TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 2 (2010) 736–750

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2010.01092.x

# Perspectives on Modeling in Cognitive Science

Richard M. Shiffrin

*Department of Psychological and Brain Sciences, Indiana University*

Received 16 July 2009; received in revised form 2 November 2009; accepted 11 November 2009

**Abstract**

This commentary gives a personal perspective on modeling and modeling developments in cognitive science, starting in the 1950s, but focusing on the author's personal views of modeling since training in the late 1960s, and particularly focusing on advances since the official founding of the Cognitive Science Society. The range and variety of modeling approaches in use today are remarkable, and for many, bewildering. Yet to come to anything approaching adequate insights into the infinitely complex fields of mind, brain, and intelligent systems, an extremely wide array of modeling approaches is vital and necessary.

*Keywords:* Models; Perspectives; Cognitive science; Mathematical

## 1. Personal modeling history

I have been asked to give perspectives on the development of modeling in cognitive science, with focus on the developments since the founding of the society 30 years ago. I take the term *modeling* to refer to models stated precisely enough to enable quantitative predictions for data. Modeling in some form has been used in psychology since the 1800s, although early modeling focused mainly on quantitative descriptions of observed data. In the early 1950s, the field of mathematical modeling began with a major emphasis upon predictions based on unobservable internal mental processes, whose functions were represented by parameters to be estimated (e.g., Estes, 1950). I was trained at Stanford as this field moved into its heyday, by professors such as Gordon Bower (my first research advisor), Dick Atkinson (my thesis advisor), Bill Estes, Pat Suppes, and a host of other professors, postdoctoral visitors, and fellow graduate students, many of whom are leaders of the field today. In 1968, I took a position at Indiana University (where I have remained since), and in

Correspondence should be sent to Richard M. Shiffrin, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405. E-mail: shiffrin@indiana.edu

1988, I started what has turned out to be one of the largest and most successful Cognitive Science programs. I have been involved in and observed modeling in its varied forms since the time I entered the field, and I appreciate the opportunity to give some perspectives. Let me apologize to the 99% of the modeling community who deserve mention but will not be cited—this short study is not a review, although by way of example I will embed a few citations to friends and colleagues.

## 2. Modeling meta-comments

Modeling in its various forms has often been misunderstood and incorrectly characterized, because scientists have a natural tendency to act as if models are either right or wrong, engaging in a kind of scientific combat to establish truth. Of course, none of our models are ever correct, even when restricted to the simplest and most controlled experimental settings. The human mind (and brain) is a vastly complex system, every task involving uncountable factors that influence performance, including those of memory, learning, attention, emotion, decision making, and the internal and external environment. Thus, the goal of modeling is to enable progress in our understanding of this vastly complex system. Even the models that do this well are in almost all cases relatively crude and provide highly incomplete approximations to reality. That being said (and probably agreed to by all members of the field), it has proved useful to act as though models are right or wrong and carry out experimental tests that either lead to new models or refinements of old models.

Modeling performed well comes with many benefits: It makes precise what are the concepts being explored. It forces the theorist to specify details and processes that are often critical to explain performance but are sometimes overlooked in theories specified only in verbal and heuristic terms. It provides correct predictions derivable from the assumptions, especially important in cases where the predictions do not flow intuitively from the assumptions. It allows precise future testing of the model and its assumptions. It makes clear what are the key assumptions responsible for the main results of interest and what the background assumptions need to produce behavior, but are not critical for the patterns of interest. It generates interest, testing, elaborations, and alternative formulations, thereby advancing understanding and the field.

## 3. Kinds of modeling and historicity

These generalities aside, there are as many kinds of modeling as there are modelers and differing goals (so that to take just a few examples, we see probabilistic, deterministic, feedforward, recurrent, analytic, and simulation models in the hands of different theorists). The nature of modeling has evolved and changed markedly over the years (and will surely continue to do so). I will use this study to give some observations on developments, as the field has encompassed approaches from the methodological to the mechanistic, from mathematical/analytic to computational, from mental to neural, and from cognitive to embodied. I will also remark on the changes in the way models have come to be evaluated, as the focus has shifted from group data to individuals, and good fit has been augmented and superseded by more general (and sensible) criteria, including complexity, predictive capability, applicability to real-world problems, and many higher level factors. Most of the developments I will mention are large fields of research in their own right, even when I mention them with only a few sentences.

The proliferation of new modeling developments could lead one to conclude, mistakenly, that the older approaches have been replaced by the newer ones. In fact, the oldest techniques remain extremely valuable today (usually in updated formulations) but have been augmented by a plethora of valuable methods. For example, the relatively simple mathematically tractable and analytic models seen in the 1950s and 1960s (e.g., the one-element model—Bower, 1961) play a very important role today (e.g., multinomial modeling—Batchelder & Riefer, 1999). This caveat applies equally to other early modeling approaches. Measurement and scaling methodology was used to characterize necessary and sufficient conditions that would apply almost universally to models for data (e.g., Luce, 1959). This approach has remained an important part of the field today (e.g., Dzhafarov, 2008; Narens, 2007). Related to such approaches are methods to analyze data that would place constraints on large classes of models in almost universal fashion. This remains an important area today, in the hands of researchers like Jim Townsend (e.g., Townsend & Wenger, 2004), providing methods to identify parallel versus serial processing, self-terminating versus exhaustive processing, types of capacity limitations, and types of independence of processes. Multidimensional scaling methods were an important early contribution used to extract meaningful patterns from data (e.g., Carroll & Chang, 1970; Shepard, 1962a, 1962b; Torgerson, 1958). As the years have passed, such methods have become even more important as the data observed become more voluminous and higher dimensional, and the methods, particularly factor analysis, have evolved to include (to name a few) principal component analysis, independent component analysis, latent semantic analysis (based on singular-value decomposition, e.g., Landauer & Dumais, 1997), topics modeling (Griffiths, Steyvers, & Tenenbaum, 2007), and a variety of nonlinear methods advancing Shepard's early contributions to multidimensional scaling.

## 4. Models and data

In general, the best perspective to take on modeling developments is based on acknowledgment of the fact that in our field, as with all fields of science, empirical findings and models advance hand in hand together, so that each new area of empirical exploration tends to bring with it a new set of models. A reviewer of an early draft of this chapter suggested that I ought to in effect ''defend'' the use of modeling in our field. I will not do so, because I believe that the co-evolution of modeling and empirical research is inevitable, every point on the gradient from purely empirical to purely theoretical research playing its own important role. If such a defense is necessary, it would have to be the subject of another paper.

Further, the complexity of mind, brain, and intelligent systems is so high that every form of modeling known to science can and probably does prove useful in one application or another. In particular, increases in computational speed and power, and in algorithms that can operate with ever-increasing efficiency, have led to an explosion of new modeling approaches of high-dimensional complexity in order to deal with high-dimensional data. This proliferation of types of modeling makes it an exciting time to be a modeler, although it is increasingly hard to keep up. In the early days, it was not difficult for a practitioner to be trained in, and become expert in, all modeling used in the field. This is no longer the case for most practitioners, and soon it will be impossible for all.

## 5. Analytic derivations and computational derivations

When I was trained at Stanford, computers were just beginning to exhibit the capacity, speed, and power that would enable predictions based on computation rather than analysis. I recall with some ''affection'' the then state-of-the-art PDP-1 whose input and output were via punched holes in paper tape (although my fond memories may have been influenced more by the ''Space War'' game we played than its utility in experimentation and modeling). In general, computation could (and can) be used in two ways: for numerical analysis and for Monte Carlo simulation. Numerical analysis includes such approaches as matrix multiplication and inversion, fast Fourier transforms, and approximate solutions to integral equations (used, e.g., in random walk and diffusion modeling). Monte Carlo methods are typically used for probabilistic models: Each probabilistic step in an assumed process is represented by a choice made with that probability, until all the processes assumed to operate on a trial have been made and some outcome produced. Such simulated trials are accumulated in large numbers, and predictions obtained by averaging. This procedure has to be carried out separately for every condition of interest. The procedure just outlined applies for a fixed set of parameters (say, each of probabilities in the set of processes). The need for computational speed and power is magnified because a modeler usually needs to find the set of parameters that best fits a set of data or (as in Bayesian approaches) the distribution of predictions across all possible parameter combinations. Suppose there are $N$ trials needed to produce predictions for one set of $K$ fixed parameter values. $N$ can be fairly large (often on the order of 1,000 simulated trials, or more depending on the data points being simulated). Suppose we approximate each of the $K$ parameters with $M$ discrete values (perhaps 100). Then the number of parameter value combinations requiring predictions in order to map the entire parameter space would be $M^K$, and the total number of simulated trials would be $NM^K$. When the number of parameters rises, even the most powerful computer one can imagine would quickly run into trouble in the face of such a demand. Fortunately, a variety of sophisticated sampling techniques from applied statistics and machine learning are under continuous development to reduce the demand to manageable levels. These techniques can be quite effective when the parameter space and models are well behaved (e.g., linear, monotonic, etc.) For one example, the TOPICS model converges to a set of best-fitting parameter values even for as many as hundreds of thousands of parameters.

## 6. Model generation

Especially, given the contents of the preceding paragraph, it would be unsurprising if nonmodelers believe that the difficulty of modeling lies in the learning of the requisite mathematical and computational skills, and the lengthy computations needed to estimate parameters. These factors certainly play a role, but most often, the greatest length of time is spent converging on the model that will eventually be presented publically. The false leads and poor intuitions that lead the modeler down fruitless pathways are never seen in the final article, giving a misleading impression that model generation is an easy matter. In my experience, one might typically spend several weeks designing a study, collecting data, and examining the results. Then the real work begins: model development and tuning that often take many months. Difficult as this may be, it is also the most interesting and creative part of the scientific process.

## 7. Changes of model complexity

The desire to produce models having analytic solutions (because computational limitations were severe) required that early models be ''simple'' in any of several ways: Models that were deterministic, linear, feedforward, and had few parameters were simpler than models that were probabilistic, nonlinear, recurrent, and many parameters. Theorists, of course, were eager to represent in their models at least some of the complexity of the systems under investigation, and they were quick to reach the limits of then available mathematical and computational techniques. Thus, Markov models for learning allowed derivations based on linear algebra and matrix multiplication (e.g., Bjork, 1968), derivations were obtained from fairly complex probabilistic trees (e.g., Atkinson & Shiffrin, 1968), nonlinear probabilistic models for memory could be analyzed with Monte Carlo methods (e.g., Raaijmakers & Shiffrin, 1980; Shiffrin, 1970), and random walk/diffusion models for response time used state-of-the-art derivations in probability theory (e.g., Link & Heath, 1975) and numerical analysis for derivations (e.g., Ratcliff, 1978).

The fact that models were relatively simple by no means implied that the predictions were transparently related to the assumptions. An example from my own research arose from the SAM model applied to part-list cuing (Raaijmakers & Shiffrin, 1980, 1981). We showed that the model predicted the part-list cuing effect, although intuition seemed to suggest the model would do the opposite. Although the model was relatively simple in its assumptions, the facts that the model was probabilistic and nonlinear and that predictions were derived with Monte Carlo methods made it difficult to understand how the model produced its correct predictions and how it did so regardless of parametric variations and deletions. It took us far more time to understand the basis for the model's predictions than to fit the model to the data. This example is not atypical of the far more complex models in use today, and it raises an important cautionary note: Because a central goal of modeling is to further human understanding of the processes at work in a domain, a modeler should almost never be satisfied simply to demonstrate a fit to data (partial exceptions, but only partial, occur when the modeler has engineering-oriented goals aimed at solving a real-world problem). Thus, a modeler must explore the model's assumptions and parameter space sufficiently to lead to an understanding of the reasons for the predictions and enable her or him to explain that understanding to other scientists, whether or not modelers themselves. We have all seen too many examples of complex models fit to data with the fit seemingly an end in itself. Such modeling has some modest value, but it falls well short of the ideal.

As computational power continuously increased, models evolved in several ways. They were applied to more complex and higher level mental processes (e.g., sentences in the HAM model—Anderson & Bower, 1973; and a more advanced version in Anderson, 1976). They were used in recurrent modeling of complex processes, including top-down feedback loops, as in the two *Psychological Review* articles by Rumelhart and McClelland on interactive activation in letter perception (in the context of word processing and reading, McClelland and Rumelhart, 1981; Rumelhart and McClelland, 1982; of course, Grossberg used recurrent modeling in his ART neural network modeling as early as the 1960s—e.g., Grossberg, 1968, 1969—but it took some time for the field to appreciate those contributions).

## 8. Modeling advances and proliferation

### *8.1. Architectures*

As a need developed to apply models to real-world settings, some researchers decided it would be most useful to move away from highly constrained studies and models aimed at of one or another subprocess of cognition. They believed that progress in applications to real settings would require a model of the entire behavioral system from perception to motor output and everything between. Whether this in fact is the case is perhaps arguable, but these researchers developed system architectures (e.g., Soar: Newell, 1990; ACT-R: Anderson, 1973, 1993; EPIC: Meyer & Kieras, 1997a, 1997b). Of course, the need to produce an entire system architecture when certain components were not yet well understood sometimes required modeling such components with what could be described as sophisticated guessing. Furthermore, each component of a system architecture is of course only an approximation to reality; if each component is only 90% accurate, one has to wonder how much noise will be in the end product when there are 100 or more such components. Nonetheless, in the hands of capable theorists, progress was not inhibited, and some excellent successes were obtained (and are continuing to be obtained) in real-world applications. These approaches are fairly far removed from my own, because the goals of developers of system architectures are somewhat different than mine. I focus on modeling data from more tightly controlled and intentionally limited studies, and I try to build models whose core assumptions are directly testable.

### 8.2. Neural networks

Neural net modeling was and is one of the more far-reaching and important advances in modeling. It has come to be identified with the publication of the PDP handbooks (McClelland, Rumelhart, & The PDP Research Group, 1986; Rumelhart, McClelland, & The PDP Research Group, 1986). These handbooks covered much territory but are best known for the introduction of deterministic, feedforward multilayer nets with multiple nodes per layer, nonlinear transformations, and perhaps most critically, a system by which error feedback could be ''back propagated'' to adjust the connection weights to produce learning and convergence. Such networks have been applied in numbers too numerous to recount, to problems in multiple domains. Neural net modeling of course had a number of predecessor contributions, dating back to ''cell assemblies'' (Hebb, 1949), ''perceptrons'' (Rosenblatt, 1958), ART (Grossberg, 1968), composite distributed BSB models (Anderson, 1973), and animal-learning models (Rescorla & Wagner, 1972), but the general use of such modeling in the field certainly took off with the PDP publications, the researchers in the PDP group, and the workshops associated with the handbook publication.

### 8.3. Noise

Neural network models were typically deterministic as information flowed from one layer of nodes to another, but of course deterministic models cannot predict noisy data. Thus, such models incorporated a probabilistic component at the last stage of response output. In effect, the noise at all the stages of processing, from perception to storage to learning to decisions (and much more) were all collapsed into noise placed at the last stage. An example would be a Luce/Shepard choice rule, by which the response would be proportional to the ''strength'' of output of a given response node relative to the other response nodes. This same idea has been used often in process models not instantiated as neural nets—for example, the exemplar models for categorization (e.g., the context model by Medin & Schaffer, 1978, and the generalized context model developed by Nosofsky, 1984). Many models of the component processes in cognitive tasks often did specify the probabilistic noise at each component stage (my own modeling efforts fell into this camp), albeit the noise assumptions were fairly simple, and somewhat ad hoc rather than motivated directly by data. For many purposes, collapsing of the noise in processing into a single final component is a fine approximation, but for other purposes, and more fine-grained analyses and inferences, it has proved useful to model noise in detail, as it applies to stages along the processing pathways. Particularly noteworthy with respect to this issue is research by Lu and Dosher (2008), who developed a general model framework (based originally on engineering principles) that specifies on the basis of observed data the noise intrusions at various processing stages, the types of noise (i.e. or e.g., additive or multiplicative), and uses the results to reach fairly general conclusions about, for example, perceptual learning and attentional effects in perception. Of course, proper account of noise has always been a core part of modeling in the sensory sciences. One could point to hundreds of researchers (e.g., a decent starting point would be from the study of Green & Swets, 1966), but for many outstanding examples I would recommend almost any set of articles by George Sperling.

### *8.4. Recurrent modeling*

Purely feedforward neural network models were not very useful for dealing with correlated temporal and sequential data, and failed to capture what was clear from behavioral and neural knowledge: Cognitive and neural systems are recurrent, with feedback loops everywhere. Thus, models with recurrent loops quickly appeared and were applied to sequential data, such as language and speech perception (e.g., McClelland & Elman, 1986).

### *8.5. Modeling of neural activity*

The many advances in brain measurement technology (PET, fMRI, magnetic resonance imaging, EEG, etc., and in nonhumans, microelectrode recordings) have brought with them a variety of methods for modeling the results, a number of which have been borrowed from the behaviorally based modeling already described, especially those capable of handling voluminous and high-dimensional data. Thus, Sajda and colleagues (e.g., Parra et al., 2008) use a multiple regression technique to analyze EEG data (e.g., finding weights that assigned to recording sites maximize discrimination). Related techniques for EEG and FMRI analysis have been used by many researchers including Haxby, Norman, and Sederberg and colleagues at Princeton (Hanke et al., 2009). Diffusion modeling has been applied to neural decision making (e.g., Ratcliff, Philiastides, & Sajda, 2009). A great deal of modeling in sensory systems has been aimed at marrying behavior to neural activity, but there are too many citations to list. Modeling of neural and oscillatory activity in awake humans is an interesting development (e.g., Kahana and colleagues: Sederberg et al., 2007). Particularly noteworthy is large-scale modeling by John Anderson using joint modeling of behavioral data and fMRI measurements from complex tasks (e.g., Anderson et al., 2008). In recent years, it has become increasingly evident that understanding of neural activity and its relation to behavior requires more than looking at the areas that change in activity level, but requires modeling of the interactions between different brain areas—for example, whole brain or system network modeling (informed partly by DTI measurements that track the density of connections of different brain regions). A recent example is summarized by Bullmore and Sporns (2009) and represented by Honey, Kötter, Breakspear, and Sporns (2007). It is also becoming evident that one must model variability and changes in variability, in addition to the level of the neural response.

### *8.6. Machine-learning approaches*

Partly in parallel with the advances in neural net modeling and partly spurred by them, more general modeling based on applied statistical and computational approaches came on the scene. The interaction between these fields is exemplified by yearly advances highlighted in the Neural Information Processing Conference (NIPS) conference (initially organized by Terry Sejnowski). This conference initially focused on neural net models, but soon generalized coverage and featured machine learning. The aim of modeling in machine learning is often applied, solving some problem with (or with the aid of) computers. Very sophisticated algorithms were developed, and continue to evolve, to deal with high-dimensional data and classify, discriminate, identify visual or auditory stimuli, make optimal decisions, and much more along these lines (one introduction to this large field is found in Bishop, 2006). These modeling techniques with applied purposes co-evolved with related modeling techniques used to model human performance (and vice versa). In particular, as data from brain measurement devices grew exponentially, an important subcomponent of machine learning has been concerned with analyzing and modeling neural data, and the relation of neural data to behavior. In fact, a number of renowned institutes worldwide have been established to focus on this intersection (the Gatsby Computational Neuroscience Unit, and the Max Planck Institute for Biological Cybernetics provide two of many examples).

As machine-learning techniques evolve, they increasingly move into areas of more complex cognition. I have already mentioned analyses of large textual databases to draw inferences about human language (e.g., Griffiths & Steyvers, 2004; Jones & Mewhort, 2007; Landauer & Dumais, 1997). This is just the tip of the iceberg, of course. One could cite the rapidly growing field of computational linguistics and its growing entanglement with psycholinguistics.

### *8.7. Causal modeling, Bayesian modeling, Bayes Nets, graphical models*

In parallel with machine-learning modeling, another critically important form of modeling developed, which could be described with the term *causal inference*, and sometimes couched in terms of graphical models. The modeling flowed from seminal work by Pearl (1988). The essential idea is the use of probabilistic inference in the form of Bayesian analysis to characterize the way one component in a network influences others, and then using sophisticated computational algorithms and sampling techniques to characterize the causal properties of the entire network. This field gained a large impetus due to the use of (and the availability of computational software to facilitate the use of) directed causal structures called Bayes Nets. Given an appropriate probabilistic characterization, algorithmic techniques allowed one to specify the value of one or more nodes in the network and derive the distribution of the remaining nodes. The technology also allows certain types of similar derivations for undirected networks (see Boltzmann machines), but these have seen somewhat less used thus far. The uses of Bayesian inference and modeling have spread too widely to summarize (the next section will mention important recent developments in symbolic Bayesian approaches to cognition), but I want to mention in particular the successful use of Bayesian modeling in the sensory sciences. There are some particularly nice examples from vision science (e.g., Geisler, 1984; Yuille & Kersten, 2006; and using Bayesian Ideal Observer Theory, Geisler & Albrecht, 2000; Geisler, Perry, Super, & Gallogly, 2001).

### 8.8. Hierarchical and symbolic Bayesian models

Using Bayesian techniques, a most interesting and promising form of modeling has occurred in recent years (somewhat in contrast with neural net approaches). It pursues inference in more symbolic form, using hierarchical Bayesian approaches (e.g., Shiffrin et al., 2008). Hierarchical modeling can generally be described as models in which some parameterized processes in a model are themselves determined probabilistically by other parameterized processes. Thus, to give one example, different participants may each have some different (multidimensional) parameters determining their performance. Rather than treat these participants independently, a hierarchical model might assume the parameters for a given participant are sampled from an assumed distribution, perhaps Gaussian with some mean and variance. The higher level parameters are usually described as hyperparameters, but all parameters are typically estimated simultaneously from the data sets for all participants. A particularly intriguing use of hierarchical Bayesian modeling involves specifying a series of levels of increasingly abstract representations, with the highest level of abstraction specifying the form of the representation to be used (e.g., linear order, tree, ring, etc.). This work is well presented by Kemp and Tenenbaum (2008, for structure inference) and Kemp, Perfors, and Tenenbaum (2007, for language development and inference).

### 8.9. Quantum probability

Bayesian analysis is couched in terms of traditional probability theory, but it is not an absolute requirement that our thought processes and behavior obey such laws. Indeed Kahneman received a Nobel Prize in economics largely for demonstrations that human decision making does not match the dictates of rationality defined by traditional probabilistic inference. Thus, very recently, we have seen a broadening of the axioms of probability to allow different forms of inference. I want to mention particularly the use of the (more general) axioms of quantum probability (e.g., Busemeyer, Wang, & Townsend, 2006; Pothos & Busemeyer, 2009).

### 8.10. Embodied cognition, complex, and dynamical systems

The recognition that brain and mind are organized in complex recurrent systems, and especially systems that involve mind, brain, body, and environment (e.g., Beer, in press), has led to modeling from a more global perspective, sometimes termed complex systems analysis, and often utilizing dynamical systems modeling. In some ways these developments represent a theoretical partner to the growing empirical field termed ''embodied cognition.'' Large-scale embodied dynamical systems exist in very high-dimensional spaces, and progress typically depends on positing (or evolving) systems with natural limit points (with phase transitions between them) that define the operative modes of the system.

### 8.11. Networks

Much of the modeling described thus far has been aimed at individual neural activity and behavior, but a great deal of important modeling is aimed at group and social behavior, especially in networked organizational form (just to take examples close to home, at Indiana University the chair of statistics, Stan Wasserman is a leader in studies and modeling of social networks, and we have a Center for Complex Networks and Systems Research, with faculty such as Randy Beer, Alessandro Vespignani, Filippo Menczer, Luis Rocha, Peter Todd Mattheus Scheutz, and others). In recent years, much social networking occurs on the World Wide Web, and it is therefore unsurprising that much of the modeling efforts are directed at the analysis of Web traffic.

### 8.12. Model selection

I would be remiss in not mentioning another important meta-modeling development: advances in the way that models are evaluated and preferred, termed *model selection*. The statistical approaches to this problem are particularly useful in the many cases where the data are noisy and limited in amount, and where the competing models differ in complexity. Then the model selection methods try to find a best balance of fit and complexity, and indirectly try to maximize the ability to predict future data from the same or similar paradigm. The chief methods are Bayesian model selection and minimum description length. A nice review (albeit with a bias toward MDL) is found in Grünwald (2007). Much research relevant for cognitive science has been carried out by Jay Myung, Mark Pitt, and colleagues (e.g., Pitt, Kim, Navarro, & Myung, 2006; Pitt, Myung, Montenegro, & Pooley, 2008).

### 8.13. Perspectives and predictions

This broad (and nonetheless selective) listing of modeling developments may strike the reader as bewildering in its variety and scope. Although most readers may regard this following comment as obvious, I will nevertheless emphasize that mind, brain, and intelligent systems are far too complex for any one approach to monopolize progress. It is essential that modeling be utilized at multiple levels of abstraction, using any and every approach from every field, and incorporating new ones as they are developed. ''Let a million flowers bloom'' is far more than a catch phrase in modeling in cognitive science. Of course, some researchers in every field and subfield (including those who model) do so well, and others do so poorly. This catch phrase is not ''Let a million weeds proliferate.'' As much fun as it might be for me to list particularly egregious uses of modeling (or for that matter of empirical research that would have benefitted from modeling), every field should be judged by their best practitioners, not their worst.

Looking to the future and predicting developments is usually a guesswork revealing more about personal biases than informed inference: Prognosticators almost always predict that the future will bring more of whatever they are doing at the moment and whatever they would like to see funded (particularly their own research). It is relatively easy to extrapolate into the near future from current trends, but virtually impossible to predict the presently unknown. Perhaps a few meta-predictions can be made. Because empirical findings and theory develop synergistically, we can expect models to move into new territory when new types of data demand this. New types of data are often driven by new technology. In our field as with many others new technology often takes the form of new tools of measurement. I do not ''predict'' this will occur, but an explosion of new research and modeling would come about should we have new technology allowing noninvasive methods of measuring neural function in extremely small regions of brain tissue with extremely accurate timing. The other main source of new model approaches is developments in other fields, particularly applied statistics and computer science because these are a good source of computational methods for noisy high-dimensional data. On a less meta-theoretical front, it seems reasonable to expect the growth of research in large-scale models that try to encompass different levels of analysis and their interactions, from the chemical/neural to social/environmental. However, this will not be a highly populated field, due to the difficulty of such large-scale modeling (I am reminded of the large-scale environmental modeling efforts by the Club of Rome). In the popular literature, it has become common to see prognostications of the development of machine intelligence that will ''overtake and pass'' humans, and do so in the near future. I am somewhat more pessimistic regarding the imminence of such an occurrence, based on what I judge to be a very large gap between even the most advanced models today and the ''true'' complex system that instantiates our minds and brains in a real environment.

## References


Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review*, *1973*, 80.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston and Sons.

Anderson, J. R., Carter, C. S., Fincham, J. M., Qin, Y., Ravizza, S. M., & Rosenberg-Lee, M. (2008). Using fMRI to test models of complex cognition. *Cognitive Science*, *32*, 1323–1348.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86.

Beer, R. D. (in press). Dynamical systems and embedded cognition. To appear in K. Frankish & W. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence*. Cambridge, England: Cambridge University Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Bjork, R. A. (1968). All-or-none subprocesses in the learning of complex sequences. *Journal of Mathematical Psychology*, *5*, 182–195.

Bower, G. (1961). Application of a model to paired-associate learning. *Psychometrica*, *26*, 3.

Bullmore, E. T., & Sporns, O. (2009). Complex brain networks: Graph-theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*, 186–198.

Busemeyer, J. R., Wang, Z., & Townsend, J. T. (2006). Quantum dynamics of human decision making. *Journal of Mathematical Psychology*, *50*, 220–241.

Carroll, J., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of the ''Eckhard-Young'' composition. *Psychometrika*, *35*, 283–319.

Dzhafarov, E. N. (2008). Dissimilarity cumulation theory in smoothly-connected spaces. *Journal of Mathematical Psychology*, *52*, 93–115.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*, 2.

Geisler, W. S. (1984). The physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A*, *1*, 775–782.

Geisler, W. S., & Albrecht, D. G. (2000). *Spatial vision: Single neurons and perception. Handbook of cognition and perception*. New York: Academic Press.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*, 711–724.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Griffiths, T., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(Suppl. 1), 5228–5235.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244.

Grossberg, S. (1968). Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *59*, 368–372.

Grossberg, S. (1969). Embedding fields: A theory of learning with physiological implications. *Journal of Mathematical Psychology*, *6*, 209–239.

Grünwald, P. D. (2007). *The minimum description length principle*. Cambridge, MA: MIT Press.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Olivetti, E., Frund, I., Rieger, J. W., Herrmann, C. S., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A unifying approach to the analysis of neuroscientific data. *Frontiers in Neuroinformatics*, *3*, 3.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Honey, C. J., Kötter, R., Breakspear, M., & Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 10240–10245.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1–37.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10687–10692.

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*(2), 211–240.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*, 77–105.

Lu, Z.-L., & Dosher, B. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review*, *115*(1), 44–82.

Luce, D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

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

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., & The PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. II). Cambridge, MA: MIT Press.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review*, *85*, 207–238.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, *104*, 3–65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review*, *104*, 749–791.

Narens, L. (2007). *Introduction to the theories of measurement and meaningfulness and the use of invariance in science*. Mahwah, NJ: Lawrence Erlbaum Associates.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Cambridge University Press.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104–114.

Parra, L. C., Christoforou, C., Gerson, A. D., Dyrholm, M., Luo, A., Wagner, M., Philiastides, M. G., & Sajda, P. (2008). Spatio-temporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks. *IEEE Signal Processing Magazine*, *25*(1), 95–115.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufmann.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, *113*, 57–83.

Pitt, M. A., Myung, J. I., Montenegro, M., & Pooley, J. (2008). Measuring model flexibility with parameter space partitioning: An introduction and application example. *Cogntive Science*, *32*, 1285–1303.

Pothos, E. M., & Busemeyer, J. R. (2009). Quantum probability explanation for violations of 'rational' decision theory. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *276*(1165), 2171–2178.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6539–6544.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64–99). New York: Appleton-Century-Crofts.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.

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., & the PDP research group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. I). Cambridge, MA: MIT Press.

Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., Tully, M. S., & Kahana, M. J. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex*, *17*, 1190–1196.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*, 125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function II. *Psychometrika*, *27*, 219–246.

Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of memory* (pp. 375–447). New York: Academic Press.

Shiffrin, R. M., Lee, M. D., Wagenmakers, E. J., & Kim, W. J. (2008). A survey of model evaluation approaches with a focus on hierarchical Bayesian methods. *Cognitive Science*, *32*(8), 1248–1284.

Torgerson, W. S. (1958). *Theory & methods of scaling*. New York: Wiley.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.

Yuille, A. L., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Neuroscience*, *10*(7), 301–308.