Clinical Aphasiology, Vol. 23, 1995, pp. 25–38

Response Variability in Naming: A Computational Study

Steven L. Small, Audrey L. Holland, John Hart, Jr., Margaret M. Forbes, and Barry Gordon

PERFORMANCE VARIABILITY

Performance variability greatly complicates both the treatment of aphasia and the design and execution of clinical studies of aphasia recovery. Two different performance issues are involved in this problem. The first issue concerns clinical neuroanatomical correlation: Two patients with apparently identical structural brain lesions can manifest grossly disparate performance in language tasks. The second problem involves individual patients themselves: At two different times of the day or under different environmental conditions, a patient can demonstrate quite variable language skills.

Neuroanatomical Variability

Despite the recent proliferation of localization arguments in neurolinguistics (Damasio & Damasio, 1989; Kertesz, 1983), fueled by advanced imaging techniques such as positron emission tomography (Petersen, Fox, Mintun, Posner, & Raichle, 1988) and functional magnetic resonance (Belliveau et al., 1991), there remain strong reasons to doubt strict neuroanatomical localization of language tasks (Jackson, 1878; Marie, 1906). Some recent studies have shown so much individual variability in the neuroanatomy of naming that the authors distinctly argue against all but the most general localization (Gordon et al., 1990; Ojemann, Ojemann, Lettich, & Berger, 1989). Variability

in neural representations extends also to the primary motor and sensory areas, long thought to be a paradigmatic example of invariability (Uematsu et al., 1992).

Behavioral Variability

Language processing variability by particular patients over time is well known (Brain, 1961). Simple environmental manipulations and/or constitutional alterations affect performance, and these have formed the basis of efforts to improve communication for such patients. However, as McNeil, Odell, and Tseng (1991) noted, neither variability nor stimulability can be explained within the predominant neoclassical theories of aphasia (Benson & Geschwind, 1989).

In her thesis, Crisman (1971) studied behavioral variability in naming after ischemic stroke. Twelve patients were equally divided between chronic (>3 months) and acute (<3 months) conditions and were tested daily for 10 consecutive days on a picture naming task. Crisman selected 10 high-frequency words and 10 low-frequency words, all denoting physical objects, and asked the patients to name each one from a black-and-white line drawing. These words are shown in Figure 1, along

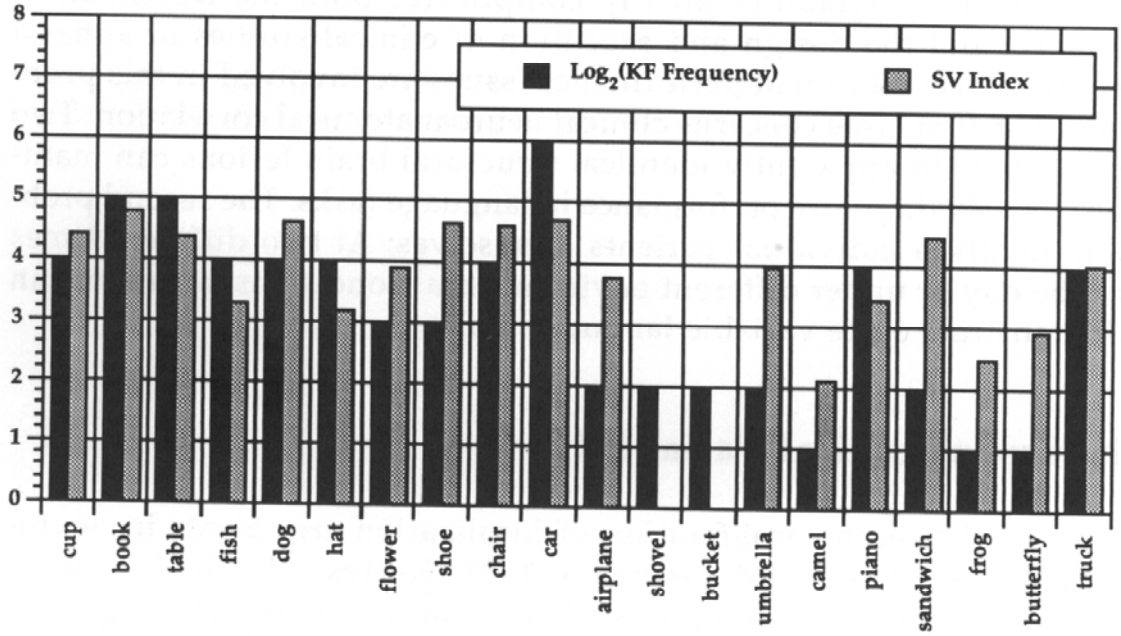

Figure 1. The corpus of words used for testing response variability in chronic aphasic subjects (Crisman, 1971) and in a damaged computer model of naming, along with indices of their frequency of occurrence. The black bars show the logarithm of the word frequencies in the written corpus of Kŭcera and Francis (1967) (KF Frequency) and the checked bars show the familiarities of the pictures in the study of Snodgrass and Vanderwart (1980) (SV Index).

with the natural logarithm of their frequencies of occurrence (Gordon & Caramazza, 1985) in standard written corpora (Kŭcera & Francis, 1967) and the familiarity of their pictorial form (Snodgrass & Vanderwart, 1980). Individual patient responses were transcribed and classified into 12 error types, and then compared statistically.

These comparisons demonstrated that (a) response variability is a prominent feature of naming performance in aphasic subjects; (b) individual patients tend to make the same types of error responses over time; (c) acute and chronic patients do not differ in their variability; and (d) patients make more errors on low-frequency words than on high-frequency words. This paper compares the response variability of Crisman's (1971) chronic aphasic subjects with that of a connectionist computer model and offers a computational neurobiological explanation that may partly explain response variability.

This approach is aimed at providing an analytical explanation for the concept of variability. Without trying to assimilate empirical data into a theory, one does not advance science. The use of a computer simulation is a way to formalize theoretical concepts. In this case, it is asserted that variability may be the result of imprecision in information processing, which is normally not evident behaviorally because of overlearning, redundant representations, and so forth.

CONNECTIONIST MODEL

Overview of Connectionist Modeling

Neural (connectionist) networks are being applied increasingly to studies in cognitive neuroscience (Sejnowski, Koch, & Churchland, 1988). Connectionist models consist of simple computational elements, or *units*, which communicate by sending their level of activation via labeled links to other elements.[1] The units may have a small number

1. The activation value is the numerical value associated with a computation unit of a neural network. In some networks, this activation is analogous to a strength of belief in the item represented by the unit. In others, it is a neuronal potential. In this model, each activation value in the input encodes the presence or absence of a particular semantic feature. Each activation value of the output encodes the belief that one physical object is encoded by the semantic features of the input. The hidden activation values do not have such clear meaning as they encode the intermediate results of distributed processing performed by the network in making its object identification. A full discussion of the construction and use of artificial neural networks for models of communication disorders can be found in Small (1994).

of states, and compute simple functions of their inputs. Associated with each link is a *weight*, indicating the significance of activation arriving over that link. The overall behavior of the model is determined by the pattern of connections, the weights on the links, and the unit functions. Figure 2 shows the main features of an artificial neural network.

In some connectionist models, particularly parallel distributed processing (PDP) models (McClelland & Rumelhart, 1986), the units are arranged into layers, with the majority of interaction occurring between adjacent layers (e.g., the top and middle layers or the middle and bottom layers in Figure 3), rather than within layers or between spatially separated layers (e.g., the top and bottom layers in Figure 3). Two layers play special roles as input and output to the network. Between the input and output are any number of intermediate layers of *hidden units*. After determining a method of encoding (i.e., a *representation*) for the desired input and output information, the network is provided with a set of examples, or teaching data from which to learn the desired mapping of input to output. For each set of input values in the teaching set, the network uses a combination of unit values and weights to

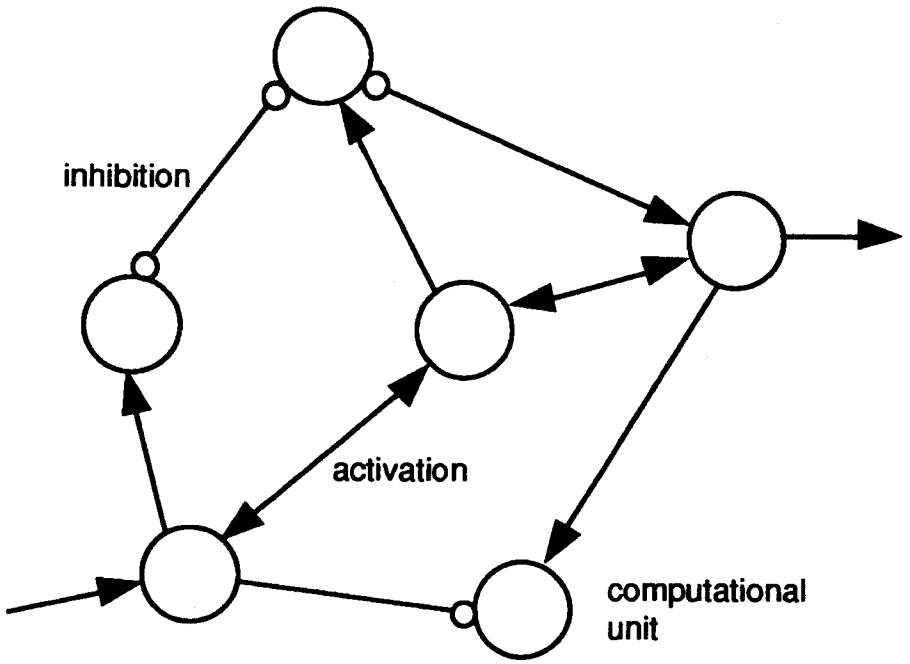

Figure 2. A neural (connectionist) network, illustrating the basic features of such networks, including units, excitatory connections, and inhibitory connections. Note that connections can be unidirectional or bidirectional, and can have different types of connections with different units.

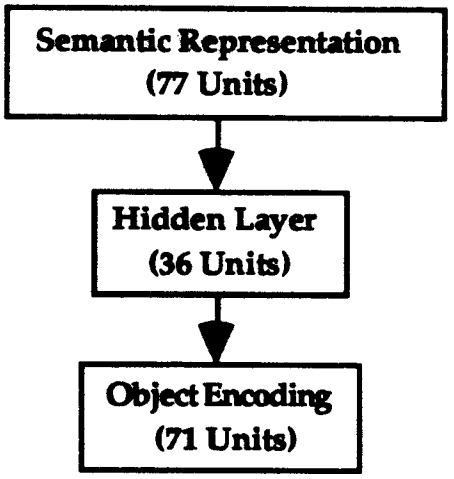

Figure 3. A three-layered feed forward connectionist network to perform object naming. The network maps a collection of semantic features (the input layer) into a single object (the output layer), using an intermediate encoding (the hidden layer) that evolves through network learning.

produce values for the next layer.[2] The same procedure is used on this layer to produce the next layer, and ultimately the output layer. Until the network has learned the correct mapping, this output will be wrong. The main trick of the PDP approach is to compare the desired (or *target*) output with the actual output, and then to adjust the weights in the network to minimize this difference. Several approaches to this "supervised learning" have been devised, but the one most commonly used is the back propagation algorithm (Rumelhart, Hinton, & Williams, 1985).[3]

When the network has finished adjusting the weights, it produces the desired output for each input. In addition, for each input, the

2. The teaching set contains the associations that the network is being asked to learn, comprising a large list of input–output associations. The input and output each consist of a number of values, corresponding to the number of units in the input layer or output layer to the network.

3. Back propagation is a "supervised learning" method (i.e., a learning method that uses an explicit teacher to explain to the network which answers are the right ones and which are the wrong ones). It can be used to teach the artificial neural network to make associations. During the process of learning associations, the units that are neither input nor output units acquire activation values that encode the results of intermediate processing stages. A full description of this algorithm in the context of the study of language disorders can be found in Small (1994) or in mathematical detail in Rumelhart et al. (1985).

network produces values across each layer of hidden units. These hidden unit values constitute an emergent feature of such networks, and they can be analyzed directly and with methods from linear algebra and statistics. Studies of such networks have demonstrated that, in solving complex problems in cognitive neuroscience, the hidden layers of these networks develop encodings that reflect interesting regularities of the input data. Examples in semantics (Hinton, 1986), natural language (Elman, 1989; Small, 1990), and spatial reasoning (Zipser & Anderson, 1988) have all led to interesting results. In the latter case, the emergent representations led to hypotheses about the anatomical structure of the parietal lobe.

Connectionist models of impaired language function involve construction of networks that perform some desired neurolinguistic mapping and then damaging them in some methodical way and analyzing the result. A number of models have been built, which account for limited disciplinary data in one area of language processing, such as lexical access (Small, Cottrell, & Tanenhaus, 1988), semantic priming (Kawamoto, 1988), acquired dyslexia (Hinton & Shallice, 1991), spatial neglect in reading (Mozer & Behrmann, 1989), and temporal processing in aphasia (Gigley, 1988). Previously, we have reported a taxonomy of artificial lesions and a postulated correspondence between artificial lesion types and human brain damage (Small, 1991).

Connectionist Model of Naming

We developed a computational model of naming (Small, Hart, & Gordon, 1992) and studied its performance under conditions of damage. The model consists of a feed forward connectionist network of three layers, in the manner of the parallel distributed processing approach described above. These layers consist of an input layer of semantic feature units, a middle (i.e., hidden) layer for computations, and an output layer representing words. The network was trained to map from a set of semantic features to specific objects. Phonological encoding, a separate and difficult computational research question (Dell, 1986), was not performed.

Seventy-one pictures, each showing a line drawing of a single physical object, were represented in terms of 77 semantic features in the style of the representation of Hinton and Shallice (1989). These included pictures used in previous naming studies (Hart & Gordon, 1992) and Crisman's (1971) pictures. The input layer represented a picture in terms of semantic features. Table 1 shows this semantic feature encoding for four example pictures in the experimental corpus, including the high-frequency words *table* and *book* and the low-frequency words *shovel*

Table 1. The Semantic Features of Four Example Words from the Experimental Corpus, Including Two High-Frequency Words (*Table* and *Book*) and Two Low-Frequency Words (*Shovel* and *Camel*)

Table	*Book*
LIMBS-4	LIMBS-0
COLOR-variable	COLOR-variable
SIZE->1-foot-and-<2-yards	SIZE-<1-foot
CROSS-SECTION-rectangular	CROSS-SECTION-rectangular
FORM-rectangular	FORM-rectangular
FORM-four-legged	MADE-OF-from-plant
MADE-OF-wood	TEXTURE-rough
MADE-OF-metal	TEXTURE-smooth
MADE-OF-glass	TEXTURE-leafy
MADE-OF-plastic	NOISE-with-input
TEXTURE-smooth	FUNCTION-esthetic
NOISE-with-input	LOCATION-bedroom
FUNCTION-esthetic	LOCATION-living-room
FUNCTION-cooking-and-or-eating	LOCATION-kitchen
LOCATION-bedroom	LOCATION-basement
LOCATION-living-room	LOCATION-farm
LOCATION-kitchen	MOVEMENT-can-be-propelled
LOCATION-basement	
LOCATION-farm	
MOVEMENT-can-be-propelled	

Shovel	*Camel*
LIMBS-0	LIMBS-4
COLOR-variable	COLOR-brown
SIZE->1-foot-and-<2-yards	COLOR-yellow
CROSS-SECTION-circular	SIZE->2-yards
FORM-cylindrical	CROSS-SECTION-circular
MADE-OF-wood	FORM-four-legged
MADE-OF-metal	MADE-OF-from-animal
MADE-OF-other-manmade	TEXTURE-hairy
TEXTURE-smooth	NOISE-on-its-own
NOISE-with-input	FUNCTION-transportation
FUNCTION-cleaning	MOVEMENT-self-moving
LOCATION-basement	MOVEMENT-grows
LOCATION-yard	
MOVEMENT-can-be-propelled	

and *camel*. The output layer consisted of one unit per word, so that naming an object consisted of causing the activation of a (single) correct output unit.

This three-layered architecture is shown in Figure 3. The network was trained such that the number of times each word was presented to the network was proportional to the natural logarithm of its Kŭcera and Francis (1967) frequency. After about 500 teaching cycles or stimulus presentations, the network had learned all of Crisman's words except *camel* and *shovel*, two of the least frequent words in the corpus. The learning curves for these two words, as well as two of the most frequent words in the corpus, *table* and *book*, are shown in Figure 4. After about a thousand cycles, all the words were learned to the output threshold of 0.6, although perfect learning (to an output threshold of 0.9) was not achieved until about 1,500 epochs. At this stage, the network was ready for the lesioning experiments.

The fully working network was lesioned by damaging the intermediate layer of hidden units (see Figure 3).[4] Network lesions involved either deleting a proportion of the hidden units or adding noise to a proportion of the hidden units. Six lesioned networks comprise the "patient population" for the current study, and are shown in Figure 5.

Unlike typical connectionist networks, this network does not contain any fixed weights or unit activation values. Instead, these numerical values that give the network its particular behavior have an intentionally imprecise (or "fuzzy") character (Zadeh, 1965), with any particular activation level or weight differing by 10% from a mean value. Although such imprecision gives these networks inherent error, it is usually subclinical in undamaged networks, leading to consistent performance. Thus, the amount of imprecision in the networks does not change after damage, but the effects of the imprecision become significant as a direct consequence of the damage.

EMPIRICAL STUDY OF THE ARTIFICIAL SUBJECTS

We used Crisman's (1971) study design to test the lesioned networks. The examiner presented the entire set of 20 pictures to each network

4. Lesioning can be either focal or diffuse, with a variety of different techniques available for damaging networks, as summarized in Small (1991). Focal lesions typically involve a spatially coherent set of connections or units (as focal brain lesions involve a set of neurons and axons in a particular location of the brain), and diffuse lesions involve damage to units and connections spread over the entire network (by analogy again with diffuse brain lesions). Damage can include deletion of units or connections, attenuation of their values, and addition of noise to their processing.

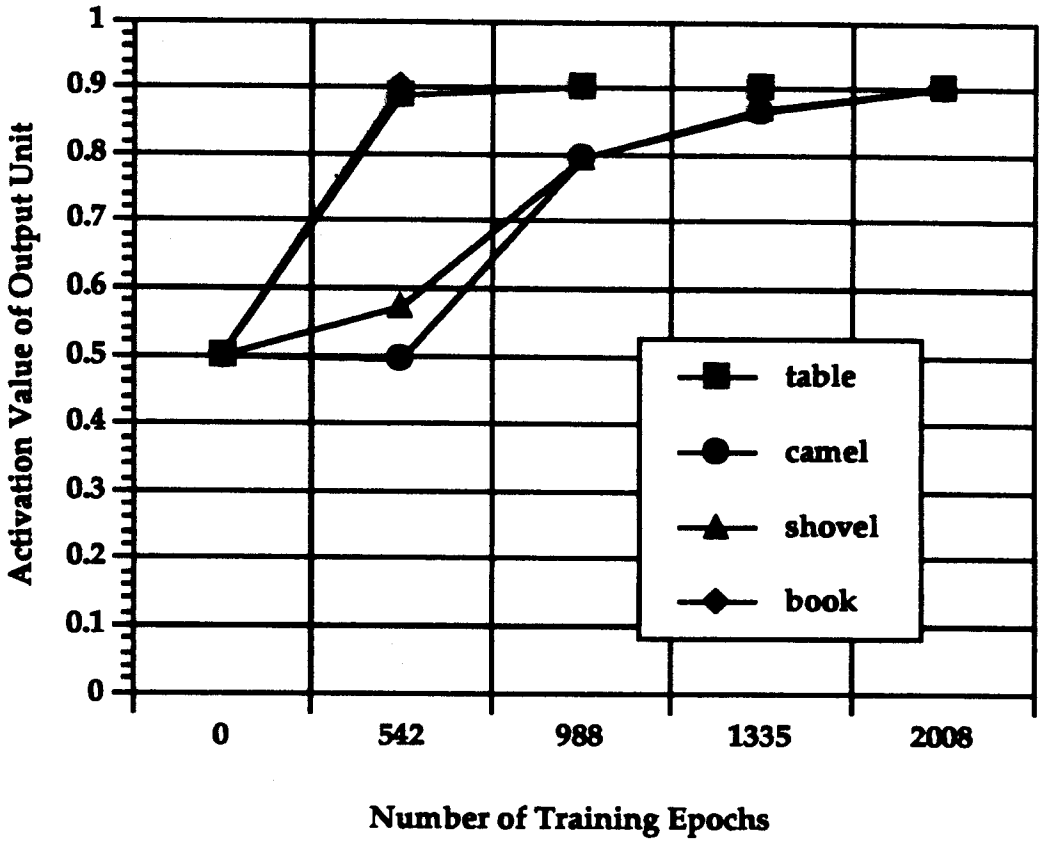

Figure 4. Graph depicting the different rates of learning for the names of two high-frequency objects and two low-frequency objects. The high-frequency words *table* and *book* are learned within the first 600 stimulus presentations (i.e., epochs), but the low-frequency words *camel* and *shovel* are not fully learned until more than 1,000 epochs. Note that an output value of 0.9 indicates a maximally learned value.

on each of 10 successive test sessions. The network named each picture in one of three ways: (a) correctly, (b) incorrectly, or (c) without a response. As we did not address issues of phonological encoding, we collapsed Crisman's 12 categories of error into these three broader classes. Analysis of the errors included (a) error frequency as a function of word frequency; (b) response variability over each of the 10 test sessions for each network; and (c) percentage agreement between the responses during the first and last test sessions.

Figure 6 plots the mean number of errors for each of the six chronic aphasic subjects and each of the six network subjects. Error bars show the standard deviations of this error. Figure 7 shows the percent agree-

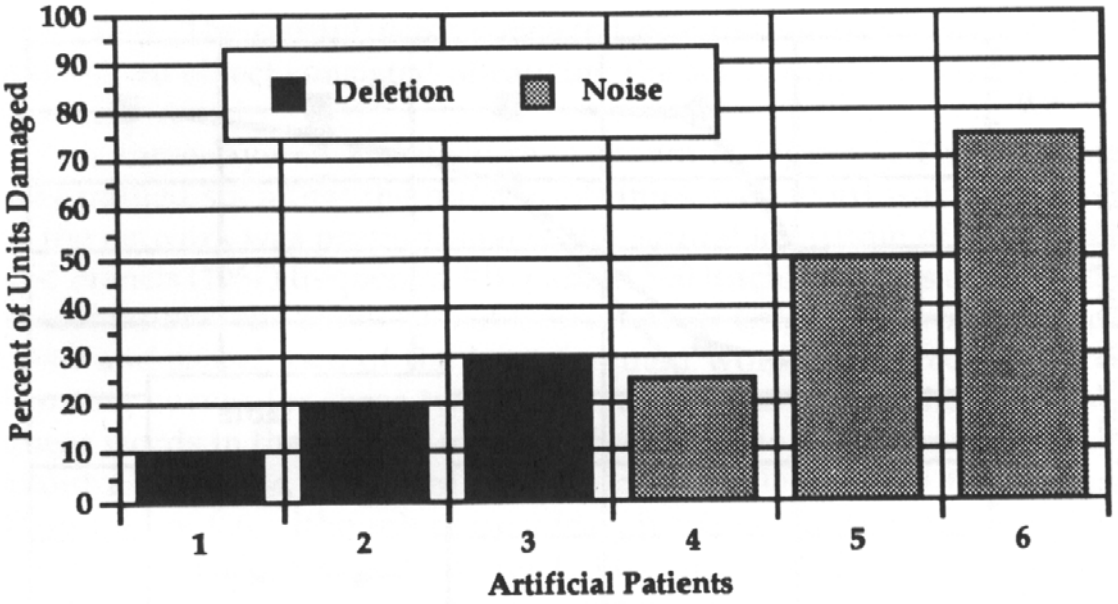

Figure 5. Six "artificial subjects" (i.e., computer networks) participated in the experiment. The black bars show the three networks from which hidden units were deleted. The checked bars show the three networks that had noise added to their hidden units.

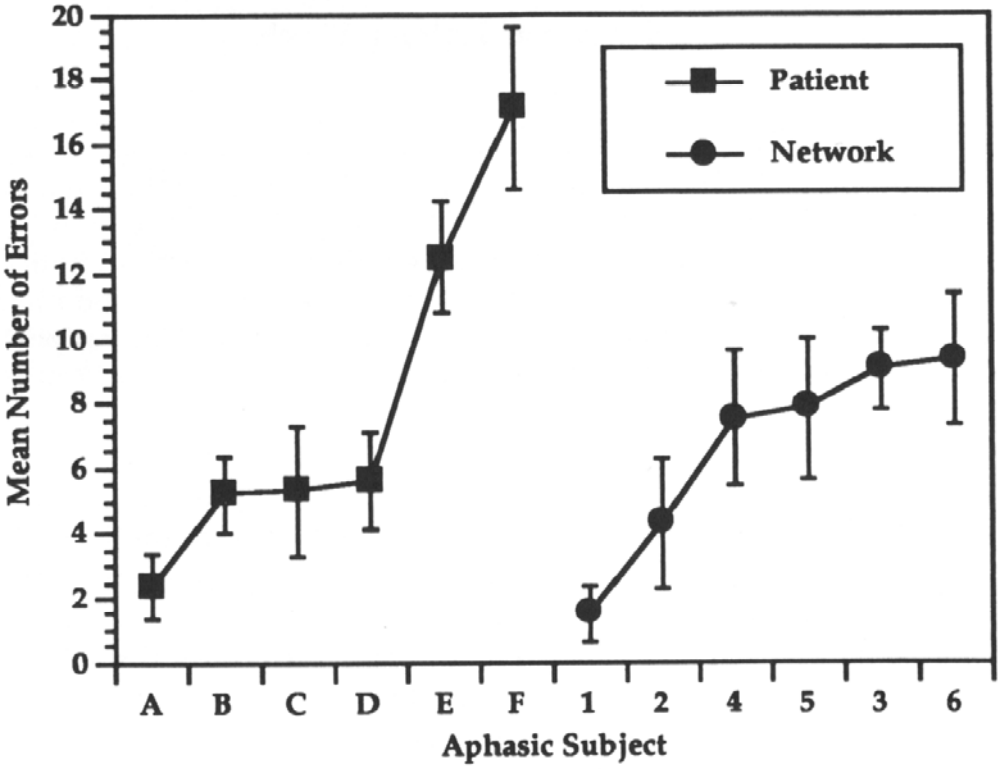

Figure 6. The results of the computational experiments demonstrate that the mean number of naming errors by the human subjects and by the networks are similar. The graph also shows that the standard deviations of this error are likewise comparable.

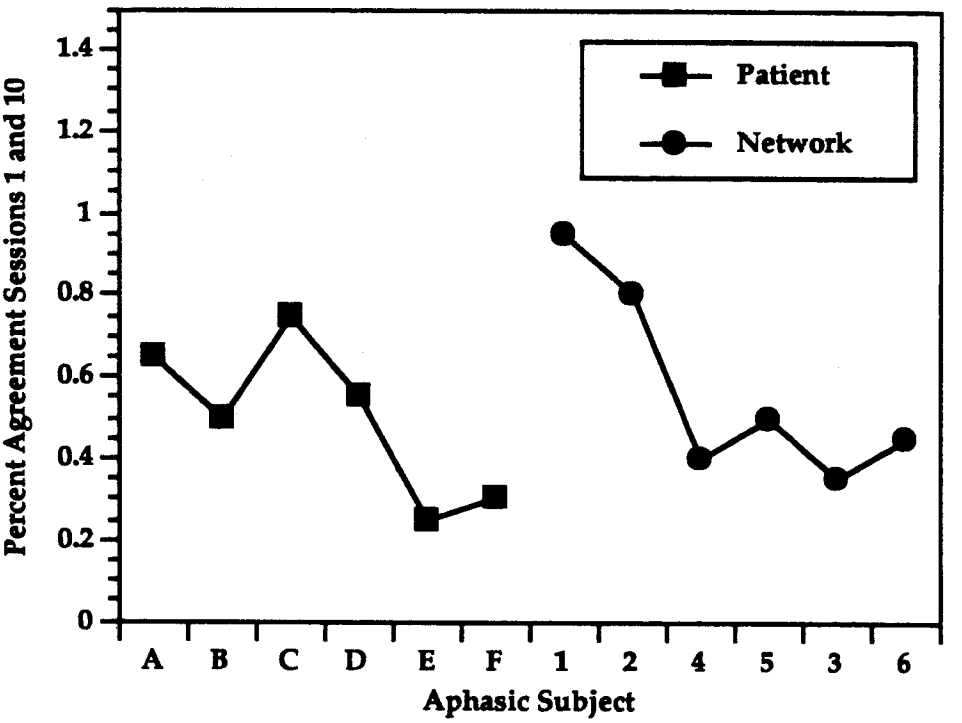

Figure 7. Response variability can also be demonstrated by comparing specific results of two different testing sessions. This graph shows the percent agreement between the 1st and 10th test sessions for the human subjects and the networks. The human and artificial subjects have comparable performance.

ment between the answers given by each subject during the 1st and 10th test sessions. These graphs demonstrate that within the constraints of a simple computational model, it is possible to simulate response variability such as that demonstrated by Crisman's subjects.

The increased error by aphasic subjects on low-frequency words was not reproduced in the model. However, Crisman's words were not normalized for length and a large proportion of the difference in number of errors was caused by sound errors, which were not addressed by the model. In addition, the data show that one human subject was significantly worse in average performance than any of the computer networks, yet the behavioral variability of this subject between Session 1 and Session 10 was similar to that of the networks.

CONCLUSIONS

The conclusions of this study depend on how the model works, and what its structure and function imply for understanding response

variability in aphasic patients. To simulate performance variability, one very simple change was made to the typical neural network architecture. Instead of using precise numerical values in the simulated neurons and synapses, we made these values imprecise, subject to a percent error in either direction.

The neurobiological speculation arising from this computational experiment is that, in the human brain, such variables as axonal conduction and synaptic integration are likewise subject to random variation from the *milieu intérieur*, or internal homeostatic environment (Bernard, 1865), and that this leads to variability in performance. Clinical efforts to manipulate this environment, such as pharmacological intervention and behavioral training, could serve to maximize the amount of time a patient spends in a more favorable state. In the unimpaired computer network, and possibly in normal people, many tasks are so overlearned that response variability does not come into play.

REFERENCES

Belliveau, J. W., Kennedy, D. N. J., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., Vevea, J. M., Brady, T. J., & Rosen, B. R. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science, 254* (Nov. 1), 716–719.

Benson, D. F., & Geschwind, N. (1989). The aphasias and related disturbances. In R. Joynt (Ed.), *Clinical neurology* (pp. 1–28). Philadelphia: Lippincott.

Bernard, C. (1865). *Introduction à l'étude de la médicine expérimentale.* Paris.

Brain, W. R. (1961). *Speech disorders.* London: Butterworth.

Crisman, L. G. (1971). *Response variability in naming behavior of aphasic patients.* Unpublished masters thesis, University of Pittsburgh.

Damasio, H., & Damasio, A. R. (1989). *Lesion analysis in neuropsychology.* New York: Oxford University Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Elman, J. L. (1989). *Representation and structure in connectionist models* (Report No. CRL-TR-8903). San Diego: University of California.

Gigley, H. (1988). Process synchronization, lexical ambiguity resolution, and aphasia. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* (pp. 229–267). San Mateo, CA: Morgan Kaufmann.

Gordon, B., & Caramazza, A. (1985). *Lexical access and frequency sensitivity: Frequency saturation and open/closed class equivalence* (Tech. Rep. No. 5). Baltimore: The Johns Hopkins University.

Gordon, B., Hart, J., Lesser, R., Schwerdt, P., Bare, M., Fisher, R., Krauss, G., Uematsu, S., & Selnes, O. (1990). Individual variations in Perisylvian language representation. *Neurology, 40*(4, Suppl. 1), 172.

Hart, J., Jr., & Gordon, B. (1992). Neural subsystems for object knowledge. *Nature, 359,* 60–64.

Hinton, G. E. (1986). Learning distributed representations of concepts. In *Proceedings of the Eighth Annual Meeting of the Cognitive Science Society* (pp. 1–12). Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Shallice, T. (1989). *Lesioning a connectionist network: Investigations of acquired dyslexia* (Report No. CRG-TR-89-3). Toronto: University of Toronto.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*(1), 74–95.

Jackson, J. H. (1878). On affections of speech from diseases of the brain. *Brain, 1*, 304–330.

Kawamoto, A. H. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* (pp. 195–228). San Mateo, CA: Morgan Kaufmann.

Kertesz, A. (Ed.). (1983). *Localization in neuropsychology.* New York: Academic Press.

Kŭcera, H., & Francis, W. N. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Marie, P. (1906). The third left frontal convolution plays no special role in the function of language. *Semaine Médicale, 26*, 241–247.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: MIT Press.

McNeil, M. R., Odell, K., & Tseng, C.-H. (1991). Toward the integration of resource allocation into a general theory of aphasia. *Clinical aphasiology, 20*, 21–39.

Mozer, M. C., & Behrmann, M. (1989). *On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia* (Report No. CU-CS-441-89). Boulder: University of Colorado.

Ojemann, G., Ojemann, J., Lettich, E., & Berger, M. (1989). Cortical language localization in left, dominant hemisphere: An electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery, 71*, 316–326.

Petersen, S. A., Fox, P. T., Mintun, M. A., Posner, M. I., & Raichle, M. E. (1988). Studies of the processing of single words using averaged PET measurements of CBF change. *Nature, 331*, 525–529.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation* (Report No. ICS-8506). San Diego: University of California.

Sejnowski, T., Koch, C., & Churchland, P. (1988). Computational neuroscience. *Science, 241*, 1299–1306.

Small, S. L. (1990). Learning lexical knowledge in context: Experiments with recurrent feed forward networks. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 479–483). Hillsdale, NJ: Erlbaum.

Small, S. L. (1991). Focal and diffuse lesions of cognitive models. In *Proceedings of the Thirteenth Annual Meeting of the Cognitive Science Society* (pp. 85–90). Hillsdale, NJ: Erlbaum.

Small, S. L. (1994). Connectionist networks and language disorders. *Journal of Communication Disorders, 27*, 305–323.

Small, S. L., Cottrell, G. W., & Tanenhaus, M. K. (Eds.). (1988). *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence*. San Mateo, CA: Morgan Kaufmann.

Small, S. L., Hart, J., Jr., & Gordon, B. (1992). *Automated boxology: A modular connectionist approach to modelling with application to picture naming*. Paper presented at the Annual Meeting of the Academy of Aphasia, Toronto, Ontario.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Uematsu, S., Lesser, R., Fisher, R. S., Gordon, B., Hara, K., Krauss, G. L., Vining, E. P., & Webber, R. W. (1992). Motor and sensory cortex in humans: Topography studied with chronic subdural stimulation. *Neurosurgery, 31*, 59–71.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.

Zipser, D., & Anderson, R. A. (1988). A back propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature, 33*, 679–684.