Chapter 1

Cognitive Science Is Not Cognitive Psychology

> Information-processing theories of the cognitive mind/brain can explain certain features of cognition that cannot be explained by means of lower level neuroscientific accounts.[1]

> Suppose that a neuroscientist has . . . arrived at your campus for a lecture. You eagerly ask him, "Tell me, Professor X, when your subject Joe images a small pine tree at an angle in the center of his imagistic 'field', what is going on at the neural level that explains the intentionality of Joe's imaging?" Professor X consults his table of psychoneural correlations and replies, in deep, serious tones: "I'm so glad you asked that. The explanation is simple. When Joe exercises his imaging capacity in that way his brain is moving form neural state N624 to neural state N1009." "Thank you, sir: that's very enlightening," you reply, as your mind draws a complete blank.[2]

This project opens with a series of biases and reactions. The most immediate and obvious is that the brain is quite important in understanding cognition. Indeed, I would daresay that the brain sciences are an *essential* component in a mature cognitive science. For those not well acquainted with the recent history of debates among some philosophers and psychologists over the nature of cognition, this may seem a pretty uninteresting and weak claim. How else do we think but with brains? On the other hand, for those intimately acquainted with recent dialogues, dialectics, and diatribes, my "bias" might appear almost heretical: one should not confuse the hardware with the software, as it were.

And certainly one should not embrace one for the other with full knowledge of forethought.

My second bias, which will turn out to be tied to my first, is that theories in cognitive science are strongly interdisciplinary. Cognitive science is not just cognitive psychology with a few additional bells and whistles. The corollary to this claim is that one should not conflate pedagogical issues in teaching cognitive science to neophytes (e.g., how best to present the material in an orderly manner) with the actual products of research by our scientific communities, who pay little mind to whether their discoveries and conclusion are palatable to those on the outside. I doubt that anyone would claim to find this bias even remotely controversial. Nevertheless, those who deny my first bias are quite often guilty (at least in their actions) of also denying the second. Allow me to explain.

Early writings in the philosophy of mind and the cognitive sciences stress two fundamental notions. First is that species of thought are species of information processing construed as formal symbol manipulations. Allen Newell and Herbert Simon are credited with the insight that the mind can be formally construed as a symbol-manipulating system.[3] They argue that we can describe human behavior by "a well-specified program, defined in terms of elementary information processing." (See table 1.1 for a general description of this sort of information-processing system.) As an instance of a Turing Machine, this general system has only a few primitive capacities but is an extraordinarily powerful machine. It can store (and recall) symbols and create new symbols to stand for those symbol structures. It can compare tokens of symbols and then classify them as the same or different. However, the most important capacity of this machine, and the capacity that gives it its true power, is its ability to act contingently—it can execute different information processes, depending on which symbol structure it encounters. In this way, the machine can respond to context. Using structures built out of symbols, it can designate environmental events and its own processes, on which it then bases its responses. This information-processing system became the dominant model for the human mind in psychology and elsewhere after Neisser's classic (1967) work, *Cognitive Psychology*, which incorporated many of these ideas.

The second defining idea is that formally described systems are multiply instantiable. As far as I know, Putnam (1960) first remarked on this fact in the context of studying the mind. He argued that the formal definitions of processes in the cognitive sciences describe events that we

Table 1.1 Postulates for a General Information Processing System

(1) A set of symbolic tokens connected by a set of relations makes up a "symbol structure," which is stored and retained in some memory register or other.

(2) If a symbol's designation is fixed by an elementary information process or by the external environment, then it is "primitive."

(3) These symbol structures then act as either inputs or outputs (or both) for an "information process."

(4) The symbol structures refer to objects if information processes that take the symbol structure as output, either:

A. affect the object itself, or
B. produce symbol structures that depend upon the object.

(5) An information processor has three basic components:

A. a fixed set of elementary information processes;
B. a short-term memory register that holds the input or output symbol structures of the elementary information processes;
C. an interpreter which, as a function of the symbol structures in short-term memory, determines the sequence of the elementary information processes to be executed by the information processing system.

(6) The symbol structure is a "program"

A. if the object a symbol structure designates is an information process and,
B. if given the proper input, an interpreter can execute the process.

can conceive *solely* in terms of input, output, and various functional or causal relations.[4] Minds just are (or can just be conceived as) things that exhibit the appropriate cause relations. Consider a mousetrap for an analogous example. What counts as a mousetrap is something that takes in a live mouse and then turns out a dead one. Mousetraps can be built out of just about any substance. If we ask what unites all legitimate token mousetraps under that type, no answer can be given in terms of underlying physical substances. Mousetraps can be made out of wood, metal springs, poison, plastic, baseball bats, and so on. What unites these objects is not the stuff from which they are made, but the purposes they serve. They all deliver dead mice when given live ones. Hence, we understand mousetraps just in terms of inputs, outputs, and certain relations between them.

These sorts of formal or functional definitions in science contrast with understanding objects or properties in terms of some developmental history. For some things, how they got to be here is important for understanding what they are. These are things that cannot be understood purely formally. For example, the property of being a codependent is not formally defined. In brief, a codependent is someone who supports addictive personalities. But the causal relations that codependents enter into might not differ much from those who are not codependent, or might not be distinguishable from the relations of noncodependents, were it not for the codependent's particular history. Codependents typically engage in solicitous behavior, usually with respect to people in trouble whom they care about, and this behavior helps boost a codependent's self-worth. Insecure kind people act in the same way. What makes codependents codependent is not their behavior per se, but the history of their relationships that leads them to act in this way.

The purely formal approach to mentality tells us that we could conceivably find minds comprised of appropriately connected neurons, or silicon chips, or zinc, or some alien substance. Since these definitions say nothing about the ontological status of mental events, the same type of mental state could be realized in any number of underlying things, and they would all count as tokens of that type as long as they give the correct outputs for specified inputs and maintain appropriate causal connections with other mental states. In other words, the study of mind is analogous to the study of mathematics: neither requires an actual, physical instantiator to get the study underway.

There is a strong methodological advantage in adopting this perspective. If we can define mental states in terms of inputs, outputs, and causal relations, then we have an easy way (relatively speaking) to explore the implications of particular theories: computer simulation. If the underlying stuff does not impact the relations under study, then theoretically at least we could design computer programs to mimic the hypothesized causal relations to test whether our model gives us the outputs we expect for the inputs we choose.

Compare this project with computer simulations of the weather. The variables that figure into determining the weather for any particular day are enormously complex; it is exceedingly difficult for any human to track them all. But we can devise computer programs that calculate the values for all the variables and their interactions for each iteration of time. These programs map the weather with an extraordinary degree of precision, at least for a few days. Something similar should be available for human cognition. In real life there are simply too many contingencies for scientists to be able to test their simple theories. Even in the fairly constrained environment of a laboratory in which subjects are only given a few inputs at a time, interactions with things that scientists cannot control can easily swamp the effects that they are trying to test. But with computer simulations, we can control all the inputs a system receives. Though these systems would then only mimic aspects of cognition, they still present a powerful tool for testing hypotheses in addition to laboratory work.

Moreover, computers can track exceedingly complicated causally interactive algorithms in ways that we cannot. Computers give us one easy and relatively inexpensive path for exploring rough and ready new ideas about how our mental states interact before designing time-consuming and expensive laboratory experiments. Explaining mental states in terms of their causal interactions radically expands the possibilities we have for exploring and testing hypotheses and theories of mental cognition.

These two assumptions taken together, that the mind can be understood as a formal symbol manipulating device and that "mental programs" can be multiply instantiated, led many early on to conclude that the underlying instantiating stuff was simply unimportant.[5] Nowadays, however, even the staunchest "formalists" would agree that the brain sciences can be useful in understanding cognition. Not surprisingly, they concede that studying the brain may tell us something about the program that it is running.

In fact, the rise of neural net modeling[6] in conjunction with concrete results from neuroscience regarding how minds actually work demonstrated three important facts to cognitive science. (1) We can model brains purely computationally at a lower level of organization than originally envisioned. This opened the question of the appropriate level of analysis for mental systems, something that had previously been taken more or less for granted. This topic is pursued in chapters 2 and 3. (2) How we actually work differs quite often from the strategies various computers use in solving problems. These differences become important not only for therapeutic reasons, but also because cognitive science at bottom is a science about *us*. Computers comprise but one tool among many we use to divine the basic principles behind thought. This theme is echoed in chapters 6 and 7. (3) It is quite difficult and artificial to separate the study of one domain from another, since everything influences just about everything else in the brain. Hence, the ways in which we had originally carved our cognitive capacities up for study might be mistaken. This point is discussed in chapter 5.

The point of disagreement between myself and others is over how useful the brain sciences are and in what respects. I believe that these facts hugely alter the face of cognitive science. However, my reading of the current state of affairs is that many philosophers of mind have simply tacked neuroscience onto their list of what counts as a cognitive science without fundamentally reorganizing their notions of the field.[7] That is, their realization of the three points listed above, plus perhaps only a rudimentary knowledge of neuroscience, have led them merely to add an entry on computational neuroscience to their mental lexicon. This, I think, is a mistake.

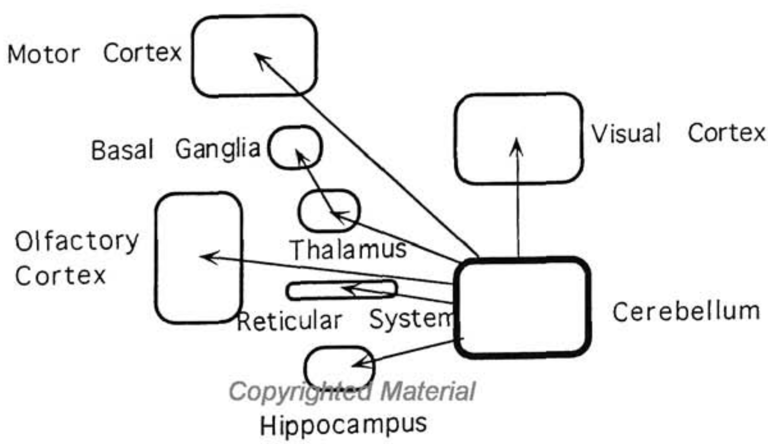
What they have done mimics the currently popular textbooks in cognitive science.[8] In these books, we find chapters devoted to computation, language, reasoning, attention, memory, artificial intelligence, and psychological development, as well as a chapter on neuroscience (usually on vision, and usually the last chapter). Pedagogically, it makes sense to break the study of the mind into these (or similar) categories, since historically each has been studied separately, and each has developed under different research paradigms. It is simply easier to learn piecemeal.

However, *thinking* of the theories of cognitive science or its subject matter in this compartmentalized manner is not prudent. For all of these topics are intimately related, and all (for us) are related to the brain. It is simply a mistake to reduce meaningful symbols to formal patterns (to reduce semantics to syntax) and then remove those patterns from their culture, history, and biology. We are living, interactive creatures. We are designed to move through this world, seeking food and mates and avoiding predators and other painful things. Any science that purports to study some aspect of us has to take these facts into consideration.

My point is that we cannot (or perhaps, should not) detach an organism's cognition from the animal's strategy for survival within particular ecological environments. Despite the claims of the early serial models, nervous systems are not general purpose computers; they evolved to accomplish a certain range of specific tasks and their architecture supports those tasks. We all know this, and yet, the early von Neumann models push us toward pretending that the brain essentially is in the symbol manipulation business. But if we consider the brain from a more biologically realistic perspective, we can see that nervous systems are designed to move organisms appropriately in their environment. As evolutionary biologists are fond of pointing out, nervous systems enable organisms to succeed in the four F's: feeding, fleeing, fighting, and reproducing.[9]

I can give no serious argument for why I think this sort of bias toward the neuroscientific study of the mind/brain is superior, but adopting this perspective does offer several advantages. For example, the general organization of the brain makes sense if we assume that the nervous system exists for creating motor output. The cerebellum, an area crucially involved with motor control,[10] is connected almost directly to all areas of the brain. (See figure 1.1.) It inputs into sensory transmissions at all levels. It is connected to our reticular activating system, the system that controls arousal and attention. It is also connected via the thalamus to the basal ganglia and forebrain. With these connections, the cerebellum could alter the frontal cortex-basal ganglia interactions. Finally, it is connected to the hippocampus, the area responsible for laying down episodic memories, and via its connections to the hippocampus, it inputs to, and receives output from, the limbic system. Not only can our "motor control" center modulate sensory transmission, arousal, attention, cortex-basal ganglia interactions, and episodic memory, it can also alter our emotionality and social responses![11] All areas of our brains seem geared almost exclusively to coping with their functions as they pertain to problems of motor control.

Figure 1.1 *Connections of the Cerebellum*



The brain puts much emphasis on the priority of motor tasks, and we should pay attention to this emphasis. If we do, then we reorient our essential understanding of what we are. Any abstract symbol manipulating we do becomes a matter of motor assembling. Because "thinking" would turn on the need to predict events in the extra-nervous world such that organisms can move successfully through that space,[12] we only process information insofar as doing so is relevant to some motor output. This, to me, is a very different picture of a cognizing organism than the one advocated by early cognitive psychologists.[13]

But I don't want to claim that cognitive science is simply cognitive neuroscience either. I do think that the brain is at the center of all of this, but theories in cognitive science are truly *interdisciplinary*[14]—they exhibit fundamental connections among previously disparate domains. To my mind, they exhibit connections between psychology and neuroscience in particular. Nevertheless, this is just my take on the current state of affairs. What is required to be an interdisciplinary theory in cognitive science is for it to span more than one traditional domain. Generally speaking, as I discuss in later chapters, this means that the theory will cover more than one level of analysis and organization in the mind/brain in its *explanans.*

To speak less from my heart: cognitive science is the study of information processing, and insofar as some discipline studies that, then it is part of cognitive science. (A computational approach, though dominant, is not required.) I shall try to do without explicating the subject matter in any detail; it changes as the science grows and develops, as we learn more about our cognitive abilities and their connections to the rest of our psyches and the world. I fear that any attempt to constrain or rein the topics would only make any analysis seem artificial and dated.[15] Having said that, let me now proceed to date this book. The disciplines currently involved to some degree or other in cognitive science include: anthropology, biology, computer science, engineering, linguistics, mathematics, philosophy, psychiatry, psychology, neuroscience, and sociology. And the list keep expanding as we realize that information processing is more complicated than artificial intelligence personnel originally thought and how many disciplines actually study this in some guise or other.

Contra Stillings et al. (1987), the convergence among these disciplines is not over the questions asked about mentality. (Indeed, as I argue below, the questions asked are very different in different domains for historical and sociological reasons.) Nor is it in a common research framework.[16] The frameworks differ as the domains differ. What they do share is an interest in explaining how we process information, period. But how they explain it and what counts as an information process is not uniform across the fields. My task is to explicate how one can take these diverse beginnings and still come out with coherent interdisciplinary theories. My answer will revolve around the common explanatory patterns one finds in cognitive science.[17] My secondary task is to convert (perhaps) a few fence sitters into believing that the brain has to be at the center of any serious cognitive science. That is, it will be to illustrate by way of example that most of the common explanatory patterns involve the brain in some way or another.

This is why I fashioned the book as a "how to" manual, for it is partly descriptive and partly normative. I describe how cognitive science is actually done, at the same time that I implicitly recommend how it should be done. Ultimately, I do not believe these two projects are that different, since instances of the successful interdisciplinary theories in cognitive science have already been done properly.

But before we can work our way to what counts as an interdisciplinary theory, I need to lay some theoretical groundwork. I begin in the next chapter by explaining a bit of the metaphysical relationship between the mind and the brain. In particular, I focus on how to understand mental causality in a physical world. I argue that the real problem is not defining causality but, instead, is deciding which level of analysis to privilege in any explanation. This decision ultimately turns on contingent facts, local to the particular investigation. I return to the question of levels of organization and analysis in chapter 3 and spend more time developing accounts of the types of hierarchies one finds in the cognitive sciences and how one should use these hierarchies in explaining mental or biological phenomena.

Chapter 4 introduces computationalism and functionalism, two popular methodologies in cognitive science. I argue that both of these have strongly pragmatic dimensions as well that belie using computer analogies as primitives in theory building. As I explain in chapters 4 and 5, much of science relies on a bootstrapping method, which in turn is determined by the history of the inquiry. So, for example, our needs and goals in developing an explanation determine the boundaries of the physical system we investigate. We then rely on previously accepted theories to determine which functions a physical system computes, as well as to give a semantic interpretation to the arguments of the function.

Indeed, most of this book emphasizes the pragmatic aspects to scientific pursuits. Theories are accepted as an answer to some specific questions, which have been posed against a background of common assumptions, and with respect to specific alternatives. In addition, available methodological and empirical techniques influence what level of description one can give a physical system and consequently influence the types of legitimate questions one can pose. The bottom line is that theories are tied to a particular scientific community, operating during a particular time, with particular players.

As I argue by example in chapters 6 and 7, these pragmatic aspects of science entail that it is difficult simply to import data or theories from one investigative domain to another. Indeed, classical notions of theoretical reduction simply can't work. It would be a rare case in which one theory from one field could explain the success of another theory in a different field, since the players, the background assumptions, the history of the discipline, and the contrast class of the questions asked would be different as well. Still, I don't want to maintain that the disciplines involved in the cognitive sciences are completely independent. Instead, something like the notion of "explanatory extension" captures the two-way dependency relations among autonomous fields. A multidisciplinary approach to problem solving means that we may use another discipline for collateral support, inspiration, and to help set the parameters of inquiry, but we cannot simply borrow data wholesale from other theories over the same state space. Such borrowings would represent an abuse of evidence because in order to import in such fashion, we must overlook important aspects of the theoretical assumptions behind the data gathering, assumptions that shape not only the very nature of the evidence but the nature of the abstract physical system as well.

I conclude that the best way to approach developing an interdisciplinary program would be to rely on "bridge sciences" for important connections. Such a program requires that at least some of the domains involved share fundamental assumptions. In the case study I examine, neuropsychology adopts the theoretical framework of some aspect of cognitive psychology in order to investigate psychological questions using neuroscientific techniques. This sort of "data borrowing" is legitimate because portions of the contrast classes, scientific audience, and theoretical vocabulary are in fact shared.

In sum: interdisciplinary theories in cognitive science are going to be messy affairs, operating on many different levels of analysis and description. In general, they will function as an overlapping set of related models whose explanatory power is based on a sort of etiological story telling of the development and occurrence of some attribute. The models refer to the exemplars of natural kinds, and the resultant theory, which is but a set of models and a list of general principles, maintains its coherency in virtue of these common principles. I explain by example how this is supposed to work. Though I focus on research in mnemonic processing for the examples, what I have to say should be taken as a general account of how to build a theory in cognitive science.

Let us now get to work.