Journal of the International Neuropsychological Society (2000), **6**, 580–583.

SYMPOSIUM

Numbers, models, and understanding of natural intelligence: Computational neuroscience in the service of clinical neuropsychology

PAUL KOCH AND GERALD LEISMAN

What we call computational neuroscience involves construction of mathematical and numerical models for understanding cognitive phenomena. This issue is devoted to showing how it can also be used to help in the analysis of cognitive defects. Although the models may seem abstract to clinicians, they are based on the reality of brain anatomy. The theoretical papers presented here are *connectionist*: They posit a network of *cells* connected by *synapses* whose *weights* are modified during *learning*. Architecture of connectionist models has progressed and ramified considerably since they were first introduced, and we include some examples of the current state of the art. The final work presented here is concerned with the connection of the constructed models with clinical experience and experiment.

In the simplest neural networks, the unit cells are very much like neurons: they "fire," that is, emit a signal to their efferent connections, if they receive a sufficient signal from their afferent connections. All of these connections are weighted, and the weights are modified by a *learning rule*. Some of them are positive (excitatory) and some negative (inhibitory). The networks are constructed in three or more layers; existence of the *hidden layers* is necessary in order to achieve computational ability. When two layers are connected, every cell in one layer is connected to every cell in the other.

Early limitations of computer abilities led to elaboration in the models. In order to be useful, for example, in pattern-recognition tasks, the nodes of the network were given a semantic content that is probably not typical of a single neuron. At the same time, there was an attempt to duplicate the *parallel distributed processing* known to be typical of the brain by attributing the content to the state of activation of all the nodes rather than a single one.

With modern computers, is it possible to construct a simple neural network, namely, one without an external semantic overlay that can exhibit real cognitive ability? If it has not been done, someone will soon try to do so, but the number of cells involved would be enormous. For example the cortex of a frog contains 6,000,000 neurons. However, the synapses in actual neural tissue have a range beyond which the connection probabilities decrease exponentially with distance. A maximum of 10,000 neurons is contained in a cortical cylinder with the radius of 1 connection range.

Thus if we are interested in the links between brain and behavior, we must abandon the simple model described above. Given this, the work can proceed in numerous directions. The first paper, by Grossberg, describes a particularly fruitful approach, known as *adaptive resonance theory* (ART), and shows how it can be applied to hallucinations, which is a very general cognitive defect. An ART network is hierarchical, containing several modules with different functions. To us, its greatest strength as a cognitive model is that the system, once constructed, functions mechanically and still has a high level of practical success. It has the weakness that both the cells and some of the mechanisms have semantic content, which must be further explained on some other level. Grossberg points to the connections between the V2 and V1 layers of the visual cortex as an example of hierarchical anatomical synapses. This gives increased credibility and power to the ART approach, particularly as regards the vertical connections within and between brain regions.

On the other hand, the existence of *on-center off-surround* lateral interlayer connections (inhibitory connection range greater than excitatory), while necessary to the functioning of ART, cannot be given anatomical significance until the semantic content of the nodes is explained.

https://doi.org/10.1017/S1355617700655078 Published online by Cambridge University Press

When the activity of neurons is measured during cognitive tasks, they are shown to respond to fragments of information, such as pleasant or unpleasant expression, rather than target entities, such as recognizable faces. The neurons whose activities are measured are perforce selected at random, but the measured results can be likened to the output of anatomical neural networks. Semantic content must somehow arise by interconnection of these fragments, simultaneously anatomical and cognitive, by definition separated from each other by distances large compared with the connection range.

Continuum theory of neural tissue, described in the second paper by Leisman and Koch, provides one possible mechanism of connecting distant neural elements. These are defined as containing large numbers of cells, so that they can well represent the above-mentioned fragments. Using well-known methods of physics, we show that the neural medium, simply described in connectionist terms, can support waves of activity that grow in time and have a well-defined wavelength considerably larger than the connection range. Fragments separated by an integral number of favored wavelengths can thus be connected into a semantic entity.

How the content is perceived cannot be explained in the theory; what we try to do is simply to explain how enough cells to exhibit semantic ability can be organized together in spite of the limitations of connection range. We apply the theory to simple models of hippocampal and cortical structure to try to examine the mechanisms of memory implantation and recall. (In relation to ART we are interested in how one of the nodes in any layer constructs and reconstructs its stored pattern from its elements.) By implication, we also touch on amnesia, in its various manifestations, as a breakdown in one or more of the elements that must be present for memories to be implanted and/or recalled.

One significant way in which the connections afforded by continuum waves differ from connections in a simple neural network is the absence of causality. The activation of one element does not directly contribute to another; rather both are made active by participation in the same wave. The network described by Geva and Peled has features suggestive of the same idea.

The semantic content in their neural network is expressed as a combination of the activation values in a collection of "neurons." Different memories are characterized by different activation patterns. There is no geometry involved in this model, but it is not far-fetched to liken these patterns to activity waves and consequently the "neurons" to continuum elements. There is an element of causality in the direct connections among the neurons, but the weights of these connections do not change and are strongly proportional to the semantic activation patterns.

The Geva–Peled model also resembles ART in that it posits an external input pattern and a set of output patterns, which can be called expectations. Its dynamics however are quite different. The firing pattern of the neurons in the network is adjusted to be equivalent to one of the preset semantic patterns by a dynamic threshold that results eventually in the extinction of those neurons that do not fit into the pattern. Asymmetric connections are devised to make expected patterns more likely to occur. By varying the threshold parameters they are able to simulate (*metaphorically*) various cognitive disturbances, as they become manifest in various cognitive tasks.

Geva, Shtram, and Policker provide an illustration of a *Boltzmann network*, in which cell activation is given a probabilistic component. The transition to deterministic connectionism is brought about by a gradual decrease in the *temperature*, which, in analogy to statistical physics, is a measure of the allowed departure from *equilibrium*. This latter state is defined as the minimum of an *energy* function, involving the weights and the expected values of the node activations (which approach deterministic values). There are several such minima, each representing a state that can be given semantic meaning. Transitions among these states (i.e., learning) can be achieved by an *annealing* process in which the temperature is given a finite value and then slowly allowed to decrease.

Their particular network is constructed for a typical purpose: connecting words to their semantic categories. A word is represented by its orthographic fragments and the elements of the semantic representation include associated descriptive features. The learning task results in a final activation state of the semantic (output) layer that contains the attributes of a word presented in terms of its graphic decomposition to the input layer. Of interest here is the effects found when artificial "lesions" are imposed on the original model.

The content attributed to the nodes in this network again suggests that the elements are actually composed of many neurons. This means that the nature of the causal connections between these nodes needs further elaboration to be given anatomical meaning. Nonetheless, the effect on learning of various defects in the connections or functions of such a model can be of importance in the interpretation of real cognitive dysfunction.

It has emerged from this discussion that there are several scales of organization in the brain that must all be accommodated in successful cognitive modeling. Locally, neurons with no semantic content other than their firing state are connected into networks capable of storing fragments of information. These fragments are in turn connected, by continuum waves or some other mechanism, into nodes or modules capable of semantic processing. The processing mechanism may very well be one of those described here. Clearly, there must also be a vertical hierarchy of connections among and within the different brain regions.

The remaining paper in this collection examines one end of this segmentation in scale. The original and convincing work by Lloyd is aimed at showing that brain dysfunction is often correlated to lesions in many brain areas, and in turn that local brain lesions are associated with many different types of dysfunction. Lloyd's results are based on analysis of many PET experiments during cognitive tasks, using

https://doi.org/10.1017/S1355617700655078 Published online by Cambridge University Press

a method of *virtual lesions* by which damage in different areas is associated with mental defects. This strongly implies cognition must indeed be most often associated with widespread organized neural activity. The large-scale organization must have a basis other than direct synaptic connections of finite range. Continuum theory and activity waves may provide the link, and it would be of value if it were given an experimental basis.

It is our belief that, when the different spatial scales are integrated, cognitive modeling on connectionist principles will lead to a unified theory of the link between brain and behavior. In the meantime, we hope that the many uses of the work presented here illustrate that this endeavor is worthwhile and worthy of the attention of clinicians.

ACKNOWLEDGMENTS

This work was funded in part by a grant-in-aid to the second author from the Foundation for Allied Conservative Therapies Research.

Paul Koch
School of Engineering
New York Institute of Technology
Old Westbury, New York

Gerald Leisman
Institute for Biomedical Engineering and
Rehabilitation Services
Touro College
Bay Shore, New York