In L. Nadel (Ed.), *Encyclopedia of Cognitive Science,* (pp. 611–615). Macmillan Reference Ltd., London. 2003.

Constraining Computational Models of Cognition

Terry Regier
Department of Psychology
University of Chicago

Definition: Computational models of cognition may be constrained in various ways. The three primary benefits of constraining a computational model are (i) parsimony, (ii) avoiding overflexibility, and (iii) potentially stronger motivation for the elements of the model.

Keywords: computational models; constraints; overfitting; motivation.

Contents:

Overview

Computational models of cognition are a critical tool in the study of the mind. One strength of such models is their explicitness. The computational researcher must specify a mental mechanism in sufficient detail to allow the resulting model to be instantiated on a computer, and run as a cognitive simulation. This often requires that theoretically important elements of the model be spelled out very clearly: the theory must be explicit enough to be described as a machine.

But what sort of machine? Computational machines range from the inflexible to the very flexible. An inflexible machine performs only a single predetermined task, such as addition, or multiplication, or accepting 65 cents in coins and returning a can of soda. A very flexible or general machine, in contrast, may be used for a wide variety of functions. Significantly, there exists an abstract machine, the Turing machine, that is so general that it has been taken as the formal equivalent of the very broad informal notion of an algorithm (Lewis and Papadimitriou, 1981); thus, a Turing machine is taken to be capable of implementing any algorithm at all. Modern programmable computers are approximations to Turing machines. Thus, the flexibility of modern computers provides an informal sense for the generality of Turing machines, and for the potential generality of computational machines as a class.

What is the significance of these issues of computational generality and flexibility for models of cognition? Often, generality is taken as a strength in a scientific theory. One would like to be able to account for a broad array of data, from a variety of sources, rather than provide a micro-theory of a very limited domain of cognition. There is, however, a danger of *over*-generality in model construction. If data are accounted for by a model that is very general and flexible, it is not always clear what to make of the model's success. The problem is that such a general model could also

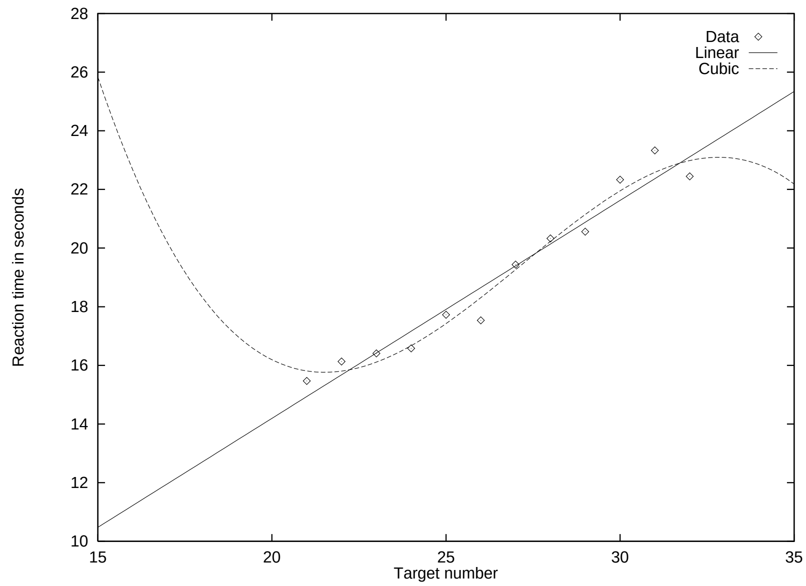

Figure 1: Mental counting

---

perhaps have fit other data, of a sort never empirically observed. In a word, it may have succeeded in accounting for the empirical data because of its extreme flexibility, rather than because of a systematic match between the structures of the model and the mental processes at play. In this sense, a model's computational power may undermine its explanatory power. A more constrained, less computationally powerful model may in some instances provide more insight into the mental processes under study.

The Benefits of Constraints

There are three primary benefits of a constrained model. These are a more *parsimonious* characterization, an avoidance of *overflexibility*, and potentially stronger *motivation* for the elements of the model. To illustrate these benefits, let us turn to a simple example.

Consider the task of silently counting from zero up to a given positive integer, and then pressing a button when done. This task taps the mental process of counting, with no overt behavioral cues to the nature of the process other than the total time required. This time is shown as a function of the target number in Figure 1. These are informally collected data, meant for illustrative purposes only. The data represent averages over five testing sessions, from one subject. The figure also shows the fits of two models, one more constrained than the other.

The obvious computational characterization of mental counting is a simple iterative loop: begin with zero, then add 1 repeatedly until the target number is reached. We may reasonably assume that the counting will assume a rhythmic character, such that each iteration requires the same amount of time. Thus, overall reaction time should be a linear function of the target number:

$$\text{Linear model: } RT = aX + b$$

Here RT denotes reaction time in seconds, X denotes the target number, and a and b are free parameters. Note that the parameter a has a straightforward psychological interpretation: the length in seconds of a single counting iteration. Similarly, b denotes the amount of time needed to count to from zero to zero – ideally this amount would itself be zero. Fitting this linear model to these data, we obtain a good fit, as shown in the figure ($R^2 = 0.9504$, $p < 0.0001$).

Let us compare this with the performance of a less constrained model, one based on a cubic function of the target number X:

$$\text{Cubic model: } RT = a_1 X + a_2 X^2 + a_3 X^3 + b$$

This more general model provides a somewhat closer fit to the data ($R^2 = 0.9713$, $p < 0.0001$). The two new terms in the model approach significance. However, the comparison between the two models also highlights the three benefits of more constrained models.

Perhaps the most obvious benefit is that of *parsimony*. The linear model is simpler than the cubic model, which contains two terms in addition to those of the linear model. Model simplicity is aesthetically pleasing, and also grants the model a greater ease of analyzability.

A related issue is that the constrained model is less likely to be *overly flexible*. In this example, the less constrained cubic model overfits the data. With the obtained parameter settings, the cubic model predicts the observed data very well. But in doing so it also fits some of the noise, such that it generalizes poorly outside the observed range. As can be seen in the figure, this model predicts that it will take longer to count to 15 than to 30 – an obviously false prediction. This problem is avoided by the more constrained linear model.

Finally, as we have seen, in this case the elements of the constrained linear model are *motivated*: they readily admit clear psychological interpretations, in the context of an iterative counting process. This is not true of the additional terms in the cubic model. The addition of such elements to a model, in the hope of increasing flexibility and thereby improving the fit to the data, comes at the price of arbitrariness.

These various benefits of constrained models need not all correlate perfectly. One could imagine, for instance, an ill-motivated model with few free parameters, or a well-motivated one with many. But the appeal of a constrained model lies in the prospect of accurately explaining a psychological phenomenon using a small number of well-motivated components.

These issues touch a range of different styles of cognitive modeling, not just the simple models used in the example. Concerns over unconstrained models have been voiced in the context of symbolic models, and also in the context of connectionist or neural network models. For example, Miller et al. (1960) advanced a symbolic mental structure known as a plan, which might be recursively created by other plans, yielding an overall system that was potentially quite complex and flexible. They anticipated that some might find their proposal too general and open-ended, and imagined their critics arguing as follows:

> "A good scientist can draw an elephant with three parameters, and with four he can tie a knot in its tail. There must be hundreds of parameters floating around in this kind of theory and nobody will ever be able to untangle them." (quoted in Seidenberg, 1993).

While Miller et al. (1960) defended their proposal against this critique, on the grounds that a complex system may in some instances be required, their sensitivity to the issue highlights an early concern with underconstrainedness in symbolic models. Another manifestation of the same concern may be found in Newell's (1990) discussion of the computational universality of his production system framework Soar, and the need for constraints in such a framework (p. 220).

Some connectionist models have evoked similar concerns. Massaro (1988) charged that the multi-layer perceptron – a commonly-used connectionist architecture – was too computationally powerful to be psychologically meaningful. This argument was based on the finding that a single connectionist model could simulate results generated by three mutually exclusive process models. Thus, the connectionist model appeared to Massaro to be overly flexible, and potentially unfalsifiable. And Cybenko (1989) showed formally that multi-layer perceptrons are in principle flexible enough to approximate arbitrarily well any continuous function over inputs that range from 0-1. Further discussion of constraints in connectionist models may be found in (McCloskey, 1991; Regier, 1996; Seidenberg, 1993; Siegelmann, 1995). We shall also return to some of these issues below.

The Source of Constraints

There are many possible sources for constraints in a cognitive model. Four particularly important ones are treated here: constraints derived from known *psychological* structure, those derived from known *biological* structure, those derived from *task* structure, and those based on the nature of the *input* to the model.

Psychological structure: It is natural for a computational model of one psychological phenomenon to be informed by – and constrained by – independent observations concerning another, related phenomenon. This would afford an explanation of the one mental process in terms of the other. For example, Gluck and Bower (1988) accounted for aspects of human categorization using a simple connectionist learning rule, the delta rule, which is ultimately motivated by studies of Pavlovian conditioning in animals (Rescorla and Wagner, 1972). This rule has known constraints; for example, it predicts blocking and overshadowing in learning. The strength of Gluck and Bower's presentation is that they find empirical evidence of these constraints in human category learning – thus mechanistically linking human categorization and animal conditioning. Another example of the use of psychological constraints may be found in Nosofsky's (1986) model of categorization. This model builds on an existing model of identification, or the discrimination of stimuli from one another (Shepard, 1957). Thus, the constraints of the original model – such as a monotonic dropoff in generalization with increasing psychological distance – are incorporated into its successor. This implicitly grounds an account of categorization in an existing account of identification.

Biological structure: Biological constraints may also be brought to bear on cognitive models. This idea holds the potential of reduction, of explaining cognitive processes in terms of the neural structure that underlies them. Wilson et al. (2000) provide a concrete example. They present a model of how humans perceive the orientation of another person's head: whether the other person is facing one straight on, or off to the side to some extent. They are able to account for their empirical findings in this domain using a model based on a population code of neurons found in area V4 of cortex, neurons sensitive to concentric and radial visual structure. Thus, the constraints of the underlying neural structure are used to explain a psychological phenomenon. Another example can be found in Regier and Carlson's (2001) study of spatial language. Participants were asked to rate the acceptability of linguistic spatial descriptions such as "The dot is above the triangle", when shown pairs of objects in various spatial configurations. These linguistic responses were well-described by a model based in part on a neurobiological finding: the representation of overall direction as a vector sum. Thus, the linguistic data are partially grounded in a neurobiological constraint. More generally, one of the major appeals of connectionism as a modeling framework has been the prospect of bringing neural constraints to bear on psychological models (Feldman and Ballard, 1982; Seidenberg, 1993).

An important source of biological constraint is *timing* (Feldman and Ballard, 1982). Complex mentally-guided behavior can occur at a timescale of seconds. Assume, for example, that someone were to ask "Do you know where the registrar's office is?" It would take only a few seconds to hear the question, extract the intended meaning, recall where you believe the office is, prepare a linguistic response to the question, and deliver that response. Thus, a good amount of cognitive computational work is accomplished in a short timespan. However neurons, which are widely assumed to be the ultimate implementation of this computational work, operate on a timescale of a few milliseconds – quite slowly by the standards of modern computers. Thus, the entire process of hearing and responding to the question must take place within only a few thousand neural computational time steps. This constrains the number of iterations a model may realistically take when accomplishing such a behavioral or cognitive task. Since very little can be computed in a few thousand time steps using serial computation, this constraint strongly motivates parallel computation in cognitive models.

Task structure: The nature of the psychological or behavioral task under study may also provide constraints on potential models. An example is the simple motor task of moving one's hand (or a pointer) to a target area of a specific size. Empirically, it has been found that the time required for this task varies as $log(D/S)$, where D is the distance from starting point to end point of the motion, and S is the size of the target region. This regularity is known as Fitts' Law. An elegant model explaining this law has been given, one based on the nature of the behavioral task. On this model, the most natural solution to the task is to launch the hand in the right general direction, and then iteratively correct during movement so as to bring the hand closer and closer to the target. This defines a recursion which, when solved analytically, yields the formula for Fitts' Law (Keele, 1968; Meyer et al., 1988; Newell, 1990). Another example is the simple mental counting task described earlier. Here, it is the intuitively obvious iterative nature of the counting task that motivates the linear model, the more constrained of the two models considered.

Input: An important source of constraint lying outside the model itself is the nature of the input supplied to it. Researchers investigating very flexible or general models often stress the constraints that reside in the model's input, at least as much as those residing in the structure of the model itself. Much connectionist modeling has this flavor. Specifically, as we have seen, multi-layer perceptrons are computationally quite general, and this generality has been a point of criticism. However, flexible mechanisms of this sort can be very scientifically useful, when considered together with the nature of the model input. A concrete example may be found in language acquisition. For many years, the study of language acquisition was dominated by the "argument from poverty of the stimulus" (Chomsky, 1986). This view holds that the linguistic input heard by the language-learning child is too sparse, too impoverished to eventually give the child full knowledge of the syntactic structure of the language. Therefore, on this account, some elements of this knowledge must be innate. This account predicts that general purpose learning mechanisms should not be able to learn the syntactic structure of natural language – as these general-purpose mechanisms lack the requisite language-specific innate structure. This stance has been challenged recently. Very flexible, general purpose connectionist networks have succeeded in learning artificial languages that resemble natural language in some important respects (Elman, 1993; Rohde and Plaut, 1999). This suggests that the input may not be as impoverished as had earlier been asserted, and that there may be no need to posit innate language-specific structure. Thus, the importance of these demonstrations lies precisely in the *un*constrainedness of the mechanism itself, and the substantial constraints present in the input.

Given this variety of possible sources of model constraints, is there a most appropriate source? Is there any reason to prefer a model constrained in one manner over a model constrained in another? The answer to this question ultimately lies in the nature of the scientific question being asked. A general mechanism with constrained input is useful in addressing the alleged necessity of innate structure. For other sorts of questions, however, it may be more informative to demonstrate that elements of already acknowledged mechanistic structure can explain a novel phenomenon, one to which they were not originally tied. In both cases, the data at hand is explained in terms of known and independently motivated constraints, whether these constraints reside in the mechanism itself, or in environmental input.

Summary

Overgenerality is a potential danger in computational models of cognition. It is conceivable that a model may account well for a set of empirical data because of its extreme flexibility, rather than because of a clear match between its structures and those of the cognitive process under study. This possibility has been a concern in both symbolic and connectionist models of cognition. However, the problem can be avoided by constraining models in principled ways. This can yield a more parsimonious, less overflexible, and more convincingly motivated model. There are several possible sources for constraints on cognitive models: existing knowledge concerning psychological or biological structure, the nature of the cognitive task itself, or the input to the model. The scientific question being posed will indicate the most appropriate source of constraint in a given modeling enterprise.

References

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use.* Praeger, New York.

Cybenko, G. (1989). Approximations by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2:303–314. also available as report number 856, Center for Supercomputing Research and Development, University of Illinois at Urbana-Champaign, 1989.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48:71–99.

Feldman, J. and Ballard, D. (1982). Connectionist models and their properties. *Cognitive Science*, 6:205–254.

Gluck, M. and Bower, G. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117(3):227–247.

Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, 70:387–403.

Lewis, H. R. and Papadimitriou, C. H. (1981). *Elements of the Theory of Computation.* Prentice-Hall, Englewood Cliffs, NJ.

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

McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological Science*, 2(6):387–395.

Meyer, D., Abrams, R., Kornblum, S., and Wright, C. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, 95(3):340–370.

Miller, G., Galanter, E., and Pribram, K. (1960). *Plans and the Structure of Behavior.* Holt, Rinehart, and Winston, New York.

Newell, A. (1990). *Unified Theories of Cognition.* Harvard University Press, Cambridge, MA.

Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1):39–57.

Regier, T. (1996). *The Human Semantic Potential: Spatial Language and Constrained Connectionism.* MIT Press, Cambridge, MA.

Regier, T. and Carlson, L. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, 130(2):273–298.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In Black, A. H. and Prokasy, W. F., editors, *Classical Conditioning II: Current Research and Theory*, New York. Appleton-Century-Crofts.

Rohde, D. and Plaut, D. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, 72(1):67–109.

Seidenberg, M. (1993). Connectionist models and cognitive theory. *Psychological Science*, 4(4):228–235.

Shepard, R. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22:325–345.

Siegelmann, H. T. (1995). Computation beyond the Turing limit. *Science*, 268:545–548.

Wilson, H. R., Wilkinson, F., Lin, L.-M., and Castillo, M. (2000). Perception of head orientation. *Vision Research*, 40:459–472.