Topics in Cognitive Science 1 (2009) 11–38

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2008.01003.x

The Place of Modeling in Cognitive Science

James L. McClelland

Department of Psychology and Center for Mind, Brain, and Computation, Stanford University

Received 13 October 2008; received in revised form 17 November 2008; accepted 17 November 2008

Abstract

I consider the role of cognitive modeling in cognitive science. Modeling, and the computers that enable it, are central to the field, but the role of modeling is often misunderstood. Models are not intended to capture fully the processes they attempt to elucidate. Rather, they are explorations of ideas about the nature of cognitive processes. In these explorations, simplification is essential—through simplification, the implications of the central ideas become more transparent. This is not to say that simplification has no downsides; it does, and these are discussed. I then consider several contemporary frameworks for cognitive modeling, stressing the idea that each framework is useful in its own particular ways. Increases in computer power (by a factor of about 4 million) since 1958 have enabled new modeling paradigms to emerge, but these also depend on new ways of thinking. Will new paradigms emerge again with the next 1,000-fold increase?

Keywords: Modeling frameworks; Computer simulation; Connectionist models; Bayesian approaches; Dynamical systems; Symbolic models of cognition; Hybrid models; Cognitive architectures

1. Introduction

With the inauguration of a new journal for cognitive science, 30 years after the first meeting of the Cognitive Science Society, it seems essential to consider the role of computational modeling in our discipline. Because the invitation suggested a consideration of the past history and future prospects of the approach, I will begin with a discussion of the forces that shaped the field. Perhaps the primary force was the invention, and rapid technological development, of the digital computer. Computers today are several million times faster than they

Correspondence should be sent to James L. McClelland, Department of Psychology, Stanford University, Stanford, CA 94305. E-mail: mcclelland@stanford.edu

were 50 years ago, and changes in computer power have been necessary conditions for the emergence of new paradigms in our field. Yet the conceptual frameworks researchers bring to cognitive modeling also play an important role. Computers are among the forces that influence these frameworks, but there are other forces at play as well.

The second section takes up the question of the role of modeling in the effort to understand human cognitive abilities, with a particular focus on the relationship between models and underlying theories, and on the process through which modeling work gets done. I argue that we should think of models as tools for exploring the implications of ideas. They can teach us things about the consequences of particular ways of construing the processes that take place when humans engage in particular kinds of cognitive tasks, with sometimes surprising consequences. Within this context I consider the role of ''keeping it simple''—simplification is essential but has its pitfalls, and different simplifications are required to explore different issues. Relatedly, successes and failures in fitting models to data need to be considered with care. A good fit never means that a model can be declared to provide the true explanation for the observed data; a poor fit likewise does not necessarily show that the core principles embodied in a model are necessarily the source of the misfit. Models are research tools that have their strengths and weaknesses, like other tools we use as scientists.

In the third section I consider a range of contemporary frameworks or paradigms for cognitive modeling, treating each as having a legitimate motivation and a natural domain of applicability and discussing my perception of some of the issues facing each.

The final section looks towards the future and asks, What will the next 1,000-fold increase in computer power bring to our discipline? According to Moore's law, such an increase should occur within about 20 years, allowing increases in the realism, content, and situatedness of future cognitive models. I will argue, however, that the effect on our discipline will not be profound unless new conceptual frameworks emerge as well.

2. The role of the computer in the early days of cognitive science

Let us just think for a minute about where the field of cognitive science would be without the computer. Can this be imagined at all at this point?

I can imagine it because my early training in psychology was, you might say, precomputational. As an undergraduate, I became involved with research in the psychology department at Columbia University. I studied several flavors of behaviorist psychology; sensory mechanisms as investigated through physiology and psychophysics; and signal detection theory. Even though I was an undergraduate in the late 1960s, and the computer was by then over 20 years old, there were still places where its influence had not reached.

The invention of the computer, though, was having a dramatic impact elsewhere, as I learned when I reached graduate school. There I encountered cognitive psychology and psycholinguistics, each of which reflected computational ideas.

According to Neisser (1967), the author of the then-definitive *Cognitive Psychology*, the invention of the computer made it possible for psychologists to overcome their reluctance to think in terms of processes they could not directly observe, and it contributed to the demise of

behaviorism. Clearly, although someone looking at a computer could not observe them, complex processes were taking place in computers when they ran, allowing computers and programs running on them to produce outputs in response to inputs. It was not too much of a leap to see that such unobservable processes might also be taking place inside people's minds.

As if to prove the point, the computer was enabling a whole new kind of scientific investigation: the computational simulation of human-like mental processes, as these were construed by researchers coming from different backgrounds. Let us consider two examples of such work from the late 1950s.

2.1. Newell and Simon's logic theorem prover

In his autobiography, Herb Simon (1991) talks of walking into class in January 1956 and announcing that ''Over the Christmas Holiday, Alan Newell and I invented a thinking machine.'' He was referring to the creation of a computer program that could prove theorems in logic based on a few axioms. Working within the dominant paradigm of the von Neumann computer, itself arising from philosophy, mathematics, and logic, proving theorems seemed to Newell and Simon to be the essence of what it was to think. The fact that not everyone agreed just made the question interesting.

2.2. Rosenblatt's Perceptron

Rosenblatt (1961) viewed his *Perceptron* in comparably grandiose terms. In his view, building a computer to execute some algorithm dreamed up by a programmer was not so important as building a computing machine that could learn from experience. He thus set out to propose a procedure that could be programmed into a computer that would allow it to learn any function anyone wanted it to be able to compute.

The contrast between the approaches of Newell and Simon, on the one hand, and Rosenblatt, on the other, could not have been starker. And, as is well known, there was a period of time when approaches like Newell and Simon's fared better than approaches like Rosenblatt's. There were surely many reasons for this, but I will focus on two that may be instructive.

2.3. Neural networks meet overpowering limitations

First, Rosenblatt was limited by his assumptions. He adopted the assumption that the computing elements in his system (neuron-like processing units) should be treated as binary switches. This idea, sometimes called the McCullogh–Pitts neuron, represented a marriage of neurophysiology and propositional logic (McCullogh & Pitts, 1943). Although the switch depended on a graded signal (the sum of a set of graded weights from a set of input switches in the ''on'' state, less the value of a graded threshold), its output was either a 0 or a 1. In consequence, it had no derivative, limiting the possibility of construing Rosenblatt's learning rule as a gradient descent procedure.

Second, Rosenblatt was limited by the available resources. Computers at that time were quite slow by modern standards, and time on them was very scarce. In one passage in his 1961 book (pp. 558–560), Rosenblatt describes a set of simulations attempting to test the utility of a procedure for adjusting the incoming connections to a set of hidden units. The simulations produced a slight effect in the right direction, but he explained that enhancements could not be explored, as the cost of running the simulations was prohibitive.

A procedure for training such connections finally came into use about 25 years later (Rumelhart, Hinton, & Williams, 1986), after overcoming both of these limitations.

Rosenblatt's simulations were carried out on an IBM 704 computer at New York University. This was the first computer produced as a commercial product, and it cost two million dollars in the mid-1950s, according to materials available on the IBM Website (2008). The computer could carry out 4,000 floating point operations per second. While the power of this computer was a source of amazement at the time, in late 2008 one can buy a desktop computer for under $2,000 that carries out 15 billion floating point operations per second. The cost per floating point operation is a factor of about 4 billion less in 2008 than it was in 1958, and calculations that would have taken a full year on the 704 in 1958 (∼32 million seconds) take under 10 seconds on a 2008 desktop.

Of course, some other people thought there were more fundamental problems with Rosenblatt's approach. Minsky and Papert (1969) argued that the limits of Rosenblatt's Perceptron were both fundamental and profound, and argued for the virtues of discrete, serial computation. Their arguments were built up from a set of premises that included some of the limits that Rosenblatt had placed on himself (some of which I have not mentioned).

Minsky and Pappert's arguments seemed compelling at the time, and the late 1960s and early 1970s represented a winter for neural networks, while symbolic approaches flourished. We will come back to issues raised during this period, but for now, we can perhaps pause to reflect for a moment.

Invention of the computer created the field of Artificial Intelligence and was essential for Cognitive Psychology as well. It enabled a way of thinking and a way of assessing the adequacy of this thinking through computer simulation. It provided tools that dramatically extended the reach of the human mind, allowing it to calculate out the consequences of long sequences of operations, which previously had been impossible. But it still had limitations. It encouraged ways of thinking that may have led away from what might have been a good idea. And, although the mind boggled every time one asked again how many arithmetic operations a computer could carry out per second, its capabilities were tremendously limited, compared to what we now take for granted. We should ask ourselves these questions, at least from time to time: What limitations do our current ways of thinking place on our ideas about the computations our minds can perform? And, what barriers might fall if computers continue their exponential improvements in performance?

2.4. The resurgence of neural networks

These points came home in the early 1980s when there was suddenly a resurgence of interest in neural networks. This time the approach had much more success, in part because

some of the thinking about neural networks had changed. Both Grossberg (1978) and James Anderson (1977)—two pioneers who explored neural networks in spite of Minsky and Pappert's dicta—relied on continuous rather than discrete activations for their neuron-like processing units. One could then compute the derivative of the activation function, and this in turn allowed Rumelhart and others to derive the learning rule for training hidden units with back-propagation.

It is important to note that Rumelhart did not allow the oft-expressed concern about the biological implausibility of back-propagation to constrain his thinking. If he had, he would not have developed his then-exciting new algorithm. Constraints on our thinking that bound our ability to exploit computational resources fully can come from any source. We should always be on the lookout for them.

Overcoming conceptual constraints was essential, but the wave of interest in neural networks could not have been launched without the increases in speed and availability of computation. In fact, back-propagation networks train slowly. They can show little progress during initial phases of training on some problems, even over many sweeps through the training set. When Rumelhart first tried back-propagation on the XOR problem (the simplest problem that the Perceptron could not solve), he did have a desktop computer, but it was still about 1,000 times slower than the computer on my desktop today, and after several hours, it had not yet solved XOR. So (as I remember the story) Rumelhart thought that his algorithm was not working, and it was only a year or so later, when he had a faster computer on his desktop, that he got around to trying it again. This time he left it running overnight and came back in the morning to find that the computer had solved the problem. On today's computers, backprop solves XOR in a fraction of a second.

Controversy still surrounds back-propagation. Many still feel it is biologically implausible and/or too slow (specifically, it takes exponentially more training time as each additional layer of units is added). But there is no doubt that it unleashed a huge wave of interest in new forms of computational modeling in the mid-1980s and still underlies many models in cognitive science. Successors of back-propagation and related algorithms are alive and well and at the forefront of Machine Learning. There has been real progress in speeding up learning, due to architectural refinements and more sophistication in the algorithms (Bengio, Lamblin, Popovici, & Larochelle, 2007; Hinton & Salakhutdinov, 2006; Ranzato, Poultney, Chopra, & LeCun, 2007).

A similar story might be told about the recent emergence of computationally demanding methods such as Markov chain Monte Carlo that allow Bayesian inference to be applied to a wide range of problems where this would previously have been impractical. Indeed, Lee and Wagenmakers (2005, p. 662) state that ''[w]ith the advent of modern computing technology, Bayesian inference has become possible in practice as well as in principle and has been adopted with great success in most scientific fields.'' Perfors, Tenenbaum, and Regier (2006) use such methods in their work demonstrating that it is possible to select which of several alternative grammars is best, based on only a few hours worth of spoken language input. Perfors (personal communication, November 12, 2008) estimates that computations carried out for just one part of the results presented in related, ongoing work (A. Perfors, J.T. Tenenbaum, and T. Regier, unpublished data) took about 352 hr, or 11 days. With computer

speed doubling every 2 years, these calculations would have taken a year and a quarter 10 years earlier. Perfors notes that the code was not optimized. Even supposing optimization would allow a four-fold speedup, the computations would still have taken several months.

The point so far is the following: Computers have enabled several phases of new thinking about cognitive processes. Their ability to do so depends on how fast they are, and also on what constraints modelers impose on their thinking. Even though Moore's law will give out at some point, it seems likely that the future will hold important breakthroughs in computing power, and this will in turn allow new paradigms for computational modeling, especially if we allow our minds to wander freely over all possible ways that computers might be effectively used.

3. The nature and roles of models in cognitive science

My goal in this section is to consider how we should think about what models are and what roles they play in cognitive science (including cognitive neuroscience and cognitive psychology). I will draw examples from my own collaborative investigations because these are the modeling studies that I know best, and examples from these studies allow me to describe specific experiences in developing these models that form the basis for my perspective on these issues. I believe, however, that the points I make here are equally applicable for all kinds of cognitive modeling, and indeed for modeling in domains far afield from cognitive science.

The main point of this section is the following: The essential purpose of cognitive modeling is to allow investigation of the implications of ideas, beyond the limits of human thinking. Models allow the exploration of the implications of ideas that cannot be fully explored by thought alone. As such, they are vehicles for scientific discovery, in much the same way as experiments on human (or other) participants. But the discoveries take a particular form: A system with a particular set of specified properties has another set of properties that arise from those in the specified set as consequences. From observations of this type, we then attempt to draw implications for the nature of human cognition. Alas, such inferences are under-constrained, and there is often room for differences of opinion concerning the meaning of the outcome of a given investigation. The implication of this is not that models are useless, but that care in interpreting their successes and failures, as well as further investigation, is usually required. This makes modeling an ongoing process, just like other aspects of science.

3.1. Using models to explore the implications of ideas

I first came upon the need to develop a model in the mid-1970s when I asked myself this question: What if processing involved a series of stages, as Sternberg (1969) had urged us to suppose, but propagation of information from one stage to another was graded in both time and value, rather than being discrete? I thought and wrote about this question in a paper I submitted to *Psychological Review* in 1975. The reviewers thought the ideas were

interesting, but they could not really understand what I was proposing, and the editor urged me to develop a computer simulation model. At about the same time, I was exposed to the neural network models of James Anderson (1977). These models allowed me to formulate my question in a simple and explicit way (McClelland, 1979). I considered a series of stages, each containing a set of simple processing units, such that the activation of a given unit i at level l was viewed as being driven by the difference between a value $v_{il}(t)$ determined by the activations of all the units indexed by j at level $l - 1$ and its current activation:

$$da_{il}/dt = k_l(v_{il}(t) - a_{il}(t)), \quad \text{where } v_{il} = \sum_j w_{ij}a_{jl-1}(t)$$

For simplicity, I considered the case where the processing rate parameter, k_l, was free to take different values at different processing levels but was fixed for all units within a level. To model a trial in a reaction time experiment, input to the first stage was switched on and stayed on at time $t = 0$ until a response was emitted at the final, response stage of processing. In modeling standard reaction time experiments, this occurred when the activation of one of the units at the response stage reached a threshold.

Exploration of this model revealed several things that were previously unknown in cognitive psychology. First, it revealed that manipulations that affected the dynamics of two separate processing stages (i.e., the rate constants k_l for two different values of l) would generally produce additive effects on reaction times, just as in discrete stage models. This was important because it made clear that evidence of additivity did not imply discreteness, contra conventional wisdom at the time.

The analysis also revealed that manipulations affecting asymptotic activation would generally produce failures of additivity when combined with each other and with manipulations affecting dynamics of any of the processing stages. This was important because it meant that Sternberg's additive factors logic no longer provided a fool-proof recipe for determining whether manipulated variables affect the same stage or different stages.

The key point here is that the model taught us something about the implications of thinking in terms of continuous process that was not previously understood by psychologists. (The model had already been in use in the study of chemical reactions, where its implications for the time course of reactions had long been known.)

3.2. The issue of simplification

I was able to obtain the results above due to the simplifications I made in formulating the cascade model. The simplifications included the stipulation that all processing is strictly feedforward, and the stipulation that the v_{il} values are always strictly linear weighted sums of the v_{jl-1} values. This made it possible to separate factors affecting dynamics of processing from factors affecting asymptotic activation, and it was crucial for the core results of the analysis described above. However, some of the simplifications I adopted have led to

problems (Ashby, 1982; Roberts & Sternberg, 1993). It is therefore worth asking, Can simplification in science be justified? Should we not always strive for verisimilitude at all times?

In an important parable that bears on this issue, Borges (1998) describes a town where there are mapmakers who are obsessed with verisimilitude in their mapmaking. Each strives to outdo the others in making his maps more detailed and realistic. Some mapmakers are criticized because their maps are too small—their scale prevents recording of many details. Others are criticized for schematic rendering of roads and intersections. The consequence is the construction of huge, life-size maps, which, of course, are completely useless because use of such a map is no easier than direct exploration of the real space that the map represents. When it comes to mapmaking, simplification is evidently crucial—the point of the map is to offer a guide, rather than a replication, of reality.

It is no different with computational cognitive modeling. We strive to understand, and in so doing, we must simplify. The more detail we incorporate, the harder the model is to understand. Good cognitive models should strive for simplicity, and consumers of these models should understand that the simplification is adopted to aid understanding. Even if our computer power keeps growing, and simulation of the brain at the level of the interactions between the participating subatomic particles becomes possible, simplification will remain of the essence in allowing understanding.

But simplification does impact on what we can conclude from our investigations. It is possible to adopt a simplification that limits the phenomena that a model can address, or that even leads it to make incorrect predictions. In fact, this happened in the case of the cascade model (Ashby, 1982; Roberts & Sternberg, 1993). Although I found it could account quite well for the relationship between accuracy and time allowed for processing in time-controlled processing tasks (tasks where, for example, the participant must respond immediately when a response signal occurs), I did not address variables other than the mean reaction time in applying the model to standard reaction time tasks. When Ashby (1982) applied the model to addressing other properties of reaction time distributions, inadequacies of the model emerged (see also Roberts & Sternberg, 1993). It is my conjecture that the inadequacies of the model stem in this case from oversimplification of the sources of variability (reasons for this conjecture are discussed below), but this remains an issue that awaits further investigation.

The point is simply this: Simplification is essential, but it comes at a cost, and real understanding depends in part on understanding the effects of the simplification. Unfortunately, this can mean that further exploration becomes more technical and complex as a result. Trying hard to add just enough additional complexity can help. Learning what simplification is the best one to use is also a part of the process. Some simplifications do a better job retaining essential properties of a process than others.

3.3. Do models embody ''assumptions''?

At the beginning of this section, I spoke of a *set of specified properties* and a *set of properties that arise as consequences*. I have studiously avoided the word ''assumption,'' which

is often used to refer to a specified property of a model, because the word seems to create the possibility to mislead. Consider this passage from Pinker and Prince's (1988, pp. 95–96) discussion of the Rumelhart and McClelland (1986) past tense (RM) model:

> These are the fundamental linguistic assumptions of the RM model: That the Wickelphone/Wickelfeature provides an adequate basis for phonological generalization, circumventing the need to deal with strings. That the past tense is formed by the direct modification of the phonetics of the root […]. That the formation of strong (irregular) pasts is determined by purely phonetic considerations.

Rumelhart and I did not intend these properties of the model as fundamental linguistic assumptions. Instead we intended to explore the consequences of a set of specified properties in our effort to address a more fundamental question about the need for linguistic rules. In fact, we specifically stated that we made no claims whatsoever for Wickelphones and Wickelfeatures other than that they sufficed to provide distinct yet overlapping representation of the 500 most frequent verbs of English and that their use conformed to our commitment to rely on distributed representations (Rumelhart & McClelland, 1986, p. 239). Let me therefore rewrite the above passage for Pinker and Prince as follows:

> The RM model explores the possibility that Wickelphones/Wickelfeatures might provide an adequate basis for representation and generalization of past tense knowledge. The model examines a simplified situation in which the past tense is formed by the direct modification of the phonetics of the root, and leaves aside the possible role of semantic factors for simplicity.

There is no doubt that Rumelhart and I saw the results of our simulation as relevant to broad issues in linguistic theory, in that the model offered an alternative to then-uncontested views about the nature of language knowledge, and of course there is no doubt that we paid less attention to the shortcomings of our model than we did to its successes. Critics can and should point out such deficiencies, as Pinker and Prince did, and it is fine for critics to conclude, as Pinker and Prince do, that our model's successes were not sufficient to overturn the conventional wisdom of linguistic theory. But, had they treated the properties of the model above as ideas to be explored, rather than ''fundamental linguistic assumptions,'' that might have facilitated further exploration of an interesting alternative to the conventional linguistic perspective.

Luckily, others took a different view of the properties of the RM model and explored improvements in the choice of representation and the inclusion of semantic as well as phonological information (Joanisse & Seidenberg, 1999; MacWhinney & Leinbach, 1991), leading to models that overcome many of the shortcomings of the RM model. These investigators have continued to explore the importance of the essential properties of the RM models—its use of connections and distributed representations instead of linguistic rules—while adapting its other properties to facilitate such explorations.

3.4. How should a cognitive model be judged?

In fields like computer science, machine learning, robotics, and artificial intelligence, one may have the goal of building a machine with specified cognitive abilities, potentially without regard to whether these conform to the ways in which humans carry out a task. Such efforts are an important source of ideas and possibly benchmarks against which to measure human performance, but they do not appear to be the main focus of modeling in cognitive science. Judging from the work presented at the Cognitive Science Society meetings and appearing in the field's journals, this work is usually concerned with the goal of capturing the essence of human cognitive abilities. Even so, it is important to consider both the sufficiency of a model as well as its ability to explain the patterns of data obtained in experiments.

The sufficiency criterion, urged by Newell (1973), reflected his engineering/artificial intelligence perspective: His priority was always to build models that actually achieve some desired computation. As Newell pointed out, many so-called cognitive models do not have that goal, and do not even consider it. Meeting the sufficiency criterion is not trivial, of course: We still lack machines that can understand simple sentences or recognize the objects present in a picture at anything remotely approaching human proficiency. These are examples of hard problems that have not been solved yet, and explorations of new ideas that help us understand how they might be solved—quite apart from how they actually are solved—are certainly important for cognitive science. Fields surrounding cognitive science, including artificial intelligence, robotics, and machine learning, are dedicated to such explorations.

More recently there has been a shift in emphasis from sufficiency as Newell framed it to rationality or optimality. I will come back to this in discussing contemporary approaches to modeling cognition below. Suffice it to say that the question: What is the rational or optimal policy in a certain situation? is itself a difficult question to answer; nevertheless, when it is possible to determine what would be optimal under a given construal of an appropriate objective function, that provides an important benchmark against which to compare human performance, and determining what is optimal has certainly become an important part of cognitive modeling and cognitive science.

Whether a model is sufficient or optimal, it may not do the task in the same way that humans do—for example, it may go through a different sequence of processing steps and intermediate states along the way to performing the task, or by using a fundamentally different procedure—in short, the process may be different. I think most cognitive scientists are concerned with understanding cognitive processes, and I count myself among this process-oriented group. Therefore, I will consider below how we can go about comparing models to evidence obtained from experiments with humans and animals to help us learn more about the nature of these processes.

Of course, the algorithms and sequences of intermediate states involved in performing any particular task are generally not directly accessible to inspection, and so, to gain evidence relevant to constraining our theorizing about what these processes might be, we generally assess their empirical adequacy to account for details of human performance. Of course, many researchers interested in the mechanisms of cognition study them using

neuroscience methods, including neuronal recordings using microelectrodes, and/or non-invasive imaging methods such as fMRI, MEG, and EEG. While neuroscience evidence is likely to continue to grow in importance for cognitive science, my comments here are largely restricted to how we use details of human performance—the patterns of overt responses people make and the patterns of response times we see as they make them—in evaluating and improving cognitive models.

I see no clear distinction between the sufficiency, optimality, and empirical adequacy criteria, in part because I view human performance as inherently probabilistic and error prone. Human abilities are far less than fully sufficient and are certainly not optimal in many tasks. Once that is accepted, if the goal is to understand human cognition, the degree of sufficiency and/or optimality of the model can be part of the assessment of its empirical adequacy: Does a model carry out some task as well as humans do, and does it achieve optimality to the same degree and deviate from optimality in the same way as humans? These are important aspects of assessing the empirical adequacy of a model, and they often guide researchers to consider particular aspects of human performance.

3.5. Assigning credit and blame, and learning from a model's failures

When a model fails to capture some aspect of human performance, it represents both a challenge and an opportunity. The challenge is to determine just what aspect or aspects of the model are to blame for the failure. Because the model is an exploration of a set of ideas, it is not clear which members of the set are at fault for the model's shortcomings. Model failures also present an opportunity: When a model fails it allows us to focus attention on where we might have been wrong, allowing real progress to arise from further investigations.

An example of this situation arose with the distributed developmental model of single word reading (Seidenberg & McClelland, 1989). This model employed a feed-forward, connectionist network that learned to assign the correct pronunciations for a set of words that were included in its training set. In our 1989 paper, Seidenberg and I compared the model to data from a wide range of studies investigating word reading and found that it fit the data very well. The model could read both regular and exception words correctly and showed strong regularity effects, as human readers do. We therefore argued, contra many people's intuitions, that a single integrated ''route''—rather than two separate routes for regular and exceptional items—provided a sufficient model of single word reading abilities.

Critics, however, argued that we had failed to test the model thoroughly enough (Besner, Twilley, McCann, & Seergobin, 1990; Coltheart, Curtis, Atkins, & Hailer, 1993). They noted that human participants can also read pronounceable nonwords and for some such items they reliably produce the same consistent answer (for some sets of nonwords, the answer conforms to a fairly simple set of rules about 95% of the time). Besner et al. tested the model on such items, and they found that it only achieved about 60% consistency with the rules. This failure of our model led supporters of the dual-route theory to argue that no single computational procedure could actually read both exception words and pronounceable nonwords correctly.

We certainly agreed with our critics that the model was in some way incorrect but suspected that the problem lay (as in the past tense model) with the particular details of the input and output coding scheme used in the implementation. Indeed, in subsequent work (Plaut, McClelland, Seidenberg, & Patterson, 1996), we found that, with an improved coding scheme, a network very similar to the original Seidenberg and McClelland model was able to read both exception words and pronounceable nonwords at human-like levels of proficiency while still exhibiting the other patterns of findings captured by the earlier model.

A similar situation also arose in investigations of the interactive activation model (McClelland & Rumelhart, 1981). The model was intended to allow the exploration of the principle of interactive, or bidirectional, propagation of activation. The model was successful in accounting (albeit in a largely qualitative way) for a wide range of preexisting (McClelland & Rumelhart, 1981) as well as new data (Rumelhart & McClelland, 1982), and similar successes were achieved by the TRACE model (McClelland & Elman, 1986), which extended these ideas to speech perception. But Massaro (1989), a proponent of a feedforward approach to perception of both printed and spoken words, pointed out, quite correctly, that the model's performance did not adhere to the quantitative form of data obtained in a number of experiments investigating the joint effects of two separate factors on letter or phoneme identification.

Massaro argued that the problem with the model lay specifically with its interactive architecture. Once again, Massaro was right that the model was flawed but incorrect in identifying the source of the problem. The problem with the model can be traced to the deterministic nature of the activation process in the model (I suspect this may also be the source of the shortcomings of the cascade model). To demonstrate this, I showed that incorporating intrinsic variability into the interactive activation model successfully allowed it to capture the quantitative form of the data (McClelland, 1991).

3.6. What exploration of models can teach us

To summarize the point of this section, models can be used to explore the implications of ideas, and in particular to assess their sufficiency, optimality, and empirical adequacy, but such explorations must be carried out with care. Because each model embodies a set of properties, it is not straightforward to determine which of these properties is at fault when a model fails. The failures are clearly important, and absolutely must be addressed, but, based on the examples above, it would appear that caution is in order when it comes to reaching broad conclusions from the shortcomings of particular models.

3.7. What exploration of models cannot teach us

In the above I have stressed things that we can learn from exploring models, and I have argued for the importance of these explorations. Here I would like to address something that a model cannot do: Even if a modeler can show that a model fits all available data perfectly, the work still cannot tell us that it correctly captures the processes underlying human performance in the tasks that it addresses.

The exploration of the fit between a model and data can only tell us whether the ideas embodied in the model are consistent with a body of facts. To the extent that it is consistent with a particular body of facts, it cannot be ruled out on that basis. But this does not mean that the model is correct, or that the properties it embodies hold the systems that it is intended to model. When a set of ideas turns out to be inconsistent with a body of data, then something about the set of ideas is in error; but when a set of ideas is consistent with a body of ideas, the only thing we can conclude is that this set of ideas cannot be ruled out.

The point I am making here is a straightforward application of Popper's (1959) philosophy of scientific inference. There appears, however, to be a great deal of confusion surrounding these issues. Perhaps more explicit acknowledgement that a good fit to data does not prove that a model is correct will alleviate some of the controversy that appears to swirl around a great deal of modeling work. Massaro (1988) and Roberts and Pashler (2000) are among those who have argued that we should not place undue credence in models just because they provide a good fit to data, and I fully agree with that. Models become candidates for further exploration (and criticism) when they account for a newsworthy amount of data (i.e., more data than it was previously thought they could account for, or data that another approach cannot account for), but no one should be fooled into thinking that the properties they embody are thereby established as correct.

3.8. Can models ever be falsified?

Some critics of modeling approaches (including Massaro, 1988 and Roberts & Pashler, 2000) have complained that the modeler has a virtually infinite toolkit of things to play with in fitting a model to data. The consequence of this, the argument goes, is that there would be no pattern of data a model could not explain, and therefore nothing would be learned from the exercise of showing that the model can fit a particular set of data.

This concern bears consideration, but in my view it is often overplayed. Consider the problem Massaro (1989) discovered with the original version of the interactive activation model. I found in following up his discovery that the problem shows up in very simple as well as very complex architectures, and it is not overcome by adjusting such things as the number of units and connections or the architecture of the network. In this and, in my experience, many other cases, it is simply not true that a given model can be fit to any pattern of data simply by twiddling its parameters. I generally find that there is some particular flaw that prevents success. In this case, attention brought to a model failure led to a better model. Progress actually can and does occur, according to the principles of Popperian science, because, as a matter of fact, achieving a good fit to a complex body of data is far from assured. Failures, which can and do occur, allow data to guide model development.

3.9. Toward increased transparency of cognitive models

As a field, we face real challenges when it comes to answering the key questions about the nature of cognitive processes, even with all the tools of computational modeling. Modeling only allows us to evaluate particular combinations of properties, which must be explored

in conjunction with deliberate simplifications as well as arbitrary choices of particular details. When a model fails to fit some pattern of actual data, we only know that something is wrong, and the problem of assigning blame is very difficult.

The fact that many models are not transparent has been a basis for dissatisfaction, particularly with connectionist models (McCloskey, 1991), but the issue arises whenever a complex model is assessed. This is not only a problem for psychology but also for many other disciplines where models are frequently used and where the reasons for their successes and failures are not always apparent. This does not mean that the models should be abandoned. Instead, they should be treated as targets of investigation in their own right so that their properties can be better understood.

How can we increase the level of transparency and understanding? One important tool our field does not exploit enough is mathematical analysis. We should strive for such analyses wherever possible, even if they require considerable simplification.

When such analyses are complemented by systematic simulations that go beyond the simplifications, it becomes possible to gain a measure of understanding. My own work has benefited tremendously from collaboration with physicists and mathematicians (McClelland and Chappell, 1998; Movellan & McClelland, 2001; Usher & McClelland, 2001) who have experience formulating tractable simplifications and using mathematical tools to analyze them. The future of cognitive modeling will continue to depend on exploiting these tools as fully as possible, and on drawing researchers into it with the relevant background.

4. Frameworks for cognitive modeling

A recent, and very useful, book, the *Cambridge Handbook of Computational Psychology* (Sun, 2008) attempts to characterize the current state of the art in the field. The book begins with a section called ''Cognitive Modeling Paradigms'' and contains chapters on connectionist models (Thomas & McClelland, 2008), Bayesian models (Griffiths, Kemp, & Tenenbaum, 2008), dynamical systems approaches (Schöner, 2008), declarative/logic-based models (Bringsjord, 2008), and cognitive architectures (Taatgen & Anderson, 2008). I prefer the terms ''framework'' and ''approach'' to ''paradigm.'' I will comment separately on each of the first four approaches, then turn attention briefly to cognitive architectures and hybrid approaches (Sun, 2002).

Attempts to compare and contrast approaches have often been framed in terms of levels (e.g., Marr, 1982). However, the question of levels does not fully capture the different commitments represented by the alternative approaches described above. For example, it is widely argued that connectionist/parallel distributed processing (PDP) models address an implementational level of analysis (Broadbent, 1985; Kemp & Tenenbaum, 2008; Pinker & Prince, 1988; Smolensky, 1988), while other approaches focus on the level of algorithms and representations or on Marr's highest ''computational'' level. My colleagues and I have argued instead that the PDP approach exploits *alternative* representations and processes to those used in some other approaches (Rogers & McClelland, 2008; Rumelhart & McClelland, 1985), and that the framework takes a different stance at the computational

level than the one taken by, for example, Kemp and Tenenbaum (in press), in their structured probabilistic models (Rogers & McClelland, 2008). Similar points can be made in comparing the other indicated approaches (for a more extended discussion, see Sun, 2008).

Each approach has attracted adherents in our field because it is particularly apt for addressing certain types of cognitive processes and phenomena. Each has its core domains of relative advantage, its strengths and weaknesses, and its zones of contention where there is competition with other approaches. In the following I consider each of the approaches in this light. The material represents a personal perspective, and space constraints prevent a full consideration.

4.1. Connectionist/PDP models

Connectionist/PDP models appear to provide a natural way of capturing a particular kind of idea about how many cognitive phenomena should be explained while also offering alternatives to once-conventional accounts of many of these and other phenomena. The natural domain of application of connectionist models appears to be those aspects of our cognition that involve relatively automatic processes based on an extensive base of prior experience; among these aspects I would include perception, some aspects of memory, intuitive semantics, categorization, reading, and language.

It cannot be denied that connectionist modelers take inspiration from the brain in building their models; but the importance of this is often overplayed. Most connectionist cognitive scientists find this inspiration appealing not for its own sake, but for its role in addressing computational and psychological considerations (Anderson, 1977; McClelland, Rumelhart, & Hinton, 1986). For example, in the interactive activation model (McClelland & Rumelhart, 1981), Rumelhart and I wanted to explore the idea that the simultaneous exploitation of multiple, mutual constraints might underlie the human ability to see things better when they fit together with other things in the same context (Rumelhart, 1977). To capture this idea, we found that a connectionist/neural network formulation was particularly useful. The computational motivation clearly preceded and served as the basis of our enthusiasm for the use of a connectionist network. Similarly, the motivation for the development of our past tense model (Rumelhart & McClelland, 1986) was to explore the idea that people might exhibit regularity and generalization in their linguistic (and other) behavior without employing explicit rules. Further motivation came from the belief that linguistic regularities and exceptions to them are not categorical as some (Jackendoff, 2002, 2007; Pinker, 1991) have suggested but fall instead on a continuum (Bybee & Slobin, 1982; McClelland & Bybee, 2007; McClelland & Patterson, 2002). Again, the inspiration from neuroscience—the use of a model in which simple processing units influence each other through connections that may be modifiable through experience—provided a concrete framework for the creation of models that allow the implications of these ideas to be effectively explored.

Both benefits and potential costs have come from the neural network inspiration for connectionist models. On the down side, some have criticized connectionist models from concerns about their fidelity to the actual properties of real neural networks. Although I have at times shared this concern, I now think debate about this issue is largely misguided. The

capabilities of the human mind still elude our most sophisticated models (including our most sophisticated connectionist models). Our efforts to seek better models should be inspired, and indeed informed, by neuroscience, but not, at this early stage of our understanding, restricted by our current conception of the properties of real neural networks.

In a similar vein, I do not wish to give the impression that relying on a neural inspiration somehow constitutes support for a connectionist model. Just as it seems to me improper to rule out a model because it does not seem biologically plausible enough, we would not want to think that a model deserved special credence at the psychological level because it adopted some specific idea from neuroscience as part of its inspiration.

In sum, PDP/connectionist models in cognitive science are offered primarily for what their protagonists see as their usefulness in addressing certain aspects of cognition. The inspiration from neuroscience has provided a heuristic guide in model development, but it should not be used either as a privileged basis for support or as a procrustean bed to constrain the further development of the framework.

4.2. Rational and Bayesian approaches

Rational approaches in cognitive science stem from the belief, or at least the hope, that it is possible to understand human cognition and behavior as an optimal response to the constraints placed on the cognizing agent by a situation or set of situations. A ''rational'' approach would naturally include Bayesian ideas, which can provide guidance on what inferences should be drawn from uncertain data, but in general it would also incorporate cost and benefit considerations. The costs include time and effort, which clearly place bounds on human rationality (Simon, 1957).

It certainly makes sense to ask what would be optimal in a given situation, thereby providing a basis for judging whether what people actually do is or is not optimal, and for focusing attention on any observed deviations. I am fully on board with this effort, and it seems clear that it can inform all sorts of modeling efforts.

There are also cases in which a rational analysis has correctly predicted patterns of data obtained in experiments. As one example, Geisler and Perry (in press) carried out a statistical analysis of visual scene structure to determine the relative likelihood that two line segments sticking out from behind an occluder are part of a single continuous edge or not, as a function of the two segments' relative position and orientation. We can intuit perhaps that approximately colinear segments are likely to come from the same underlying edge; Geisler and Perry went beyond this to uncover the details of the actual scene statistics. In subsequent behavioral experiments, Geissler and Perry found that human observers' judgments of whether two segments appeared to be from the same line conformed to the relative probabilities derived from the scene statistics. Both the scene statistics and the judgments did not match predictions based on a number of other prior proposals. Here is a clear case in which knowing what is optimal led to interesting new predictions, subsequently supported by behavioral data, and indicating that, in this case at least, human performance approximates optimality. Similarly, rational (Bayesian) approaches have been useful in predicting how human's inferences concerning the scope of an object or category

label is affected by the distributional properties of known exemplars (Shepard, 1987; Xu & Tenenbaum, 2007).

One reaction to some of this work is that it explores Bayesian inference in a very specific inferential context. For example, in Kemp and Tenenbaum (in press) the goal of learning in a cognitive domain is construed as being to discover which of several preexisting types of knowledge structure best characterizes observations in a particular domain. While the authors eschew any commitment to the particular procedures used for calculating which alternative provides the best fit, their approach specifies a goal and a set of alternative structures to consider. Theirs is not, therefore, simply a Bayesian model but a particular model with a number of properties one might or might not agree with, quite apart from the question of whether the mind conforms to principles of Bayesian inference. What separates some of the models my collaborators and I have explored (e.g., Rogers & McClelland, 2004) from these approaches may not be whether one is Bayesian and the other connectionist, but whether one views learning in terms of the explicit selection among alternative prespecified structures.

Considering the ''rational models'' approach more broadly, one general problem is that what is ''rational'' depends on what we take to be an individual's goal in a given situation. For example, if a human participant fails to set a hypothesized response threshold low enough to maximize total reward rate in a reaction time experiment, we cannot judge that the participant is behaving nonoptimally, because we may have missed an important subjective element in the participant's cost function: to the participant, it may be more important to be accurate than to earn the largest possible number of points in the experiment. If we allow a free parameter for this, we unfortunately can then explain any choice of threshold whatsoever, and the appeal to rationality or optimality may lose its force, unless we can find independent evidence to constrain the value of the free parameter. Another problem is that what is rational in one situation may not be rational in another. We then confront a regress in which the problem for rational analysis is to determine how we should construe a given situation.

Appeals to evolution as an optimizing force are often brought up in defense of optimality analysis. The difficulties with this were clearly spelled out by Gould (1980) in *The panda's thumb*: Evolution selects locally for relative competitive advantage, not for a global optimum. The forces operating to determine relative competitive advantage over the course of evolutionary history are themselves very difficult to determine, and the extent to which whatever was selected for will prove optimal in a given situation confronting a contemporary human is also difficult to judge. It is possible to agree fully that selection shapes structure and function while stopping short of a conclusion that the result is in any way optimal, much less optimal in the contemporary world, which is so different from the one we evolved in.

Thus, as with seeking one's inspiration from the properties of biological neural networks, seeking one's inspiration from a rational analysis is no guarantee of success of a cognitive model. In my view, we should evaluate models that come from either source for their usefulness in explaining particular phenomena, without worrying too much about whether it was neural inspiration or rational analysis that provided the initial motivation for their exploration.

4.3. Dynamical systems approaches

The dynamical systems approach begins with an appeal to the situated and embodied nature of human behavior (Schöner, 2008). Indeed, it has mostly been applied either to the physical characteristics of behavior (Does a baby exhibit a walking gait when placed in a particular posture? Thelen & Smith, 1994), or to the role of physical variables in determining behavior (Does the physical setting of testing affect the presence of the A-not-B error? Thelen, Schöner, Scheier, & Smith, 2001). The presence or absence of the behaviors in question has previously been taken by others to reflect the presence or absence of some cognitive or neural mechanism; thus, dynamical systems approaches have clearly brought something new into consideration. For example, babies held in a standing position exhibit a stepping gait at a very early age but do not do so when they are a little older. A dynamical systems approach explains this not in terms of the maturation of top-down inhibitory mechanisms, but in terms of the greater weight of the legs of the older baby (Thelen & Smith, 1994).

The idea that dynamical systems ideas from the physical sciences might usefully apply to cognitive and psychological modeling is still a fairly new one—partly because the analysis of physical dynamical systems is itself relatively new. While there can be little doubt that the approach has led to some interesting models of aspects of performance in physical task situations, the idea that it will be possible to build an interesting theory of cognition within this approach, as Schöner (2008) proposes, remains open at this stage.

To date at least, the dynamical systems approach has also been relatively silent about the processes that give rise to changes in behavior in response to experience and over the course of development. Continuous underlying change in some parameter is often assumed to account for developmental differences, and the occurrence of discontinuous changes that can result from such changes has been demonstrated using simulations based on this approach. But the source of change often comes from outside the model, as an externally imposed ''control variable.'' Finding ways to allow the changes in these variables to arise as a result of experience and behavior will greatly enhance the dynamical systems framework. Extending the framework to deal with less obviously concrete aspects of behavior will also be an interesting challenge for the future. The importance of gesture, at least as a window on more abstract mental process, and perhaps even as a factor in these processes (Alibali & Goldin-Meadow, 1993), suggests that the extension may well have considerable value.

4.4. Symbolic and logic-based approaches

The notion that thought is essentially the process of deriving new propositions from given propositions and rules of inference lies at the foundation of our philosophical traditions, and thus it is no accident that a focus on this type of process played a prominent role in the early days of cognitive modeling. Fodor and Pylyshyn (1988) articulated the view that the fundamental characteristic of thought is its ability to apply to arbitrary content through the use of structure- rather than content-sensitive rules.

It may well be that some of the supreme achievements of human intelligence have been the creation of inference systems of just this sort. These systems, once developed, have allowed such things as proofs of very general theorems and construction of beautiful systems for mathematical reasoning such as geometry, algebra, and calculus. It is therefore perhaps only natural that many cognitive theorists have sought a basis for understanding human thought as itself being essentially such a formal system. The acquisition of formal systems (arithmetic, algebra, computer programming) is a central part of modern education, and many of the errors people make when they use these systems can be analyzed as reflecting the absence (or perhaps the weakness) of a particular structure-sensitive rule (for an analysis of the errors children make in multicolumn subtraction, see Brown & VanLehn, 1980). Therefore, the framing of many cognitive tasks in terms of systems of rules will surely continue to play a central role in the effort to model (especially the more formal) aspects of human thinking and reasoning.

It seems at first glance very natural to employ this type of approach to understanding how people perform when they are asked to derive or verify conclusions from given statements and to provide explicit justifications for their answers (Bringsjord, 2008). Surely, the field of cognitive science should consider tasks of this sort. Yet there seem to me two central questions about the approach. To what extent does this approach apply to other aspects of human cognition? And to what extent, even in logic and mathematics, are the processes that constitute insightful exploitation of any formal system really processes that occur within that system itself?

The claim that the approach is essential for the characterization of language was explicitly stated by Fodor and Pylyshyn but has been challenged both by linguists (Bybee, 2001; Kuno, 1987) and by cognitive scientists (Elman, in press; McClelland, 1992). Fodor and Pylyshyn do not take the position that the structure-sensitive rules are accessible to consciousness—only that the mechanism of language processing must conform to such rules as an essential feature of its design. Indeed it has been thought that the supreme achievement of linguistics has been to determine what these rules are, even though they must be inferred from what is and what is not judged to be grammatical (Chomsky, 1957). This is not the place to review or try to settle this debate, but only to remind the reader of its existence, and perhaps just to note the agreement about the implicitness of the knowledge in this case—here, it is not assumed that native speakers can express the basis for their linguistic intuitions. Other domains in which such questions have been debated include physical reasoning tasks such as Inhelder and Piaget's balance scale task (van der Maas & Raijmakers, 2009; Schapiro & McClelland, in press; Siegler, 1976). Here the question of explicit justifications is much more complex. People often do provide an explicit justification that accords with their overt behavior but this is far from always true, making the question of whether the rule is used to guide the behavior or only to justify it a very real one (McClelland, 1995).

The idea that a logic-based approach might be applicable to explicit logical inference tasks (those in which conclusions are to be derived and verified along with an explicit justification for them) seems almost incontrovertible, and yet one may still ask whether these rules themselves really provide much in the way of explanation for the patterns of behavior

observed in such task situations. The pattern of performance on the Wason Selection Task (Wason, 1966) is perhaps the best known challenge to this framework. As Wason found, people are very poor at choosing which of several cards to turn over to check for conformity to the rule ''If a card has a vowel on one side, then it has an even number on the other.'' Performance in analogs of this task is highly sensitive to the specific content rather than the logical form of the statement (e.g., ''If the envelope is sealed, then it must have a 20 cent stamp on it''), and attempts to explain the pattern of human performance in this task are not generally framed in terms of reliance on abstract rules of logical inference (Cosmides, 1989; Oaksford and Chater, 1996).

The approach described in Bringsjord (2008) is one variant of a symbolic approach, and it differs in many ways from other symbolic approaches. Most such approaches are not limited to modeling explicit reasoning tasks and indeed are not generally construed as necessarily conforming to principles of logic. While the best-known current framework for symbolic cognition (ACT-r, Anderson & Lebiere, 1998) employs explicit propositions and production rules that capture explicit condition and action statements, these propositions and productions are not constrained to apply only to explicit reasoning tasks and, even when they do, can make reference to problem content and context, thus avoiding the problems facing an essentially logic-based approach. Furthermore, the content of production rules is not generally viewed as directly accessible, allowing for dissociations between the processes that actually govern performance and the verbal statements people make in explaining the basis for their performance.

Furthermore, both the declarative content and the production rules in contemporary production system models have associated strength variables, making processing sensitive not only to the content and context but also to frequency and recency of use of the information and to its utility. Such models generally can and often do exhibit content- as well as structure-sensitivity, contra the strong form of the idea that human thought processes are well captured as an instantiation of an abstract formal reasoning framework (Marcus, 2001). The inclusion of graded strengths and mechanisms for adjusting these strengths can make it difficult to define empirical tests that distinguish production system models from connectionist models or models arising in other frameworks. It seems likely that further developments will continue to blur the boundaries between these approaches.

4.5. Cognitive architectures and hybrid systems

The fact that each of the four approaches discussed above has its own relative strengths appears to underlie both hybrid systems (Sun, 2002) and contemporary integrated cognitive architecture-based approaches (Jilk, Lebiere, O'Reilly, & Anderson, 2008). The motivation for the cognitive architectures approach is to try to offer an instantiation of a complete human-like cognitive system. Newell (1994) explicitly advocated broad coverage at the possible expense of capturing data within each domain in all of its details, perhaps reflecting his desire to achieve a useable engineered system, in contrast to the goal of accurately characterizing the properties of human performance within a single domain.

The building of cognitive architectures seems especially natural if, as some argue, our cognitive systems really are composed of several distinct modular subsystems. In that case, to understand the functions of the system as a whole one needs not only to understand the parts but also to have an understanding of how the different parts work together.

It seems evident that there is some specialization of function in the brain, and that characterizing the different roles of the contributing parts, and the ways in which these parts work together, is worthy of consideration. Indeed, I am among those who have offered proposals along these lines (McClelland, McNaughton, & O'Reilly, 1995). I would argue, however, that some hybrid approaches take too much of an either–or approach, assigning some processes to one module or another, rather than actually considering how the components work together. I see this as an important area for future investigations. A small step in the right direction may be the ACT-R model of performance of the balance scale task (van Rijn, van Someren, & van der Maas, 2003), which uses declarative knowledge as a scaffold with which to construct production rules. I would urge consideration of models in which differentiated parts work together in a more fully integrated fashion. One example is Kwok's (2003) model of the roles of hippocampus and cortex in episodic memory for meaningful materials. In this model, reconstructing such a memory is a mutual constraint satisfaction process involving simultaneous contributions from both the neocortex and the hippocampus. Sun (2008) also stresses the importance of synergy between components in his hybrid CLARION architecture.

To return to the main point of this section, it seems clear to me that each approach that we have considered has an appealing motivation. As each has its own domains of relative advantage, all are likely to continue to be used to capture aspects of human cognition.

5. The future of cognitive modeling

According to the inventor and futurist Raymond Kurzweil, exponential growth in the power of computers will continue unabated indefinitely. In *The singularity is near* (2005), he predicts that personal computers will have the same processing power as the human brain by the year 2020. By 2045, he predicts, we will reach what has been called the technological singularity, a transition to essentially autonomous artificial intelligences, far smarter than humans, that are capable of inventing and producing superior future versions of themselves. Kutzweil's predictions in 1990 for the first decade of the 20th century were in several cases on the mark, and though one can question both the continued exponential growth in computer power and the impact that this will have on machine intelligence, there is little doubt that computing will become more and more powerful and ubiquitous. It is an interesting exercise to ponder what these developments will hold for modeling in cognitive science.

I am reluctant to project the exact future rate of expansion of the power of computers, but if Moore's law holds up, we will in 20 years have computers 1,000 times faster than today's computers. What would this power allow us to achieve as cognitive scientists? I consider three kinds of extensions of cognitive modeling that will become more and

more developed as computing power increases. Researchers are already moving in all three directions.

5.1. Models with greater fidelity to properties of real neural networks

Today I can buy a desktop computer for about $2,000 that carries out about 15 million floating point operations per second. It takes two such operations per synapse to propagate activity through a neural network (and another two for calculating weight changes) so, roughly speaking, I can now simulate a neural network with one million connections on such a computer in real time, with a temporal granularity of one update every quarter of a second. A factor of 1,000 would allow 40 times the connections and 25 times the time fidelity, that is, 40 million connections updated 100 times per second. This size network corresponds roughly to the size of the rodent hippocampus, and the time-scale is sufficient to model the timing of neuronal firing at different phases of the theta rhythm, a waxing and waning of neural activity in the hippocampus with a frequency of 8 cycles per second.

Simulating the rat hippocampus in real time at the fidelity described above should allow much fuller development of current theories of hippocampal storage and retrieval. These models (Hasselmo, Bodelón, & Wyble, 2002; Mehta, Lee, & Wilson, 2002) rely extensively on the relative timing of firing of different neurons within the theta rhythm to store associations between successive events. Another factor of 10 would allow simulation of the fast (200 Hz) oscillations that occur during hippocampal sharp waves, in which, it is proposed, time-compressed stored memories are replayed for self-maintenance and neocortical consolidation (Buzsaki, Horvath, Urioste, Hetke, & Wise, 1992).

Simulation at this scale is not merely a matter of achieving realism for its own sake. It would allow the possibility of testing whether the hypothesized mechanisms really work as well as they are claimed to by those who have proposed them, and of further elaborating the putative mechanisms whereby the brain stores, retrieves, and consolidates information.

5.2. Models with far richer knowledge bases

The example above may seem to relate primarily to neural network/connectionist and dynamical systems approaches, but the remaining cases are likely to apply to other approaches. Simply put, most existing models work on highly restricted knowledge bases and impoverished item representations. For cases like reading words aloud, where items consist of letters and phonemes, existing models can now approach realistic vocabularies including words of any length (Sibley, Kello, Plaut, & Elman, 2008). It is even possible to build up representations of word co-occurrences that allow extraction of an approximation to a representation of meaning (Griffiths, Steyvers, & Tenenbaum, 2007; Landauer & Dumais, 1997). But the human storehouse of semantic information is built up from the flux of auditory and visual experience with real objects and spoken words. Relying only on text to provide a gloss of the content of such experience, about 100,000 words a day might be about right. A 1,000-fold increase in computing capacity would greatly facilitate the effort of extracting structure from that much experience.

Again the ability to explore models of this type will be more than an exercise in achieving greater realism per se. Our existing models of human knowledge are very unlikely to have the power to use all of the information contained in such a stream of data, and they will be forced to evolve beyond their current limited scope to assimilate the vast range of content.

5.3. Increased complexity and capacity of situated cognitive agents

One exciting area of future development for cognitive modeling is in the control of situated cognitive agents interacting with other agents and the real world. Today researchers are already able to use Sony's ambulatory humanoid robot to explore simplified worlds containing small numbers of high-contrast objects that the agents can learn to approach, avoid, manipulate, and describe to each other in a simplified spoken (and heard) language (Steels, 2007). The work, just one facet of the rapidly burgeoning field of robotics, is still in its infancy, and the capabilities of the agents remain highly primitive. But as computing capability increases, such robots will come to contain many processors, some dedicated to processes within a sensory modality and others dedicated to integration of information across modalities. These robots will continue to improve in dexterity, acuity, and in their cognitive and linguistic abilities. The 1,000-fold increase in computer power will be essential to these developments.

6. What might future frameworks for cognitive modeling be like?

Future projections are notoriously difficult but, based on current trends, we can perhaps anticipate some future developments. It seems likely that existing frameworks will continue to develop and that there will be a continuing convergence and overlap of approaches. I especially look forward to a blurring of the boundaries between Bayesian/probabilistic and connectionist approaches. I think that there is a far closer correspondence between these approaches than is generally assumed (see McClelland, 1998), and I myself have recently been involved in exploring models that might be seen as hybrids between the approaches (Vallabha, McClelland, Pons, Werker, & Amano, 2007). I also see the prospect of further integration of connectionist and dynamical systems approaches (Spencer, Thomas, & McClelland, 2009). These approaches have a great deal in common, and many existing dynamical systems models are neural networks with particular properties (including intrinsic variability, topographic organization, and recurrent excitatory and inhibitory connections).

It is less clear to me how symbolic and propositional approaches will develop. I would like to think that the symbolic and propositional framework will come to be understood as providing an approximate description of a richer, more dynamic and continuous underlying process akin to that simulated in a connectionist model. Yet it seems likely that the more abstract framework will survive as a useful framework for the succinct representation of cognitive content.

Finally, I would hope that, sometime in the next 20 years, an entirely new and currently unanticipated framework will arise. It seems likely that the emergence of such a framework

will require both continued increases in computing power and fundamental new insights taking our models of the nature of cognition beyond the bounds imposed on it by our current frameworks for thinking and modeling.

Acknowledgments

The author thanks Tim Rogers for relevant discussions contributing to the formulation of some of the ideas presented in this paper, Mike Lee and Amy Perfors for input on the role of computers in Bayesian modeling, and Wayne Gray, Karalyn Patterson, and an anonymous reviewer for helpful comments on drafts of this article.

References

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture–speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, *25*, 468–523.

Anderson, J. A. (1977). Neural models with cognitive implications. In D. LaBerge, & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 27–90). Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review*, *89*, 599–607.

Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007) Greedy layer-wise training of deep networks. In B. Schölkopf, J. Platt & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS)*. Cambridge, MA: MIT Press.

Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the connection between connectionism and data: Are a few words necessary? *Psychological Review*, *97*, 432–446.

Borges, J. L. (1998). On the exactitude of science. In *Collected fictions*, trans. A. Hurley (p. 325). London: Penguin.

Bringsjord, S. (2008). Declarative/logic-based cognitive models. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 127–169). New York: Cambridge University Press.

Broadbent, D. (1985). A question of levels: Comment on McClelland and Rumelhart. *Journal of Experimental Psychology: General*, *114*, 189–192.

Brown, J. S., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, *4*, 379–426.

Buzsaki, G., Horvath, Z., Urioste, R., Hetke, J., & Wise, K. (1992). High frequency network oscillation in the hippocampus. *Science*, *256*, 1025–1027.

Bybee, J. (2001). *Phonology and language use*. New York: Cambridge University Press.

Bybee, J., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language*, *58*, 265–289.

Chomsky, N. (1957). *Syntactic structure*. The Hague: Mouton.

Coltheart, M., Curtis, B., Atkins, E., & Hailer, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.

Cosmides, L. (1989). The logic of social exchange: Was natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.

Elman, J. L. (in press). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, forthcoming.

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

Geisler, W. S., & Perry, J. S. (in press). Contour statistics in natural images: grouping across occlusions. *Visual Neuroscience*.

Gould, S. J. (1980). *The panda's thumb*. New York: W. W. Norton.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 59–100). New York: Cambridge University Press.

Griffiths, T. L., Steyvers, M. X., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*, 211–244.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes,-maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology*, Vol. 5 (pp. 233–374), New York: Academic Press.

Hasselmo, M. E., Bodelón, C., & Wyble, B. P. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation*, *14*, 793–817.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*, 504–507.

IBM Website. (2008). *704 Data processing system*. IBM Archives. Available at: http://www-03.ibm.com/ibm/history/exhibits/mainframe/mainframe_pp704.html. Accessed November 13, 2008.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, England: Oxford University Press.

Jackendoff, R. (2007). Linguistics in cognitive science: The state of the art. *The Linguistic Review*, *24*, 347–401.

Jilk, D. J., Lebiere, C., O'Reilly, R. C., & Anderson, J. R. (2008). SAL: An explicitly pluralistic cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, *20*, 197–218.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 7592–7597.

Kemp, C., & Tenenbaum, J. B. (2008). Structured models of semantic cognition. Commentary on Rogers and McClelland. *Behavioral and Brain Sciences*, *31*, 717–718.

Kemp, C., & Tenenbaum, J. B. (in press). Structured statistical models of inductive reasoning. *Psychological Review*, forthcoming.

Kuno, S. (1987). *Functional syntax: Anaphora, discourse, and empathy*. Chicago: University of Chicago Press.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York: Viking Penguin.

Kwok, K. (2003). *A computational investigation into the successes and failures of semantic learning in normal humans and amnesics*. Doctoral Dissertation, Department of Psychology, Carnegie Mellon University.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.

Lee, M. D., & Wagenmakers, E. J. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review*, *112*, 662–668.

van der Maas, H. L., & Raijmakers, M. E. J. (2009). Transitions in cognitive development: Prospects and limitations of a neural dynamic approach. In J. Spencer, M. S. C. Thomas, & J. L. McClelland (Eds.), *Toward a new grand theory of development: Connectionism and dynamical systems theory re-considered* (pp. 229–312). Oxford, England: Oxford University Press.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *40*, 121–157.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language*, *27*, 213–234.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, *21*, 398–421.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, *23*, 1–44.

McClelland, J. L. (1992). Can connectionist models discover the structure of natural language? In R. Morelli, W. M. Brown, D. Anselmi, K. Haberlandt, & D. Lloyd (Eds.), *Minds, brains & computers* (pp. 168–189). Norwood, NJ: Ablex Publishing.

McClelland, J. L. (1995). A connectionist approach to knowledge and development. In T. J. Simon, & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 157–204). Mahwah, NJ: LEA.

McClelland, J. L. (1998). Connectionist models and Bayesian inference. In M. Oaksford, & N. Chater (Eds.), *Rational models of cognition* (pp. 21–53). Oxford, England: Oxford University Press.

McClelland, J. L., & Bybee, J. (2007). Gradience of gradience: A reply to Jackendoff. *The Linguistic Review*, *24*, 437–455.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*, 724–760.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, *6*, 465–472.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. I, pp. 3–44). Cambridge, MA: MIT Press.

McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological Science*, *2*, 387–395.

McCullogh, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.

Mehta, M. R., Lee, A. K., & Wilson, M. A. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, *417*, 741–746.

Minsky, M. L., & Papert, S. A. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Movellan, J. R., & McClelland, J. L. (2001). The Morton-Massaro law of information integration: Implications for models of perception. *Psychological Review*, *108*, 113–148.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton Century Croft.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York: Academic Press.

Newell, A. (1994). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, *103*, 381–391.

Perfors, A., Tenenbaum, J. T., & Regier, T. (2006). Poverty of the stimulus: A rational approach. In R. Sun (Ed.), *Proceedings of the 28th annual meeting of the Cognitive Science Society*, Vancouver, BC, Canada, July 26–29 (pp. 663–669). Mahwah, NJ: Erlbaum.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530–555.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73–193.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.

Popper, K. (1959). *The logic of scientific discovery*. New York: Harper.

Ranzato, M., Poultney, C., Chopra, A., & LeCun, Y. (2007). Efficient learning of sparse representations with an energy-based model. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS)* (pp. 1137–1144). Cambridge, MA: MIT Press.

van Rijn, H., van Someren, M., & van der Maas, H. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science*, *27*, 227–257.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing *Psychological Review*, *107*, 358–367.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611–653). Cambridge, MA: MIT Press.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rogers, T. T., & McClelland, J. L. (2008). Precis of *Semantic Cognition, a Parallel Distributed Processing Approach*. *Behavioral and Brain Sciences*, *31*, 689–749.

Rosenblatt, F. (1961). *Principles of neurodynamics*. Washington, DC: Spartan Books.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance* (Vol. VI, pp. 573–603). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. I, pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The context enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.

Rumelhart, D. E., & McClelland, J. L. (1985). Levels indeed! A response to Broadbent. *Journal of Experimental Psychology: General*, *114*, 193–197.

Rumelhart, D. E., & McClelland, J. L. On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. II, pp. 216–270). Cambridge, MA: MIT Press.

Schapiro, A. C., & McClelland, J. L. (in press). A connectionist model of a continuous developmental transition in the balance scale task. *Cognition*, forthcoming.

Schöner, G. (2008). Dynamical systems approaches to cognition. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 101–126). New York: Cambridge University Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Sibley, D. E., Kello, C. T., Plaut, D. C., & Elman, J. L. (2008). Large-scale modeling of wordform learning and representation. *Cognitive Science*, *32*, 741–754.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481–520.

Simon, H. (1957). *Models of man*. New York: Wiley.

Simon, H. (1991). *Models of my life*. New York: Basic Books.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11* (1), 1–23. discussion: 23–74.

Spencer, J. P., Thomas, M. S. C., & McClelland, J. L. (Eds.) (2009). *Toward a unified theory of development: connectionism and dynamic systems theory re-considered*. New York: Oxford University Press.

Steels, L. (2007). Fifty years of AI: From symbols to embodiment—and back. In M. Lungarella, F. Iida, J. Bongard, & R. Pfeifer (Eds.), *50 Years of artificial intelligence, essays dedicated to the 50th anniversary of artificial intelligence, LNAI 4850* (pp. 18–28). Berlin: Springer-Verlag.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315.

Sun, R. (2002). *Duality of the mind*. Mahwah, NJ: Erlbaum.

Sun, R. (2008). Introduction to computational cognitive modeling. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 3–19). New York: Cambridge University Press.

Taatgen, N. A., & Anderson, J. A. (2008). Constraints in cognitive architectures. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 170–185). New York: Cambridge University Press.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioural and Brain Sciences*, *24*, 1–86.

Thelen, E., & Smith, L. B. (1994). *A dynamics systems approach to the development of perception and action*. Cambridge, MA: MIT Press.

Thomas, M. S. C., & McClelland, J. L. (2008). Connectionist models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 23–58). New York: Cambridge University Press.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: the leaky competing accumulator model. *Psychological Review*, *108*, 550–592.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 13273–13278.

Wason, F. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth: Penguin.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272.