complex and, hopefully, more accurate representations of how the brain really does things.

There is no universally recognized definition for what exactly a “neural network” is. The most widely accepted definition is probably the one developed in 1988 by a team of neural network experts that were part of a study in the United States commissioned by the Defense Advanced Research Projects Agency (DARPA).⁵

NOT JUST CLEVER COMPUTERS

a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes. ... Neural network architectures are inspired by the architecture of biological nervous systems, which use many simple processing elements operating in parallel to obtain high computation rates.

What is meant by the *nodes* or *elements* in this definition? In the work of Warren McCulloch, one of the field's pioneers in the 1940s, nodes were conceived to be analogous to single neurons (brain cells). But as the field developed, scientists more often conceptualized the nodes as large groups of neurons or as regions of the brain. This change occurred for several reasons. First, the number of neurons in the human brain is extremely large, of the order of a trillion, so a neuron-by-neuron "wiring diagram" would be impractical. Second, experiments from neurophysiology laboratories have suggested that the electrical patterns of single neurons and the biochemistry of the connections (called *synapses*) between neurons aren't very regular in their organization. But if some of the irregularities at the levels of single neurons and synapses are averaged out across large groups or brain regions, regular connection patterns emerge that are important for behavior and for mental functions.

Sometimes, a node in a *model* neural network represents not a known brain area but the brain's encoding of a particular concept — for example, "the letter A," "the hunger drive," "the rule that classifies cards by color." In one network nodes represent stimuli used in a conditioning experiment. Some biological purists object to that, since using current knowledge they can't localize such a node to electrical patterns in a particular region of the brain. But this is just the "tunnel" being built from the psychological side of the problem. If you want to construct a computer model of a complex behavior, such as classifying cards, you first need to break it down into simpler behaviors, such as perceiving features like color or shape of the designs on the cards. So some of these neural network nodes may represent the "sub-functions" necessary to understand a more complex, larger mental function.

COMMON SENSE AND COMMON NONSENSE

Now what does a typical piece of “tunnel” built from the neurobiological side look like? The mathematical relationships within a set of neural network nodes and connections are designed to be somewhat analogous to those in the real nervous system— regardless of whether the nodes are interpreted as brain regions or as representations of concepts. The process by which neurons transmit signals to other neurons is described in many textbooks,⁶ and only a sketchy description is given here. Each neuron consists of three major parts (see Figure 2.1). These are the *cell body*, containing the nucleus which all biological cells, including neurons, possess; the *axon*, a long thin filament projecting out from the cell body; and the *dendrites*, a large number (in the thousands for each neuron) of smaller branches going into the cell body. The typical neuron receives electrical signals from other neurons at the dendrites, then processes these signals at the cell body. If their combined strength is large enough, it is translated into another signal that travels down the axon to one of the synapses it makes with other neurons. At the synapse, the mechanism changes and processes involving chemical transmitters take over.

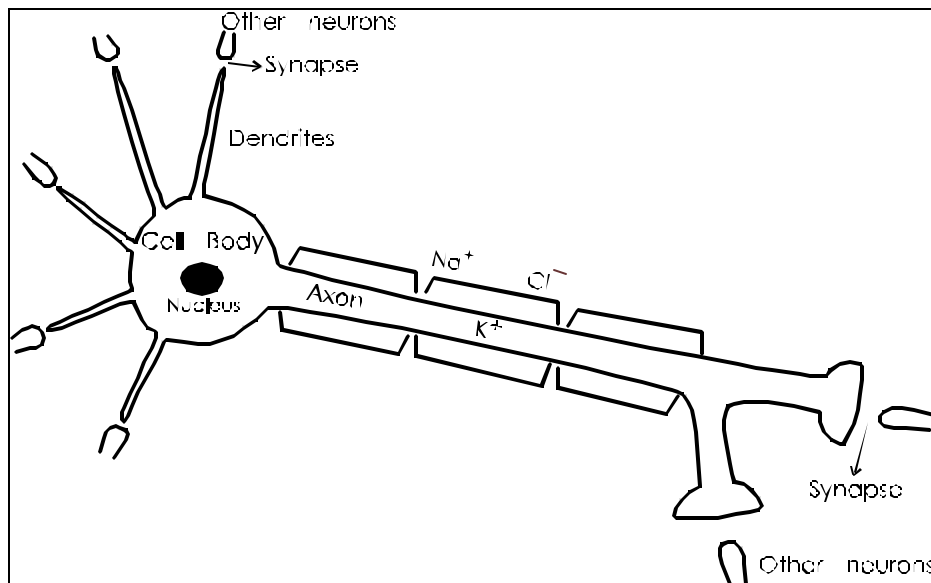


Figure 2.1. Schematic neuron. Main parts (axon, cell body, dendrites, synapses) are labeled. Na^+ (sodium), Cl^- (chloride), and K^+ (potassium) are the ions (electrically charged atoms) that play roles in electrical impulses generation and transmission. (Reprinted with permission from Levine, Daniel S., *Introduction to Neural and Cognitive Modeling*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.)

But what is the electrical signal? It consists of a temporary change in the voltage across a membrane which covers the cell, all along the dendrites, cell body, and axon. This is done by means of movement across the membrane of some electrically charged atoms or *ions* — sodium, potassium, and chloride ions. By processes that aren't yet fully understood, release of a chemical transmitter substance from synapses leading to a neuron can either raise or lower the probability of this exchange of electrically charged atoms across the membrane. The changed voltage, by other processes that are also not completely understood, in turn causes the release of a certain amount of chemical transmitter from some synapses going from that neuron to other neurons. The process of voltage change, if it is sufficient to transmit a signal, is called an *action potential*.^{*} It is also sometimes referred to as the neuron *firing* or, because of the characteristic shape of the curve representing the voltage over time (see Figure 2.2), *spiking*.

In most computer models of artificial neural networks that perform behaviors, the nodes or units are interpreted as groups of a large number of neurons, maybe several thousand. At this stage of development of the models, there isn't enough precise knowledge to assign each of these nodes to a specific, measurable brain area (even though, at times, a rough general location in the brain is indicated). For this reason, the details of electrical signals and electrically charged atoms don't appear in the equations for typical computer networks. Instead, the electrical signals are averaged out into variables called the *activity* of each node. The most common biological interpretation of "activity" is of the current average frequency of action potentials for a group of neurons in some area over some window of time. Some readers will no doubt be uncomfortable with using an abstract notion such as activity which can't be fully defined yet in real-world terms. But scientists in many fields, for most of this century, have used abstract constructs like this to gain understanding of the real physical world by studying a simplified, idealized version of the world.

^{*} The word "potential" here is used in the sense of electrical potential, which is a synonym for voltage.

COMMON SENSE AND COMMON NONSENSE

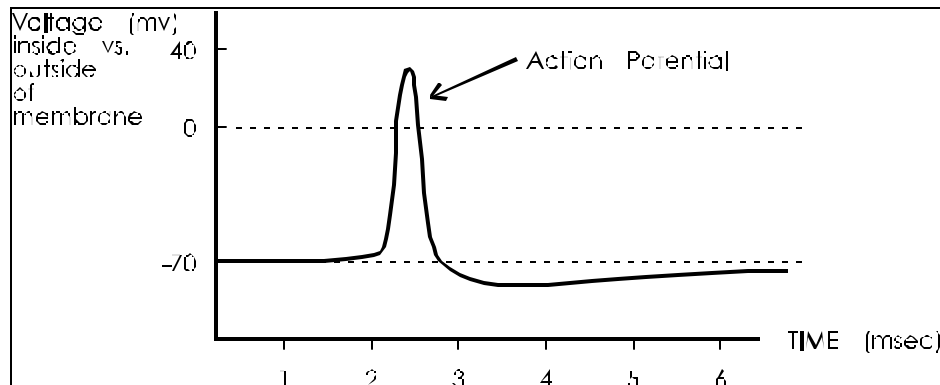


Figure 2.2. The action potential recorded across the membrane of the giant axon of a squid. (Reprinted with permission from Thompson, Richard, *Foundations of Physiological Psychology*, New York: Harper and Row, 1967.)

Connections between neurons can either be *excitatory* (tending to increase the probability of an action potential) or *inhibitory* (tending to decrease the probability of an action potential). Likewise, a connection between nodes in a neural network is excitatory if a signal produced by activity in one node tends to increase activity in the other node. The connection is inhibitory if a signal produced by activity in one node tends to decrease activity in the other node. Both excitation and inhibition perform cognitive functions in artificial as well as biological neural networks. Excitation is important for creating associations between concepts (e.g., for Pavlov's dogs, hearing the bell "excites" the thought of food). It also plays a key role in causing either an emotional drive, or a reasoned plan, to stimulate action. Inhibition is important for making decisions between alternatives, because a person or animal needs to do one thing and *not* do another thing. It can cause us, for example, to engage in one behavior and not engage in a competing behavior. Likewise, inhibition can make us attend to one part of a perceived stimulus but not attend to another part. Inhibition is also important for controlling the intensity of brain activity, that is, for keeping excitatory signals from overwhelming the network with epileptic-like discharges.

There is a great deal of variety in the mathematical rules for neural networks. Most of them involve changes over time in the activities of interacting nodes and, often, in the strengths of connections (sometimes

called *connection weights*) between nodes. Connection weights are idealized biochemical variables, just as node activities are idealized electrical variables. They are thought to represent amounts of chemical transmitter substances or properties of certain molecules on the cell membranes that are *receptors* for these transmitters — that is, they bind with the transmitters and so cause the receiving neuron’s electrical properties to change. By now there is considerable neurophysiological evidence that strengths of many synapses between pairs of neurons change when both neurons are repeatedly electrically active at the same time. Psychologists interested in learning (including Sigmund Freud) suggested the idea of changes at synapses long before it was actually observed by neurophysiologists. This kind of flexibility of connection strengths seems to be required not only for learning but for the brain’s overall function of mediating between the rest of the body and the outside environment.

For the purposes of this book, the technical details of how neural networks are simulated computationally and what they correspond to in biological neurons are mostly unnecessary. What is important is that they represent the dynamics of interactions between nodes that are identified either with regions of the brain or types of neurons within given brain regions; with representations* of particular mental objects such as percepts, actions, memories, emotions, plans, or concepts; or with both simultaneously.

Like all mathematical models of real-world events, neural network models of the brain can be thought of as *caricatures* of what they model. That is, a model doesn’t represent *everything* about the system it’s reproducing, only those features needed to understand particular behaviors of the system. But just as the caricatures in cartoons can bring alive some tendencies of the characters they portray, the caricatures in neural network models can yield some valuable intuitions for what types of brain structures are likely to produce certain behaviors. The two “take-home messages” about neural networks are:

* There is now a great deal of academic debate among philosophers, linguists, and modelers as to what exactly a “representation” of a mental object consists of. Since this debate isn’t essential to the main points of this book, it won’t be discussed further here, but the interested reader can consult Stephen P. Stich and Ted A. Warfield, *Mental Representation: A Reader*, Oxford, UK: Blackwell, 1994.

COMMON SENSE AND COMMON NONSENSE

(1) Neural networks are a metaphor for the fact that all of mental life is dynamically interrelated. Perceptions, categorizations, beliefs, emotions, plans, and actions can't be separated from each other, but instead form what some religions call "an interconnected web." Later chapters will discuss results from experimental psychology showing, for example, that cognitive ambiguity can lead to emotional discomfort, and that emotional biases can influence how categories are chosen.

(2) Some specific neural network architectures can function as useful metaphors for specific human attitude tendencies. Later, for example, I will introduce a neural network that serves as a metaphor for the human tendency to get stuck in entrenched, unrewarding behaviors. Another network serves a metaphor for jumping between polar emotional opposites (such as love and hate). But we will also discuss a neural network representing the creative process that encourages self-actualization!

For the reader who wants a more detailed grounding in how a neural network is organized, some examples are shown in the appendix. The simplest example that I start with there is a model of Pavlovian conditioning. Other examples of networks represent other types of psychological processes such as memory, learning, sensory perception, reward and punishment, motor control, decision making, and categorization.

Current Uses of Neural Networks in Neuroscience and Psychology

Neural networks are designed to simulate various sorts of mental functions, wherever those functions appear. As a result, in addition to their biological and psychological applications, neural networks have found wide usage in industrial and engineering applications that call for some form of "intelligent" functioning. These include, for example, visual pattern processing, speech signal processing, robotics,

NOT JUST CLEVER COMPUTERS

manufacturing analysis, financial forecasting, and many other applications⁷. This has been a growth area for high technology since the mid-1980s.

The applications of neural networks to understanding actual human mental processes, which are much more emphasized in this book, have lagged behind the industrial applications. But now, in the late 1990s, is a time of rapid growth for these biological and behavioral models. This growth is happening in part because more biologists and psychologists than ever before have access to high-speed personal computers. It is also spurred by many advances in experimental neuroscience, such as positron emission tomography (PET) scanning, which are enabling a more complete (though not yet perfect) account of the actual metabolism of brain tissue during the performance of cognitive tasks. The assumption is that those areas that are most active, in terms of metabolism and blood flow, are the parts of the brain being most used in the current task. All these advances are making it seem possible for brain science to be given a solid theoretical framework. For this reason, many psychology or neuroscience laboratories are hiring or collaborating with researchers whose expertise is primarily theoretical rather than, or as well as, experimental. These researchers often combine knowledge of the relevant neuroscientific or psychological literature with training in computer science, engineering, mathematics, or physics, all of which help them understand system dynamics and build quantitative models.

Let me mention a few scientific events from the 1990s as signs of the times. In 1995, two conferences took place — one in London and one in College Park, Maryland — on neural network models of mental and cognitive disorders, such as schizophrenia, epilepsy, Alzheimer's disease, depression, stroke, and aphasia. These were the first major conferences devoted to that recently developed subfield. The Maryland conference, which has been made into a book,⁸ was physically crowded because it was planned as a tiny workshop but drew over a hundred attendees, including many practicing psychiatrists and neurologists. There have also been symposia on neural networks at conferences such as annual meetings of the American Psychological Society and American Psychological Association. There are also at least four ongoing annual international conferences devoted to neural networks themselves; some of these meetings include both sessions devoted to neuroscience and other sessions devoted to industrial

COMMON SENSE AND COMMON NONSENSE

applications. In addition, one of these conferences had between 1994 and 1996 a session devoted to studying and modeling consciousness.

Government support of neural network research has begun to be integrated with support for neuroscience. For instance, there has been some funding by several European governments (including Great Britain, France, Belgium, Germany, Switzerland, Sweden, and Finland) for parts of an effort to map out the structure of the whole human brain, known as Project PSYCHE. This project includes neural network modelers along with neurophysiologists, electroencephalographers, clinical PET scanners, and others. The National Institutes of Health in the United States has been funding a looser program with similar aims, called the Human Brain Project, and there are others in Japan.

The best known early successes of neural network modeling were in the field of perceiving, classifying, and categorizing sensory patterns. This gave insights into how the brain's perceptual systems, particularly its visual system, work. These pattern classification networks also have a wide variety of applications to the performance of "intelligent" functions by computers. Among these are medical diagnosis, wherein the visual display of a particular organ for someone with a disease is different from the same display without the disease. It has also been used to classify radar signals as coming from different emitters⁹ or handprinted numerals in zip codes as being specific digits.¹⁰ Neural network modeling of the psychological process of conditioning, has also been an active field of research since the 1970s.

More recently, this same methodology has come closer to understanding the most complex human cognitive processes and their characteristic breakdowns with brain damage or mental illness. There have been models of the effects of frontal lobe damage on the ordered planning of behaviors.¹¹ There have also been preliminary models, for example, of Alzheimer's disease,¹² one type of dyslexia,¹³ and a qualitative neural network theory of manic-depressive illness.¹⁴ Also, there has been at least one computational model of the disruption of motor behavior by Parkinson's disease.¹⁵

The Microstructure of Cognition

Neural network models of complex cognitive processes are often developed by breaking these processes into simpler subprocesses for the purposes of analysis. For example, neural networks have been used to model of categorization of sensory patterns (see the appendix for more details). Specifically, handwritten characters such as might appear on a postal envelope, and at times are written imprecisely, are categorized as to what letter of the alphabet they are closest to. In order to model that categorization process, we need models of at least two subprocesses. One of these subprocesses is learning, because the representation of Roman letters in the brain isn't hard-wired: the same neural structures are equally capable, for example, of learning Japanese, Hindi, Hebrew, or Russian letters. Another subprocess is deciding between two alternatives. For example, a sloppily written letter might look somewhere in between an "E" and an "F," so we need to be able to decide which letter is more likely, enhance our mental image of that letter, and suppress our image of the other letter.

The neural network modelers David Rumelhart and James McClelland, who are also psychologists, called this type of analysis an *exploration of the microstructure of cognition*.¹⁶ Another group of modelers, headed by Stephen Grossberg, has applied this kind of analysis to a range of cognitive and behavioral processes including categorization, conditioning, visual perception, word recognition, and speech encoding.¹⁷ The computational theories of Grossberg and his colleagues suggest that similar types of subprocesses are components in all these different things that our brains do. For example, the same principles of associative learning and perceptual decision are used both to model the process of categorization and also, in a different form, to model of the role of selective attention in conditioning.¹⁸

What these various neural modelers have done is develop a "tool kit" consisting of different parts, or subblocks, of neural networks, that can be used repeatedly and in different combinations. Is this, as I believe, roughly the way our brains are really constructed? That would mean that just like a few base substances account for all of our rich genetic code, a few types of characteristic neural connections repeat in many if not all parts of the nervous system, from the spinal cord and midbrain reflex centers up to the

COMMON SENSE AND COMMON NONSENSE

frontal lobes and other association areas of the cortex. A book in press by the neural network modeler John Taylor develops a series of interrelated neural network theories for many different areas of the brain as they relate to high-level processes.¹⁹ These include the organization of planned sequences of behavior — based on a combination of rational analysis and emotional preferences²⁰ — and the way memory is involved in consciousness.

If Rumelhart, McClelland, Grossberg, Taylor, and their colleagues are on the right track, their methodology can be fairly universal when it comes to mental processes. They hint that the same kinds of neural structures that handle relatively simple processes like visual perception and conditioning can also, *in different combinations* and *with greater complexity*, handle much more complex processes like reason, inference, and decision making. Ultimately, sufficient understanding of mental processes can even lead to theories of self-actualization, ethical behavior, and how we decide right from wrong.

This emboldens me to apply this kind of “microstructural” analysis to the behavioral processes this book deals with. Neural networks give us a framework in which to think more systematically about self-actualization and its absence, the “imp of the perverse,” the gap between intention and action, and the deviations of much human behavior from optimal.

I am sometimes asked what a neural network, or computational, approach can add to our understanding of human psychology over and above what can be gained by just thinking intelligently about mental processes. Many people seem to think that approaching problems via a new discipline automatically leads to a different viewpoint. One student who read a draft of this book, in fact, expressed an apprehension that I was trying to quantify love. My answer was, “No, I am trying to love-ify quanta!”

Neural networks don’t yield a dramatically different view of the brain and behavior than would be had without them. They merely help us tackle problems of human behavior using what has come to be called a *systems* approach.²¹ This means that each of our personalities, like any other complex system, is seen as a web of different subsystems (in this case emotion, cognition, reason, memory, perception, motor action, et cetera), all influencing each other dynamically but each somewhat autonomous. In such a web, it is common nonsense to say that some parts of our personality structures are “better” or more

NOT JUST CLEVER COMPUTERS

“basic” than other parts. Our neural networks are studied through the mathematical theory of dynamical systems (also sometimes known as chaos theory), which applies to a wide range of physical and social systems.

There are no gimmicks here, and the reader may ask “So what’s new?” But in later chapters when we discuss personal and social change, the dynamic systems approach will turn out to have a lot of surprising implications for human relations. This does *not* mean, as some fear, that people will need degrees in computer science or mathematics to understand what this book says! Some conclusions in this book may be difficult for some people to integrate because they will challenge some of our comfortable social norms. But these conclusions will be stated in everyday and not technical language as much as possible, and so will suggest standards for human behavior that are within reach for almost everyone. So for people willing to break new ground, scientific approaches will provide hope. Let us now return to the gap between human potential and human reality.

Chapter 2: Neural Networks: Not Just Clever Computers

1 Ausubel, 1948.

2 There are some textbooks on the process of mathematical modeling, which is applicable to all the natural and social sciences. One of the best I have seen is Bender, 1991.

3 Voss, 1995, 42.

4 Voss, 1995.

5 DARPA Neural Network Study (1988). Alexandria, VA: AFCEA International Press, p. 60.

6 Kandel and Schwartz, 1985; Shepherd, 1983.

7 Miller, Walker, and Ryan, 1989; Miller, Sutton, and Werbos, 1990; Arbib, 1995.

8 Reggia, Ruppin, and Berndt, 1996; Parks, Levine, and Long, 1998; Stein, 1998.

9 Anderson, Penz, Gately, and Collins, 1988; Levine and Penz, 1990.

10 Kamangar and Cykana, 1988.

11 Bapi and Levine, 1994; Cohen, Dunbar, and McClelland, 1990; Cohen and Servan-Schreiber, 1992; Dehaene and Changeux, 1989, 1991; Levine and Prueitt, 1989.

12 Hasselmo, 1994; Parks and Levine, 1998..

13 Plaut and Shallice, 1995.

14 Hestenes, 1992.

15 Contreras-Vidal and Stelmach, 1995.

16 Rumelhart and McClelland, 1986.

17 Grossberg, 1988.

18 Grossberg and Levine, 1987.

19 Taylor, in press.

REFERENCES

20 Bapi, Bugmann, Levine, and Taylor, submitted.

21 von Bertalanffy, 1968.