The Role of Computational Models in Investigating Typical and Pathological Behaviors

Prahlad Gupta, Ph.D.[1]

ABSTRACT

The use of computational models to simulate unimpaired human psychological behavior is now fairly common, and use of such models to simulate impaired behavior has also increased. In this article I discuss the relation of computational models to behavioral investigation and to theory, with a view to clarifying what a computational model is, and what its value may be in investigating unimpaired and pathological human psychological behavior.

KEYWORDS: Computational models, simulation, cognitive behavior, language pathology, multiple determination

Learning Outcomes: As a result of this activity, the reader will be able to (1) explain the value of using a computational model of language or other cognitive abilities as a tool to understand the organization of the ability, and (2) explain how a computational model can be used to investigate unimpaired or pathological cognitive behavior.

Computational models are now fairly commonly used as a means of investigating human psychological behaviors. A computational model is constructed, it is demonstrated that the model succeeds in simulating certain behaviors in the domain of interest, and then the investigators suggest that the model offers a good account of the cognitive mechanisms underlying that particular human performance. (For example, a computational model that yields simulated processing times for reading *garden path* sentences such as "The horse raced past the barn fell" might be constructed and show a good correspondence of its simulated processing times at different parts of the sentence with human eye-fixation durations for reading the same sentences.) It is not uncommon in such work for some aspect of the model to be manipulated, and for the effect of the manipulation on the model's output to be viewed as a simulation of a pathology in that behavioral domain. (For example, in the hypothetical computational model just cited, the functioning of some processing element

[1]Department of Psychology, University of Iowa, Iowa City, Iowa.

Address for correspondence and reprint requests: Prahlad Gupta, Ph.D., Department of Psychology, University of Iowa, Iowa City, IA 52242 (e-mail: prahlad-gupta@uiowa.edu).

Dynamic Interactions of Language with Other Cognitive Processes; Guest Editor, Nadine Martin, Ph.D.

Semin Speech Lang 2008;29:211–225.
DOI 10.1055/s-0028-1082885. ISSN 0734-0478.

might be deliberately "impaired," and it might be shown that the model's simulated processing times under such conditions show a good correspondence with eye-fixation durations of individuals with Broca's aphasia when they read garden path sentences. It might then be proposed that the impairment in the model is suggestive of an underlying impairment in Broca's aphasia.)

Such use of computational models is regarded as a powerful tool for psychological investigation. However, two general questions about this approach should be answered. First, what is the value of demonstrating that a particular computational model can simulate a particular set of behaviors? Second, what implications do such demonstrations have for pathological conditions? Although most researchers do not ignore these points, they are usually tied to the specifics of the particular model, often somewhat implicitly.

In this article, I discuss these questions both explicitly and generally. I use examples from my own work as illustrations. My goal is not to argue for or defend a particular model but rather simply to use it as a means of illustrating broader points about computational models in general.

THE ESSENCE OF A COMPUTATIONAL MODEL

How do computational models differ from more typical behavioral investigation? Suppose an investigator hypothesizes that the phonological similarity of the words in a sentence affects how accurately the sentence can be comprehended. If this is an accurate hypothesis, then it would suggest that the cognitive processing mechanisms that underlie sentence comprehension use a representational code that includes phonological information.

To investigate this hypothesis behaviorally, the researcher might construct two sets of sentences. In one set, words in each sentence would have high phonological similarity to each other (Set A). In the other set (Set B), words in each sentence would have low phonological similarity to each other. Subjects' comprehension would be assessed for the two sets of sentences. If comprehension was found to be significantly better for Set A, this would constitute evidence for the original hypothesis, assuming there were no other differences in the experimental procedure that could explain the results. This result would support the idea that sentence comprehension processes are based, at least in part, on a phonological encoding. Note that the focus would be on examining the effect of an independent variable (within-sentence phonological similarity) on a dependent variable (the measure of comprehension performance), and on attempting to control all sources of variability other than the independent variable.

To investigate the same hypothesis through computational modeling, the researcher would construct a computational model that simulates sentence comprehension, with the aim of examining whether simulated sentence comprehension in this model is affected by the within-sentence phonological similarity of words. Before actually examining the effect of phonological similarity, the researcher would seek to demonstrate that the model provides a plausible account of sentence comprehension, not just with respect to any effect of phonological similarity, but *in general*, by comparing its performance with that of humans under some selected variety of conditions *other* than variation in phonological similarity. This is the first aspect of a computational model, as follows:

Aspect 1 To argue for its plausibility, the model's output (i.e., behavior) is typically compared with that of humans, across a range of situations in the psychological domain of interest.

Once the model's plausibility is established in this manner, the researcher next examines whether the effect of within-sentence phonological similarity on the model's comprehension performance matches the effect of phonological similarity on human sentence comprehension performance. That is, the researcher examines whether the particular independent variable of interest (within-sentence phonological similarity) has the same effect on the dependent measure (sentence comprehension performance) in the model as it has been demonstrated to have in humans. This is a second characteristic that is commonly an important aspect of a computational model. (I use the qualifier "commonly" because it is possible for a computational investigation to be aimed at demonstrating plausibility of the model across a range of behaviors, without special interest in a particular independent variable. That is, the goal might be encompassed by Aspect 1.) In the present example, however, there would also be a second aspect of the model:

Aspect 2 (Often) Compares the effect of a specific independent variable on a model's behavior with the effect of that independent variable on human behavior.

The ability to make the comparisons of model behavior with human behavior in Aspects 1 and 2 depends on data from human behavioral experiments. Thus computational investigation cannot replace behavioral investigation as a methodology for studying human behavior; rather, it supplements behavioral investigation. However, it is important to consider carefully the nature of this "supplement."

The supplement is that the researcher constructs a computational processing model of the behavioral task from which the measure(s) of interest is/are obtained in humans. That is, the researcher constructs a theory of processing in the behavioral task. This theory must be very highly specified: This is what makes it "computational." It is important to clarify here what the term "computational" means. Because the use of computational models to investigate cognition is so strongly correlated with the use of computers, the involvement of a computer is often perceived as the defining aspect of a "computational" model of cognitive behavior. That is, "computational model" is often thought to have an intrinsic and essential connection to computers, perhaps also incorporating the "computer metaphor," which likens the human mind to a conventional, serial-processing computer (often termed a "von Neumann computer"). However, the term "computational" actually does *not mean* either "involving the use of a computer" or "involving the computer metaphor." It means simply that the account (i.e., theory) of performance in the task is specified well enough that the theory's prediction of what the behavioral response (i.e., the dependent measure) will be in a given situation can be determined numerically. That is, it can be precisely *calculated*. That is, it can be *computed*.[1] A *simulation* is the process of actually *making* these calculations for a particular set of inputs/circumstances. However, whether or not an abacus or calculator or computer is needed to actually make the calculation/computation is a separate matter, and it is not what makes the theory or model or simulation "computational." The behavior predicted by a computational theory for a particular case might be determined or derived (i.e., simulated) by making calculations on the back of an envelope, and the account would still be a "computational" theory or account or model, and the simulation would still be a computational one.

In practice, the calculations required to derive the behavioral prediction of a computational model (i.e., to conduct a simulation) would usually take impossibly long to conduct by hand. Therefore, a speedier calculating device is used, such as a computer. This is why the use of "computational models" has become prevalent only since the advent of computers. This is also why the use of computational models is so highly associated with the use of computers and why a "computational model" is frequently taken to be something whose essence (and possibly theoretical basis) is that it is run on a computer. But this is not the case, for the reasons just noted. In fact, for these same reasons, the frequently used term "computer model" is a misnomer. There are "computational models/theories" and there are "computational simulations." If the computational simulation is conducted on a computer, it is additionally a "computer simulation." But, as just discussed, the use of a computer is not what makes a model or theory or simulation "com-

[1] *It may be worth noting that a mathematical model is a particular type of computational model, in which the behavior of the model can be fully described in terms of an equation or set of equations. This property is not typically true of other types of computational models. There are also certain other typical differences between mathematical and other computational models, but discussion of these is beyond the scope of the present article.*

putational." Thus when a model is termed a "computer model," it should more correctly be called a "computational model."

Although "computational specification" has nothing to do with the use of a computer or any other calculating device, it nevertheless constitutes a crucial point of difference between the computational and behavioral investigations. In the latter, there may be a prediction that the independent variable will affect the dependent variable in a particular way—for instance, that greater within-sentence phonological similarity will lead to a lower sentence comprehension score. This prediction is based on theoretical ideas, so that the behavioral investigation is based, more or less explicitly, on a theory. However, this theory does not constitute a computational account: It does not provide a basis for determining numerically what the behavioral performance will be under each condition. Although it may provide quantification on an ordinal scale (comprehension performance is predicted to be greater in the low than the high phonological similarity condition), it does not provide quantification on an interval or ratio scale. Thus in the behavioral investigation, there is no computationally specified account of performance in the task from which the dependent measure is obtained. In terms of the present example, there would be no computational model of sentence comprehension.

This difference has some important consequences. To construct a computational model of performance in the task (in the present example, sentence comprehension), it is necessary to consider what factors may be expected to determine performance in the task. For any except the simplest human behaviors, multiple factors determine performance. For the task in which sentence comprehension is measured, a computational model necessarily has to provide some account of how words are represented, how their meanings are represented, and how the overall meaning of the sentence is derived. This account must be sufficiently well specified as to be computational, as defined earlier. Creating such an account entails consideration of (and computational specification of) a great many variables that enter into and determine performance (e.g., the state of existing syntactic and semantic information in the model, processing speed, and some account of working memory). A crucial aspect is that the account has to specify computationally *how these factors work together to determine task performance.* That is, it must consider and specify the joint and separate impact of multiple variables on task performance. This is true even for a much simpler behavioral task, such as phoneme identification, for which a computational account would provide a computational specification of how the input acoustic signal is represented, what processes operate on it, and how a categorization is made.

Thus, in general, virtually every computational investigation must take into account many cognitive variables in constructing the computational model of task performance, and crucially, the taking into account must take the form of *computationally specifying* how they determine task performance, severally and jointly. In contrast, in a behavioral investigation of the same cognitive phenomenon, it is crucial to be sure that the situation under which the sentence processing task is undertaken varies *only* with respect to the independent variable(s). The many other variables that determine sentence processing performance, and the relationship of these variables to each other, are not the focus of interest in a particular behavioral investigation, although several of them may be examined systematically in a successive set of investigations. But computational specification of the relationship of these variables to each other in determining task performance is not inherently necessitated by the nature of the investigation. The important consequence of this difference is that the computational investigation is capable of examining the manner in which behavioral performance in the task of interest is influenced by the *simultaneous* effect of multiple factors. That is, the computational investigation encourages, if not forces, us to recognize that the behavior under consideration is *multiply determined* (I have adopted this term from the work of Gathercole[1]). This is much less highlighted in or by experimental investigation.

This difference between behavioral and computational investigations is particularly relevant for the investigation of pathology. There are multiple ways in which the model's performance can be impaired. Thus pathological performance in the task could be caused by impairment of any one of, or a combination of, many different underlying determinants. This point is sufficiently important that it bears some elaboration. A computational model typically embodies an account of multiple determinants of the simulated behavior, in a manner that behavioral investigation typically does not. In addition, a computational model affords the opportunity to *impair* the functioning of each of these multiple determinants, in a manner that behavioral investigation typically does not. These two points combine to yield a difference between computational and behavioral investigation that has considerable significance for the investigation of pathological behaviors: A computational model provides a means of simulating impairment of multiple determinants of a behavior (one or more determinant at a time) in a manner that is usually impossible (practically and ethically) in behavioral investigation. For instance, in our sentence comprehension example, because the model would necessarily have to incorporate accounts of how words are represented, how their meanings are represented, and how the overall meaning of the sentence is derived, it also provides a means for examining how impairments of sentence comprehension can arise from various combinations of impairment in any or all of these factors. In effect, it would provide an "animal model." Recent thinking in the field of speech language pathology has emphasized the importance of using animal models where applicable.[2] It has been particularly emphasized that knowledge of learning, memory, and neuroplasticity in animals can and should be taken into account in thinking about remediation of impaired performance in humans because such remediation necessarily requires learning and memory to effect improvement in the patient. For example, principles of animal learning indicate that such things as the timing of treatment delivery and the frequency and intensity of training are relevant to regimens of therapy.[2] However, there are a great variety of human behaviors (and aspects of behaviors) that an animal model cannot address. For example, an animal model cannot address the question of whether the impaired learning in a particular syndrome arises more from impairment of syntax or of semantics. A computational model, however, can enable investigation of such questions and can thus provide a *pseudo-animal model* that is a powerful extension over what is possible through behavioral investigation. Of course, the plausibility of what is suggested by such a computational investigation of impairment depends on whether the model's account of *unimpaired* performance is plausible. As we have seen, establishment of this depends on relevant human data to compare against, and such data can only be obtained through behavioral investigation. Thus we return to the fact that computational investigation cannot and should not be viewed as a replacement for behavioral investigation. However, it can provide an extremely powerful supplement that is highly relevant for investigation of pathology.

The key points made in the preceding discussion can be summarized as follows:

Aspect 3 The term "computational" does not indicate any necessary involvement of a computer. Rather, it indicates that the model is specified in sufficient detail that the behavior predicted in a given situation can be numerically calculated, or computed. It incorporates a computationally specified account (or model, or theory) of the cognitive processing that underlies performance of a particular behavioral task.

Aspect 4 Investigation of cognitive behavior using a computational model (i.e., computational investigation) can supplement behavioral investigation but does not and cannot replace it. This is because the model must be assessed by comparing its behavior with the human data, which can only be obtained through behavioral investigation.

Aspect 5 However, the behavioral investigation that provides comparison data for a computational model would not typically include a computationally specified account of task performance. Thus a computational model of that task's performance does provide something that behavioral investigation typically does not.

Aspect 6 In a computational investigation, the creation of a computational account or model of task performance necessitates consideration of the fact that, for almost any task, behavior has multiple determinants.

Thus its focus is on specifying the relationship between these factors in determining task performance. In behavioral investigation, it is critical to ensure that all such factors other than the one(s) chosen for manipulation do not vary in task performance. The nature of the investigation does not necessitate specification of the relationship between the various determining factors. As a result, the manner in which the behavior is multiply determined is typically less salient.

Aspect 7 A computational model necessarily incorporates a computational account of how the multiple determinants of performance in the task of interest interact in determining the observed behavior. Thus it emphasizes that impaired behavioral performance can arise from impaired functioning of one or many of these determinants. If the functioning of these determinants can be "impaired" in the model, then it provides a means of examining what behavioral impairments are caused by impairment of different underlying determinants and, conversely, a means of examining what underlying impairment (or combination of impairments) best explains a particular behavioral impairment. To the extent that the model's unimpaired behavior can be shown to be a plausible account of human unimpaired behavior (as in Aspect 1), the causes and consequences of impairment observed in the model will constitute plausible hypotheses regarding impaired human behavior. As a result, a plausible computational model of normal human behavior in which the various determinants of performance can be systematically impaired offers a powerful means of investigating pathology of the behavior.

Finally, it is worth identifying some of the typical goals of computational investigations. The following rough classification is not an exhaustive list, but it may be helpful as a brief summary of some of the motivations that commonly underlie computational investigations. The accompanying brief discussion of examples also indicates that these motivations are not mutually exclusive, and models often have more than one of them.

1. Perhaps the most basic motivation for constructing a computational account of performance in a particular task would be to explain the human behavior observed in that task (most commonly, the focus is on unimpaired behavior). This would provide a mechanistic account that had presumably not been provided before. The account might also bring coherence to what had previously appeared to be a diverse set of behavioral results. Even models that may be motivated primarily by the other motivations noted later must satisfy this requirement of providing a reasonable account of several psychological phenomena, to a greater or lesser extent.
2. A further motivation might be to test whether a *particular* variable in the model has a particular hypothesized effect on (unimpaired) behavior, in a way that matches up with human data.

   The model by Dell,[3] for example, could be characterized as having primarily the first motivation. This model presented a computational theory of aspects of sentence production, with a view to explaining how typical patterns of speech production errors might arise. It thus aimed to provide a mechanistic account of human speech error production and bring coherence to diverse results in the field. In a specially designed experiment, the results from which were also simulated, this work also examined the effect of several specific variables on human as well as the model's performance, as a means of testing the model, corresponding with the second motivation just noted.
3. An additional motivation might be to examine a successful model's working so as to operationalize or clarify some poorly understood psychological construct in terms of the model's functioning.
4. The goal might be to show that a particular *type* or *style* of model *can* simulate a particular task. This would typically be meant to advocate for that type or style of modeling, as a means of advocating for some broader theoretical or meta-theoretical framework/approach.
5. The motivation might be to challenge an existing theoretical account of a phenomenon/set of phenomena by showing that an

account framed in terms of quite different psychological constructs can offer an alternative view. This might additionally, more or less directly, advocate for an alternative (meta-) theoretical framework.

The model by Gupta and Cohen[4] could be characterized as having primarily the third motivation—namely, to clarify and operationalize the psychological constructs of skill learning and repetition priming, which the authors considered to be poorly understood in the literature. Relatedly, this work aimed to advocate for an alternative view of behavioral phenomena that are construed in terms of these constructs, corresponding to the fifth motivation just outlined. The model of Rumelhart and McClelland,[5] which presented a computational account of English verb past-tense formation, was intended to challenge the then-standard account, which was framed in terms of morphological rules such as "add-*ed*." This motivation corresponds to the fifth one here. An important additional goal of this work, however, was to advocate for the meta-theoretical framework that is known as *parallel distributed processing* (e.g., Rumelhart and McClelland[6]) or *connectionism*, by showing that a computational model embodying the principles of this framework *could* simulate past-tense formation. This corresponds to the fourth motivation here.

6. The motivation might be to offer an account of *impaired* behavior in some psychological domain. In most cases, establishing the plausibility of the model's account of impaired performance requires providing a plausible account of the unimpaired behavior as well, although this latter account might not be as extensive as for a model with primarily the first of the motivations listed here. Such models have been considerably fewer in number.

The models of Hinton and Shallice[7] and Dell et al[8] are two influential models of impaired functioning that had primarily the sixth motivation noted here. The first was a model that aimed to examine how the constellation of word reading errors observed in what has been termed *deep dyslexia* could arise; this work also had the fifth motivation, in that it sought to reconceptualize thinking about this and other impairments.[7] The second was a model that sought to account for patterns of impaired picture-naming in aphasic patients as well as controls.[8]

The remainder of this article is divided into two parts. In the first part, I introduce the specific model that I use as the example for discussion. This model offers a computational account of how the sound patterns of words are learned and of how novel words are repeated. I discuss what I take the value of this model to be with respect to psychological investigation. In the second part, I discuss how aspects of processing in the model can be impaired, to create impairment of the model's behavior and simulate impaired processing in humans. I discuss how the fact that the model's behavior is multiply determined is an important aspect of the insight it can bring to the study of pathology. I outline a general strategy for the use of this model in investigating human pathology in word sound pattern learning and novel word repetition. I point out how this strategy is likely to be powerful for computational investigation of pathological behaviors in general.

AN EXAMPLE MODEL

Connectionist Models

The model I use here as an example is drawn from my own work (Gupta and Tisdale, unpublished data, 2008). It embodies a type of computational approach that has been termed *parallel distributed processing* or *connectionism*. It may therefore be useful to say something about this type of computational approach in general. Parallel distributed processing or connectionist or neural network models of psychological phenomena are comprised of simple processing elements usually termed "units," which are to be thought of as highly abstract neuron-like elements ("artificial neurons"). Such a unit has connections to and from other units. The outgoing connections from a unit are thought of loosely as axonal projections from a neuron. These projections make contact with other units, with such contacts being thought of loosely as synapses. The strength of any given "synapse" from unit *a* to unit *b* in such a model is instantiated as a *weight* on the connection from *a* to *b*. Thus a unit in such a model receives input from other units via the weighted incoming connections, summates its input, and, if the summed input crosses some threshold, transmits an output on its outgoing connections, thus providing input to other units to which it is connected. The input received by unit *b* from unit *a* is the product of the output emitted by *a* and the weight on the connection from *a* to *b*. The connection weights in such a model at any given point are thought of as encoding the model's *long-term knowledge* at that point. They start at some initial random value and are adjusted following each stimulus processing event. The gradual adjustment of connection weights in this way modifies the long-term knowledge, and is therefore thought of as *learning*. A variety of weight-adjustment or learning procedures exist for such models.

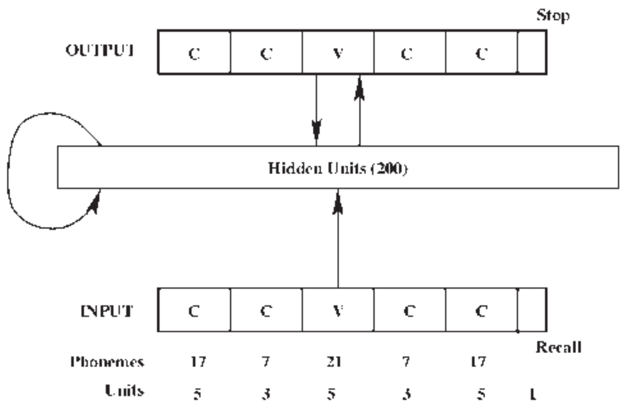
Typically, such models have one set of units that is termed an *input layer*. These units are thought of as being activated by external input stimuli. Thus the pattern of activation present at these units is thought of as representing an externally presented stimulus. There is also a set of units termed an *output layer*. The pattern of activation present at these units constitutes the model's output, or behavioral response. Thus the model responds to each input stimulus it is presented with. The goal of the learning procedure is to adjust the initially random-valued connection weights so that the model's behavioral responses to a set of input stimuli comes to approximate those that would be produced by humans when performing the task that is sought to be modeled.

The Example Model

Throughout the rest of this article, I use the term "phonological word form" to denote a phonotactically legal phonological form (i.e., sound pattern) of English, applying the term both to the phonological forms of real words of English and to the phonological forms of possible nonwords of English. The present model takes as its input a sequence of one or more syllables constituting a monosyllabic or polysyllabic phonological word form, with each phoneme in each syllable being represented individually. After the input has been presented, the model attempts to produce as its output the entire word form, as the correct sequence of syllables (including correct phoneme representations). That is, it receives as input and attempts to repeat phonological word forms that are each presented as a sequence of syllables.

The structure (*architecture*) of the model is shown in (Fig. 1A) and is an adaptation of an architecture introduced by Botvinick and Plaut.[9] The model has an input layer, at which a representation of an entire syllable is presented, and an output layer that uses the same representation scheme, at which the model's output is produced. The representation of a syllable, at both the input and the output layers, is in terms of a CCVCC (i.e., Consonant-Consonant-Vowel-Consonant-Consonant) template. That is, a syllable is represented at the input layer across a set of units that are divided into five slots. Activation of units in the first slot denotes the first C (if any) of the syllable, activation of units in the second slot denotes the second C (if any), and so on. Within each of these slots, the various phonemes that are legal for that slot for English are represented as different patterns of activations across a set of units. For example, for the encoding scheme used, 17 different phonemes of English are legal for the first C slot. These phonemes were represented as different patterns of activation across five units constituting the first C slot. Similarly, the 21 phonemes that are possible in the V slot were represented as different patterns of activation across a set of five units constituting the V slot; and so on, for the various slots shown at the input and output layers in Fig. 1A.

The model also has an intermediate layer of 200 units. Such an intermediate layer, which does not directly receive the model's input or directly produce the model's output, is usually termed a "hidden layer," with the units it contains typically termed "hidden units" (see Fig. 1A). All units in the input layer project to all units in the hidden layer, and all units in the hidden layer project to all units in the output layer, as is common in such connectionist models of cognitive phenomena. An additional aspect of the architecture is the self-connections on the hidden layer (indicated by the circular



A

Time step	INPUT	TARGET OUTPUT
1	flug	flug
2	wish	wish
3	"Recall"	flug
4	"Recall"	wish
5	"Recall"	"Stop"

B

Figure 1 (A) Architecture of the model. (B) Processing regimen in the model, illustrated for the word form *flugwish*.

arrow from the hidden layer back to itself), which denote a connection from every unit in the hidden layer to itself and to every other unit in the hidden layer. These connections are termed "recurrent" connections. The model's task is to accept as input a *sequence* of syllables and to produce as output a *sequence* (for this model, the same sequence) of syllables. It is well established that for connectionist models that perform sequential processing of this kind, the presence of recurrent connections is critical. Thus the recurrent connections in the present model are crucial for it to be able to perform the task of inputting and repeating phonological word forms presented as sequences of syllables.

Figure 1B illustrates the regimen of presentation and desired output in the model, for the example word form *flugwish*. The procedure is the same, irrespective of whether or not the word form has been presented to the model previously. Following presentation of the first syllable *flug* at the input, the model's task is to produce that same syllable at the output. The model's actual output may be either correct or incorrect. Either way, after the model has produced an output, the activation pattern at the hidden layer is transmitted across the recurrent connections, thus transmitting information to *itself* that will arrive at the next time step, so that when the second syllable *wish* is presented, the model's hidden layer actually receives input from two sources: the input representing *wish*, and input from its own previous state. When presented with this second syllable at the input, the model's task, as for the first syllable, is to produce the input syllable at the output. Again, the output may be correct or incorrect, and the hidden layer activation pattern is transmitted across the recurrent connections to be available at the next time step. The input at this next step is actually an indication of the end of input, denoted by activation of the "Recall" unit in the input layer. At this point the model's task is to produce at the output layer the *entire* sequence of syllable representations previously presented at the input layer (i.e., *flug* followed by *wish*), and then activate the "Stop" unit at the output layer, to signify the end of production of the word form. Because this repetition must be performed in the absence of any external input representing the word form, the network must necessarily have encoded some internal memory representation of the word form to allow it to now produce it in correct sequence (i.e., to perform immediate serial recall of the word form). At each point during recall, the model's hidden layer receives input from activation of the Recall unit, and from its own state at the previous time step. (Note that the activation of the Recall unit is only a cue and carries no information about the specific word form that was presented because this same unit is activated as a cue for *all word* forms). Overall, the model attempts to match its own production (i.e., repetition) of a syllable sequence constituting a word form with the observed linguistic sequence provided by the environment. At the end of presentation and repetition of one word form, the model's connections weights are adjusted using a learning procedure for neural networks with recurrent connections known as *back propagation through time*, whose details are beyond the scope of this article (for further discussion, see references 9 and 10).

PHENOMENA SIMULATED BY THE MODEL

The aim of the present discussion is simply to provide a sense of the behaviors simulated by the model; Gupta and Tisdale, unpublished data, 2008). The most basic behavior exhibited by the model was *phonological vocabulary learning*. Via connection weight adjustments following repeated presentations of a set of word forms representing those occurring in the environment, the model developed more and more detailed internal phonological representations of them, learning to produce them with increasing accuracy. At each point in this learning trajectory, the word forms that could be correctly produced constituted the model's *phonological vocabulary*. Importantly, this was a phonological vocabulary of over 100,000 word forms of English, not a small set as is common in such models. Thus the model learned to correctly produce (i.e., repeat) over 100,000 word forms of English, outputting their syllables in correct order and correctly producing all phonemes within each syllable.

The model was also able to repeat novel word forms that had not been part of its training vocabulary. That is, it was able to repeat what to it were nonwords. In a series of simulations, the model's nonword repetition exhibited several key phenomena that have been documented in human nonword repetition, as follows:

1. The model's nonword repetition accuracy was poorer for longer than for shorter nonwords, matching human data.[11]
2. The model's nonword repetition accuracy, when assessed by syllable serial position within nonwords, exhibited better recall for syllables in beginning and ending serial positions, and poorer recall for syllables in middle positions (so-called *primacy* and *recency*, or *serial position effects*), matching human data.[12,13]
3. The model's nonword repetition accuracy was greater for high phonotactic probability nonwords than for low phonotactic probability nonwords, exemplifying an effect of phonological knowledge on nonword repetition performance, matching human data.[14,15]
4. The model's patterns of errors in its nonword repetition matched those of human subjects repeating the same stimuli. For instance, when a syllable of a nonword was incorrectly produced, the syllable structure of the target syllable was nevertheless overwhelmingly preserved, in the model as in the human data. As another example, the proportion of incorrectly produced syllables that incorporated single versus multiple phoneme errors in the model was closely similar to that in the human data.

VALUE OF THE MODEL

For present purposes, two aspects of the model are valuable for furthering our understanding of vocabulary learning and nonword repetition. In recent years, much attention has focused on reported links between human nonword repetition accuracy and phonological vocabulary learning (for a recent summary, see Gathercole[1]). As noted earlier, the model simulated both these behaviors, matching human performance in several ways that suggested the model provided a plausible account of human nonword repetition. This meant that examining the internal details of how the model achieved its behaviors would provide plausible hypotheses about what might underlie human nonword repetition performance.

This is a matter that has been the subject of considerable debate. In brief, there have been two opposing views (for further discussion, see Gathercole[1]). One view, which Gupta and Tisdale termed the "PSTM account" (unpublished data, 2008), posits that PSTM causally determines nonword repetition performance as well as phonological vocabulary learning ability, and it thereby explains the well-documented link between the two. Another view, which we termed the "linguistic account," maintains it is phonological vocabulary size that determines nonword repetition performance. The intuition motivating this account is that an individual with a larger vocabulary has greater phonological knowledge, and that greater phonological knowledge supports better performance in processing and repeating novel phonological forms such as nonwords. Phonological vocabulary size is, of course, determined by phonological vocabulary learning. Thus the observed correlations between phonological vocabulary learning and nonword repetition would in this view be mediated by phonological vocabulary size, rather than by any necessary common involvement of *phonological short-term memory* (PSTM). (For a review of the evidence for each view, see Gathercole[1]).

The first aspect of the model's value highlighted here is that it provides a very clear operationalization of the construct of PSTM. As noted earlier, the recurrent connections are crucial for sequential processing of word forms. We were able to show that the information maintained by these recurrent connections from one step of processing to the next is, in fact, PSTM: it is information about the network's states during the past, and therefore *memory*; it is information that is available only temporarily, and therefore *short-term memory*; and it is information that encodes phonological information, and is therefore *phonological short-term memory*. The value of the model in this case is its clarification of an important psychological construct that has remained poorly understood (corresponding to the third motivation of computational modeling noted earlier).

Gupta and Tisdale (unpublished data, 2008) were further able to show that elimination of this information in the model completely disrupted its ability to produce nonwords and to learn a phonological vocabulary. This indicated that PSTM was crucial for nonword repetition and for vocabulary learning, supporting the PSTM account. However, it was also shown that nonword repetition ability increased as a function of phonological vocabulary size, without change in the model parameters governing maintenance of PSTM information. That is, the model indicated that the postulates of *both* the PSTM and the linguistic accounts are correct. This was further supported by the model's simulation of the various phenomena described earlier because some of these have been viewed as evidence for the PSTM account and others as evidence for the linguistic account. Because of the model's overall plausibility, its suggested reconciliation of the linguistic and PSTM accounts can credibly be viewed as applicable to human performance.

This resolution of a theoretical debate is the second valuable aspect of the model highlighted here What enabled the model to achieve this resolution? It was the fact that the model incorporates a detailed (*computational*) specification of the mechanisms underlying each of the behaviors of interest. It was this computational specification that enabled operationalization of PSTM, demonstration that PSTM is crucial for nonword repetition and phonological vocabulary growth, and demonstration that accuracy in nonword repetition increases as a function of vocabulary size even without any change in PSTM parameters. Because the model was computationally specified, it was possible to see and demonstrate unambiguously unambiguously that *both PSTM* and vocabulary size are determinants of nonword repetition.

As noted earlier, the computational specification of how the behaviors of interest arise forces awareness of the fact that these behaviors are jointly determined by very many aspects of the model's functioning. The consequences of this focus on *multiple determination* for examination of pathology are what I turn to next.

INVESTIGATION OF PATHOLOGICAL BEHAVIOR

One aspect of the examination of pathological behavior in the model has been mentioned earlier: the investigation that showed elimination of PSTM information abolished nonword repetition and vocabulary learning ability. But this fails to speak to individual differences, which would require showing that *graded* manipulation of aspects of the model yield graded effects on performance, some in the "normal" range and some in an "impaired" range. What would such evidence look like? This is where the model's multiply-determined nature is relevant. Gupta and Tisdale (unpublished data, 2008) noted that impairment of numerous parameters in the model would lead to impaired nonword repetition performance. In particular, nonword repetition performance in the model was affected by manipulation of a parameter corresponding to PSTM but also by manipulation of a parameter corresponding to *learning efficiency*. Each of these constructs (PSTM and learning impairment) has been proposed as a primary cause of deficits in *specific language impairment* (SLI).[16,17] The findings from the model indicate that manipulation of *each* of these variables affects nonword repetition performance. This is also true of many other processing variables the model clearly defines that correspond with psychological or cognitive constructs. Thus, although it is true that nonword repetition performance depends on PSTM, it is not the case that an observed nonword repetition deficit necessarily implies an underlying PSTM deficit. Likewise, for the many other processing elements of the model whose impairment can cause nonword repetition problems, the impaired repetition performance of the model does not in itself specify which processing element/s is/are impaired. To the extent that the model is plausible, this is likely to be true in the human case as well. Nonword repetition in the model is multiply determined; the fact that the model is computationally specified enables unequivocal determination of this important point.

But where does this leave the investigation of what *does* underlie a particular observed behavioral impairment syndrome or pathological condition? A model such as this still offers an extremely powerful means—perhaps the only means—of addressing this question seriously.

Let us consider behavioral impairment syndrome *X*. For concreteness, let us suppose that *X* is conduction aphasia. This pathological syndrome is, like most other syndromes, characterized by more than one behavioral deficit (e.g., impaired word and nonword repetition, phonological paraphasias in production, reduced phonological short-term memory). The recommended approach to investigation of this syndrome would be to examine *multiple behavioral variables*. This approach has several components:

A. Identify multiple behavioral measures that together characterize performance of those with conduction aphasia (CA) and of control subjects (CS). For conduction aphasia, some of these measures would be word and nonword repetition scores; rates of phonological errors in repetition; immediate serial list recall scores; scores on tests of phonological perception; scores on tests of semantic ability; measures of novel word form learning.
B. Identify the correspondence of processing elements in the model with cognitive constructs of interest. For instance, in the Gupta and Tisdale model (unpublished data, 2008), there are clear correspondences of model processing parameters to a PSTM parameter, a phonological learning rate parameter, and a phonological discrimination ability parameter. There are also clear correspondences of *structural* processing elements of the model to other psychological constructs such as input phonological processing and output phonological processing.
C. Use a computational model (or a small number of computational models that incorporate the same processing elements) to simulate behavior in each of the tasks (the ones from which the behavioral measures are obtained), matching the model's performance to CS performance (schematized, for example, in Fig. 2A). This requires determining which combination of values of the model's processing elements of

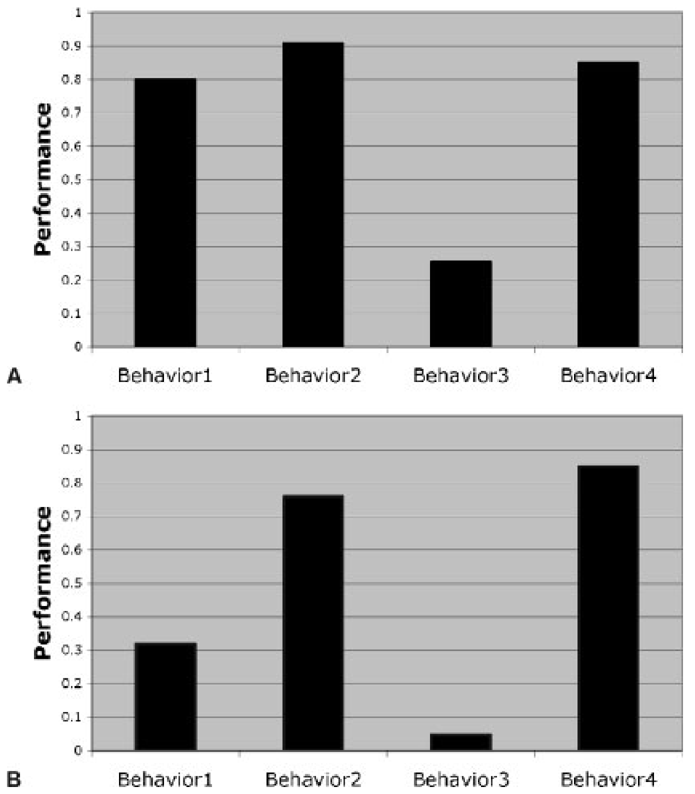


Figure 2 Schematic graph of performance on multiple behavioral measures for (A) control subjects (CS) and (B) individuals with conduction aphasia (CA).

interest (determined in the previous step) produces a set of behaviors that best fit the CS performance profile in Fig. 2A.

D. Match the model's profile of performance to the CA behavioral profile (Fig. 2B). This requires configuring the processing elements of interest in the model so that the behaviors of the model exhibit the CA performance profile depicted in Fig. 2B. This process is essentially identical to that involved in configuring the model to simulate the CS performance profile in step C. Note that in addition to varying parameters of the model, simulated damage to various structural components of the model could also be investigated—for instance, by varying the amount of damage to the input phonology versus output phonology connections in the model or to the hidden layer units in the model.

E. At this point, the models simulates a variety of behaviors in both CS and CA, all based on variability in a small number of processing variables of interest, which correspond to psychological constructs, and/or based on simulated damage ("lesions") to structural components of the model that also correspond to psychological constructs. The difference between the CS and CA values of these variables/ structural components constitute the model's hypothesis about what is different in CA as compared with CS. To the extent that the model is plausible with respect to unimpaired (i.e., CS) performance, this hypothesis can be considered a strong one.

CONCLUSION

In this article, I have aimed to concretize how computational models can make a genuine contribution to our thinking about cognitive behavior, both unimpaired and impaired.

With respect to computational models of unimpaired cognitive behavior, a common suspicion is that they are merely exercises in parameter fitting (like, for example, a regression model). As a result, it is argued, a model's ability to simulate a given behavioral phenomenon does not establish its processing mechanisms as a good account of human psychological/cognitive processing, any more than the calculations underlying the fitting of a regression model would be. Two points here are worth noting. First, a computational model is, by definition, a *highly* specific hypothesis about processing underlying the behavior in question. It is this very specificity that opens it to the charge of having no necessary correspondence to actual human processing. In contrast, a noncomputational theory does not posit any mechanisms, at least, not in a rigorously testable way, and therefore escapes the criticism simply by virtue of its lack of specificity. Second, however, it is true that the mere fact a computational model can simulate a certain behavior does not mean it is a good account of that behavior.

Three principles of modeling can offset the potential limitations of computational models noted earlier and realize their considerable power as tools for studying human cognitive processing.

The first principle is that the model must establish its plausibility by demonstrating that *the same model* can simulate a range of phenomena in the domain of interest. If a single model can give an account of multiple behavioral phenomena, without any changes in parameter values or other structural properties, this greatly enhances the credibility of the model and offsets the charge of parameter fitting.

The second principle is that the model must account for behavior that it was *not designed to account for*. The Gupta and Tisdale model (unpublished data, 2008), for example, was certainly designed to perform nonword repetition, and its performance was calibrated so it would match with human adult performance (in overall nonword repetition accuracy) at the end of its training. However, no aspect of the model's design was included with a view to simulating other phenomena it exhibits, such as the effect of nonword length on nonword repetition accuracy, the effect of phonotactic probability on nonword repetition accuracy, or the error patterns it produces.

A computational model that incorporates these two principles can go a long way to addressing the legitimate criticisms noted earlier and becomes a strong candidate to be a credible account of human cognitive processing. The most influential computational models, many of which have fundamentally altered our thinking about the psychological phenomena involved, all incorporate these two principles.

A third principle is relevant to the understanding of *individual differences* and *impaired* cognitive behavior and related to the previous two principles. If processing elements in a model can be identified (e.g., processing components, such as the hidden layer, or processing parameters, such as the learning rate parameter) that correspond plausibly with cognitive constructs, and if manipulation of these variables (e.g.., through simulated damage to processing components or systematic variation of processing parameters) can be shown to yield individual differences analogous to those in human behavior, in both normal and impaired ranges, then the model has simultaneously achieved two things: (1) it has established its candidacy as a plausible account of the *impaired* behavior and *individual differences*, and (2) it has enhanced its credibility overall as an account of the relevant human cognitive behavior(s). This latter aspect is because the human individual differences in performance and human impaired performance constitute additional data that the model was not designed to account for; and accounting for them strengthens its adherence to the second principle just described (in this case, with variation of the model, but with such variation being principled).

With the adoption of this third principle, computational models begin to realize their potential as tools for understanding individual differences and impaired human cognition (that is, understanding of *pathology*), as well as to further enhance their power as models of human behavior in general. Although few computational models have as yet taken this third step, especially with regard to individual differences, it has been my aim in this article to show that this approach is certainly possible and should more often be taken in computational modeling.

ACKNOWLEDGMENTS

Jamie Tisdale conducted the simulations for the computational model used here as an example. This work was supported in part by grants NIH R01 DC006499 to Prahlad Gupta and NIH RO1 DC01924 to Nadine Martin, Prahlad Gupta, and others.

REFERENCES

1. Gathercole SE. Nonword repetition and word learning: the nature of the relationship. Appl Psycholinguist 2006;27:513–543
2. Raymer AM, Beeson P, Holland A, et al. Translational research in aphasia: from neuroscience to neurorehabilitation. J Speech Lang Hear Res 2008; 51:S259–S275
3. Dell GS. A spreading activation theory of retrieval in sentence production. Psychol Rev 1986;93:283–321
4. Gupta P, Cohen NJ. Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. Psychol Rev 2002;109: 401–448
5. Rumelhart DE, McClelland JL. Learning the past tenses of English verbs: implicit rules or parallel distributed processes? In: MacWhinney B, ed. Mechanisms of Language Acquisition. Mahwah, NJ: Erlbaum; 1987
6. Rumelhart DE, McClelland JL. Parallel Distributed Processing. Vol. 1: Foundations. Cambridge, MA: MIT Press; 1986
7. Hinton GE, Shallice T. Lesioning an attractor network: investigations of acquired dyslexia. Psychol Rev 1991;98:74–95
8. Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in normal and aphasic speakers. Psychol Rev 1997;104:801–838
9. Botvinick MM, Plaut DC. Short-term memory for serial order: a recurrent neural network model. Psychol Rev 2006;113:201–233
10. Rumelhart D, Hinton G, Williams R. Learning internal representations by error propagation. In: Rumelhart D, McClelland J, eds. Parallel Distributed Processing Vol. 1: Foundations. Cambridge, MA: MIT Press; 1986
11. Gathercole SE, Willis CS, Emslie H, Baddeley AD. Phonological memory and vocabulary development during the early school years: a longitudinal study. Dev Psychol 1992;28:887–898
12. Gupta P. Primacy and recency in nonword repetition. Memory 2005;13:318–324
13. Gupta P, Lipinski J, Abbs B, Lin P-H. Serial position effects in nonword repetition. J Mem Lang 2005;53:141–162
14. Gathercole SE. Is nonword repetition a test of phonological memory or long term knowledge? It all depends on the nonwords. Mem Cognit 1995; 23:83–94
15. Munson B, Kurtz BA, Windsor J. The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetition of children with and without specific language impairment. J Speech Lang Hear Res 2005;48:1033–1047
16. Gathercole SE, Willis C, Emslie H, Baddeley AD. The influences of number of syllables and wordlikeness on children's repetition of nonwords. Appl Psycholinguist 1991;12:349–367
17. Ullman MT, Pierpoint EI. Specific language impairment is not specific to language: the procedural deficit hypothesis. Cortex 2005;41:399–433