Computer Models:
A New Approach to the Investigation of Disease

James A. Reggia
Depts. of Computer Science & Neurology, Inst. Advanced Computer Studies
A.V. Williams Bldg., University of Maryland, College Park, MD 20742, USA
Email: reggia@cs.umd.edu

Eytan Ruppin
Depts. of Computer Science & Physiology, Tel Aviv University
Tel Aviv 69978, Israel
Email: ruppin@math.tau.ac.il

Rita Sloan Berndt
Dept. of Neurology, University of Maryland School of Medicine
Baltimore, MD 21201, USA
Email: rberndt@umabnet.ab.umd.edu

Abstract: During the last several years there has been a growing interest in developing computational models of phenomena associated with brain and cognitive disorders. Work in this area has included neural modeling studies of Alzheimer's disease, aphasia and dyslexia, epilepsy, stroke, Parkinson's disease, schizophrenia, depression, and related problems. Here we suggest that this computational work represents a new research paradigm for the understanding of disease, complementing traditional methods such as clinical studies and animal models. While computational models have so far focused most prominently on neurological and psychiatric disorders, there is no reason that this approach cannot be extended productively to any area of medicine.

Acknowledgement: Work supported by NINDS awards NS35460 and NS29414, and NIDCD award DC00699.

1. Introduction

During the last few years neural modeling methods have gained increasing visibility in clinical computing. Perhaps most widely appreciated is their use as a method for supporting medical decision making. For example, neural networks have been used to classify ultrasound images [1], diagnose acute myocardial infarctions [2], localize neurological damage [3], and aid in the development of new cancer chemotherapeutic agents [4], to mention just a few examples. These applications view neural models as simply one of a growing number of methods for computationally processing clinical data.

There is a second way in which neural modeling is being used in clinical computing: to investigate brain and cognitive disorders. This work uses "lesioned" neural models to study various brain and cognitive disorders from a computational point of view. The term "lesioned" is interpreted broadly here to mean that an intact model's structural or functional mechanisms are disrupted in some fashion. The goals of this research have been to construct computational models that can explain how specific neuroanatomical and pathological changes can result in various clinical manifestations, and to investigate the functional organization of the symptoms that result from specific brain pathologies. Table 1 lists examples of topics that have been studied in this fashion during recent years.

Neural models necessarily simplify the biological phenomena occurring in the nervous system and are generally constrained in size. The simulated lesions in such models are also substantial simplifications of pathophysiological events occurring within the brain and/or in cognitive processes. Nevertheless, such computer-based models complement traditional methods of studying brain disorders in substantial and important ways. Lesion size and location can be controlled precisely and can be systematically varied over arbitrarily large numbers of experimental "subjects" and information processing tasks. Further, computational experiments are open to detailed inspection in ways that biological systems are not, permitting one to determine the underlying mechanisms and to trace the behavioral changes following a lesion.

Table I: Some Previously Modeled Disorders

memory impairment	random lesions to neural networks amnesia agnosia Alzheimer's disease semantic memory loss structural vs. functional damage
symptoms of language impairment	aphasia acquired dyslexia acquired dysgraphia
focal structural damage	cortical deafferentation focal cortical damage, stroke
"functional" neurological disorders	epilepsy Parkinson's disease migraine aura delerium
psychiatric disorders	schizophrenia paranoia depression

In June 1995, about 150 individuals from around the world met at a workshop we organized at the University of Maryland to discuss and debate issues in this area. The shared enthusiasm for this new approach to studying clinical disorders led to a collection of twenty-two papers based on presentations made at the meeting [5]. The publisher of this collection, World Scientific, has kindly given us permission to present three of these papers here in an edited and abbreviated form. In the following, we briefly summarize some past work so that the reader can examine these papers in the context of related work.

2. What has been done?

Table I indicates the wide range of brain and cognitive disorders that have been studied computationally during recent years. We have recently reviewed this work in detail [5; Chapter 1], and provide only a brief summary here. For convenience we divide the models that have been studied into four categories involving memory, language, neurology, and psychiatry.

Memory Disorders

Substantial efforts have gone into developing models of memory disorders. Perhaps the earliest systematic lesioning study of brain models was done with simple associative memories [6]. These models were found to be fault tolerant in that small lesions tended to have little effect. As lesion size increased, performance gradually declined. On the other hand, removal of any specific neuron produced a roughly similar deficit in performance as removal of any other neuron. These results led to the conclusion that lesioning of associative networks of simulated neurons produced effects resembling Lashley's "mass action" and "equipotentiality". Subsequent work with similar but more sophisticated networks of neurons demonstrated that if lesions were made incrementally rather than all at once while synaptic modifiability was allowed to continue, then there was much less deterioration in performance for a given lesion size. This latter result is intriguing because it is consistent with observations that slow brain damage produces less of a behavioral deficit than rapid damage of equal size.

Many models of memory impairment have simulated aspects of Alzheimer's disease, the most common dementing illness. Recent neuroanatomical investigations have repeatedly demonstrated that the progress of Alzheimer's disease is accompanied by considerable synaptic changes, including synaptic deletion and compensation. While *synaptic deletion* is manifested in a reduction of the number of synapses per unit of cortical volume, a concomitant increase in the size of the remaining synapses has also been observed, i.e., *synaptic compensation*. Motivated by these experimental observations, some neural models have examined how the interplay between synaptic deletion and compensation determines the observed patterns of memory deterioration, and what strategies of increased synaptic efficacy could best maintain memory capacities in the face of synaptic deletion [7]. This work has shown that the deterioration of memory retrieval due to synaptic deletion can be much delayed by strengthening the remaining synaptic weights by a uniform compensatory factor. Variations on the rate and exact functional form of synaptic compensation were used to define various compensation strategies, and these could account for the observed variation in the severity and progression rate of Alzheimer's disease.

Other modeling studies have taken a different approach by examining *runaway synaptic modification* [8]. This forms the subject of the paper by Michael Hasselmo in this issue. Runaway synaptic modification denotes a pathological exponential growth of synaptic connections that may occur due to interference by previously stored memory patterns during the storage of new patterns. This interference occurs because, when a new memory pattern is being stored in the network, the resulting network activity is not only guided by the new pattern but also by all the previous memory patterns which are engraved in the synaptic matrix. Thus, previously memorized patterns tend to bias the activation in "their direction" during new storage. This inherent reinforcement may lead to exponential synaptic growth and to a pathological increase in the number of synapses.

What is the hypothesized role of synaptic runaway in the pathogenesis of Alzheimer's disease? Analysis shows that there is a critical storage capacity beyond which interference during learning cannot be prevented and synaptic runaway is unavoidable [8]. Several factors can lead to the initiation of synaptic runaway, such as a decrease in the level of cortical inhibition and reduced synaptic decay. Once synaptic runaway occurs, it is claimed that its increased metabolic demands or excitotoxic effects could be sufficiently severe to cause neuronal degeneration, parallel to that found in Alzheimer's disease. This work on synaptic runaway is an excellent example of research that combines experimental physiological studies with computational modeling. It demonstrates how a computational model can raise a quandary (how are patterns actually stored without being accompanied by synaptic runaway?) that motivates an experimental study (the differential effects of acetylcholine on internal and external synapses).

Language Disorders

Neurologists and neuropsychologists have identified a wide variety of language disorders that are associated with both focal and diffuse brain lesions. Neural models have been developed to simulate normal language functions and then lesioned to simulate the ways in which language is disrupted by brain injury. Although these cognitively-oriented "connectionist models" often use a neuron-like processing structure, they are generally *not* brain models in

the sense that they do not directly model specific neuroanatomic structures and neurophysiological functions. They are best viewed as simulations of cognitive functions and related phenomena that emanate from brain function.

Several computational models of *aphasia* (disordered spoken language) now exist. Although some of the earliest attempts to model aphasic language focused on disturbances of sentence processing, more recent work has dealt with simple transcoding of information at the single word level. Many of these efforts have been concerned with the issue of how information of different types (e.g., phonological, semantic) serves to trigger word responses under different conditions of stimulation: picture naming, word repetition, or written word naming. For example, Dell and colleagues simulated the individual error type distributions of 21 fluent aphasic patients by altering a neural model's parameters such as the decay rate [9]. Once the model was individually fitted to each aphasic patient's data, predictions were made about a number of other aspects of patient performance.

Much more computational work has been done involving reading disorders. Four distinct types of acquired *alexia* (or *dyslexia*) have been simulated using connectionist models. Two of these types are characterized by special difficulty reading unfamiliar or "made-up" words; when such "non-words" are misread, patients frequently produce a visually similar real word. Among these patients with non-word reading impairment, *deep dyslexics* also demonstrate a variety of problems reading aloud real words, particularly abstract words, and they often produce errors that are either semantically or visually related to the target. *Phonological dyslexics*, who are also impaired in non-word reading, do not produce semantic errors when reading words, and in fact may read real words very well. In contrast to these patterns of reading symptoms, *surface dyslexic* patients can read aloud non-words and many real words, but have difficulty reading words with irregular spelling/sound correspondences.

Several of the characteristics of deep dyslexia have been simulated in a connectionist model that links orthography (word form) and meaning (semantic features) [10, 11]. In the initial implementation, this model did not attempt to simulate non-word reading (or its impairment), but did perform lesion simulations that succeeded in producing errors that were

semantically related to the targets. The lesioned model also produced other interpretable error patterns that have been argued to occur in deep dyslexia, including errors visually related to the target and a higher proportion of "mixed" (semantic/visual) errors than would be expected by chance.

A different model of normal word recognition and pronunciation has been lesioned to simulate some of the characteristics of surface dyslexia [12]. The initial lesion simulations succeeded in producing better performance with regular than irregular words, but failed to reproduce the preserved non-word reading and regularization errors that are characteristic of surface dyslexic reading.

Still another approach involved using a dual-route reading model [13]. This model takes input in the form of pre-segmented "grapheme" units - one or more printed characters that serve as the written representation of a single phoneme. It had two parallel "routes" through which activation flowed in the network. The non-lexical route provides for the mapping of graphemes onto their corresponding phonemes by weighted links based of the probability of association between specific graphemes and phonemes. The lexical route consists of connections from the same grapheme nodes to a set of word nodes, and connections from the word nodes to phoneme nodes. Most importantly, performance patterns characteristic of both surface and phonological dyslexia were produced by the model when the lexical and non-lexical routes, respectively, were degraded, consistent with current theory.

Neurological Disorders

We now turn to models of neurological disorders such as stroke, epilepsy, migraine and parkinsonism. This class of models is more closely based on anatomical structures and physiological processes occurring in the brain than the models discussed in the preceding section. For example, there have been numerous efforts to develop computational models of the cerebral cortex. These models typically are concerned with simulating map formation in primary sensory cortex. Each primary sensory region of the cerebral cortex has a "map" of relevant aspects of the external world (e.g., the homunculus in somatosensory cortex).

It has repeatedly been demonstrated experimentally that such maps are highly plastic in adult animals: they undergo reorganization in response to loss of sensory inputs and focal cortical lesions [14]. For example, following a peripheral nerve lesion that deprives part of the somatosensory cortex of its input, the somatosensory map reorganizes to reuse the area of cortex that has lost its primary input. The map shifts so that the deafferented part of the cortex comes to represent other parts of the body surface. Recent noninvasive studies have suggested that such plasticity is also found in human cortical maps.

Computational models of the hand region of primary somatosensory cortex have been partially deafferented by removing a portion of their input connections to cortex, e.g., by removing input from the palm surface of the first two fingers to simulate a median nerve lesion [15]. When such deafferentation is done, the cortical map spontaneously reorganizes in a fashion similar to what is observed in experimental studies. Similar models of cortical deafferentation have been used more recently to support a theory of *phantom limb* experiences. Manfred Spitzer has combined a model of focal cortical deafferentation with input noise to account for various observations related to phantom limbs. This work is described in his paper later in this issue.

Recent work has also developed successful computational models of acute focal lesions in cortex (simulated strokes), again involving the hand region of primary somatosensory cortex [16]. When a focal lesion is introduced into the simulated map, the model reorganizes such that the sensory surface originally represented by the lesioned area spontaneously reappears in adjacent cortical areas, as has been seen experimentally in animal studies. Two key hypotheses emerged from this modeling work. First, that post-lesion map reorganization is a two-phase process, consisting of a rapid phase due to the dynamics of neural activity and a longer-term phase due to synaptic plasticity. Second, that increased perilesion excitability is necessary for useful map reorganization to occur. Recent work has shown that these hypotheses remain valid in a much more complex computational model of primary motor cortex that controls the positioning of a simulated arm in three-dimensional space [17].

Other neurological disorders that have been modeled include epilepsy, parkinsonism, and

migraine. A model of part of the hippocampus, a brain region in which highly synchronized electrical activity and EEG spikes occur with certain types of epilepsy, has been modeled successfully [18]. This work provided insight into the mechanisms underlying abnormal neuronal bursting and afterdischarges. Other modeling work has shown how repeated simulated electrical stimuli can generate a new focus of abnormal, epileptogenic activity due to synaptic changes [19]. It has also been possible to simulate the emergence of parkinsonian tremor as an oscillatory state in a recurrent neural network [20], and to simulate movement abnormalities in both parkinsonism and Huntington's disease [21]. Finally, a neural model of the visual hallucinations occuring during the migraine aura has been studied [22].

Psychiatric Disorders

Neural models have also been created for a wide range of psychiatric disorders. Most attention has been given to schizophrenia, a clinically heterogeneous disorder with a broad spectrum of manifestations. Its symptoms are diverse, and include both "positive symptoms" such as hallucinations, delusions, disorganized speech and behavior, and "negative symptoms" such as loss of fluency of thought and speech, impaired attention, abnormalities in the expression and observation of emotion, and loss of volition and drive. The pathogenesis of schizophrenia is unknown.

Neural modeling of schizophrenia has taken multiple paths. The first avenue concentrated on modeling schizophrenic positive symptoms in the framework of an associative memory attractor network [23]. This work pointed to a possible link between the appearance of specific neurodegenerative changes and the emergence of 'parasitic attractors', states in which a neural network's normal processing becomes locked in dysfunctional patterns of activity. These parasitic states can be interpreted as the model's correlates of schizophrenic delusions and hallucinations. They are generated when the model's memory capacity is exceeded. In this current journal issue, Ralph Hoffman describes how excessively pruned attractor networks behave in a fashion suggestive of loose associations and delusions, and how pruned backpropagation networks can produce spontaneous outputs suggestive of hallucinated voices. A similar neural model has recently been used to examine a theory that schizophrenia is as-

sociated with reactive sprouting and synaptic reorganization taking place in the projection sites of degenerating temporal neurons [24]. In this model, it is shown that while preserving memory performance, compensatory synaptic regenerative changes may lead to adverse, spontaneous activation of stored patterns (i.e., simulated delusions/hallucinations). The incorporation of Hebbian activity-dependent synaptic changes into the model leads to a *biased* retrieval distribution that is strongly dominated by a single memory pattern.

A different approach has been taken by Cohen and Servan-Schreiber [25]. They have presented a comprehensive modeling study of the performance of normal subjects and schizophrenics in three attentional and language processing tasks. The effects of dopamine on information processing (which may play a major role in the pathogenesis of schizophrenia) were modeled as a global change of the input gain. These simulations demonstrated that a change in the gain of neurons can quantitatively account for the differences between normal and schizophrenic performance in the tasks examined.

Neural networks have been used to model a variety of psychiatric problems other than schizophrenia. For example, manic hyperactivity has been modeled as an increase in noise levels resulting in an enhanced rate of transition between attractors [23]. Some aspects of major depression have also been the focus of recent modeling work [26].

3. Prospects

As the above summary has illustrated, the last several years have witnessed the development of a wide variety of interesting computational models of neurological, neuropsychological and psychiatric disorders. These models have already demonstrated that a small set of fairly simple assumptions can account for phenomena seen in recovery from brain damage and impaired cognitive information processing. They have tested various hypotheses about underlying pathophysiological processes, at times making testable predictions, and at times they have even suggested ideas for therapeutic/rehabilitative steps.

Of course, as noted earlier, this past work is not without its limitations. Most previous neural models greatly simplify the true psychological/biological phenomena being studied.

Many of the models discussed here are also quite small in scope. Even with these simplifications, many are still computationally very expensive to run on conventional computer architectures (e.g., they require very large amounts of processing time). Further, past work in this area has only attempted to model a small fraction of the neurological and psychiatric problems that are potentially amenable to study with neural networks.

With all of these limitations, why has there been such a rapidly growing interest in using neural models to study brain and cognitive disorders? We believe that there are a number of reasons for this. Traditional medical investigative techniques (clinical trials, animal experiments, etc.) are expensive and face substantial barriers in terms of what can be practically or ethically studied. Computational models, on the other hand, are relatively inexpensive and provide for large numbers of "subjects" without ethical concerns. They allow one, in theory, to vary any aspect of a system to assess its effects, and to record virtually any variable from a system without interfering with the system's behavior. Ultimately, computational models may prove the most effective way to understand some of the underlying mechanisms of brain and cognitive disorders. They may become important factors in suggesting new treatments or in their preliminary assessment. To a great extent, it is this hope of deeper understanding of pathophysiology and of potential therapeutic guidance that motivates much of the current modeling work in this area.

In the past, studying disease has largely meant either clinical investigations involving human subjects, or the use of animal models. It seems apparent that computational models offer a new paradigm for the investigation of disorders that complements these more traditional experimental methods. By evaluating physiological or therapeutic hypotheses in detail, computational models offer a genuinely new approach to the study of disease. While the ultimate answers will still lie with clinical/biological experimentation, computational models may help lead us to those answers more quickly and less expensively. Further, while we have considered only neurological and psychiatric disorders here, it is reasonable to suppose that similar approaches could be used effectively in any area of medicine. Of course, some modeling work has been under way in other areas, such as the investigation of cardiac ischemia [27]. Much remains to be done though, and we expect exciting developments in

this area during coming years.

8. References

[1] Pan H. and Chen Y. Liver Tissue Classification by Artificial Neural Networks, *Pattern Recognition Letters*, 13, 1992, 355-368.

[2] Baxt W. Use of an Artificial Neural Network for Data Analysis in Clinical Decision Making, *Neural Computation*, 2, 1990, 480-489.

[3] Tuhrim S., Reggia J., and Peng Y. High-Specificity Neurological Localization Using a Connectionist Model, *Artificial Intelligence in Medicine*, 6, 1994, 521-532.

[4] Weinstein J., Kohn K. and Grever M. et al. Neural Computing in Cancer Drug Development: Predicting Mechanism of Action, *Science*, 258, 1992, 447-451.

[5] Reggia J., Ruppin E. and Berndt R. (editors). *Neural Modeling of Brain and Cognitive Disorders*, World Scientific, 1996.

[6] Wood C., Variations on a Theme by Lashley: Lesion Experiments on the Neural Model of Anderson, Silverstein, Ritz and Jones, *Psychological Review*, 85, 582-591, 1978.

[7]Horn D., Ruppin E., Usher M. and Herrmann M. Neural Network Modeling of Memory Deterioration in Alzheimer's Disease. *Neural Computation*, 5:736–749, 1993.

[8]Hasselmo M.E. Runaway Synaptic Modification in Models of the Cortex: Implications for Alzheimer's Disease. *Neural Networks*, 7(1):13–40, 1994.

[9] Dell G, Schwartz M, et al. A Connectionist Model of Naming Errors in Aphasia, in [5], pp. 135-156.

[10] Hinton G. and Shallice T. Lesioning an Attractor Network: Investigations of Acquired Dyslexia, *Psychol. Review*, 98, 1991, 74-95.

[11] Plaut, D.C. and Shallice, T. Deep Dyslexia: A Case Study of Connectionist Neuropsychology, *Cognitive Neuropsychol.*, 10, 377-500, 1993.

[12] Patterson K., Seidenberg M. and McClelland J. Connections and Disconnections: Acquired Dyslexia in a Computational Model of Reading Processes, in P. Morris (Ed.), *Connectionism: The Oxford Symposium*, Cambridge University Press, 1989.

[13] Reggia J., Marsland P. and Berndt R. Competitive Dynamics in a Dual-Route Connectionist Model of Print-to-Sound Transformation, *Complex Systems*, 2, 1988, 509-547.

[14] Jenkins W. & Merzenich M. Reorganization of Neocortical Representations After Brain Injury: a Neurophysiological Model of the Bases of Recovery from Stroke, in *Progress in Brain Research*, Vol. 71, Seil F., Herbert E. & Carlson B. (eds.), Elsevier, 1987, 249-266.

[15] Pearson J., Finkel L. and Edelman B. Plasticity in the Organization of Adult Cerebral Cortical Maps, *Journal of Neuroscience*, 7, 1987, 4209-4233.

[16] Sutton G., Reggia J., Armentrout S., and D'Autrechy C. Cortical Map Reorganization as a Competitive Process, *Neural Computation*, 6, 1993, 1-13.

[17] Goodall S, Reggia J, Chen Y, Ruppin E & Whitney C. A Computational Model of Acute Focal Cortical Lesions, *Stroke*, 1997, in press.

[18] Traub R. Models of Synchronized Population Bursts in Electrically Coupled Interneurons, *J. Comp. Neurosci.*, 2, 1995, 283-289.

[19] Mehta M., Dasgupta C. & Ullal G. A Neural Network Model for Kindling of Focal Epilepsy, *Biol. Cybernet.*, 68, 1993, 335-340.

[20] Borrett D, Yeap T & Kwan H. Neural Networks and Parkinson's Disease, *Canad. J. Neurol. Sci.*, 20, 1993, 107-113.

[21] Contrevas-Vidal J., Teulings H, & Stelmach G., *A Neural Network Model of Movement Production in Parkinson's Disease and Huntington's Disease*, in [5], pp. 377-392.

[22] Reggia J. & Montgomery D. A Computational Model of Visual Hallucinations in Migraine, *Comput. Biol. Med.*, 26, 1996, 133-141.

[23] Hoffman R. Computer Simulations of Neural Information Processing and the Schizophrenia-Mania Dichotomy, *Arch. Gen. Psychiatry*, 44, 1987, 178-188.

[24] Ruppin E, Reggia J & Horn D. Pathogenesis of Schizophrenia Delusions and Hallucinations: A Neural Model, *Schizophrenia Bulletin*, 22, 1996, 105-123.

[25] Cohen J. and Servan-Schreiber D. Context, Cortex and Dopamine: A Connectionist Approach to Behavior and Biology in Schizophrenia, *Psych. Rev.*, 99, 1992, 45-77.

[26] Luciano J, Cohen M & Samson J. Modeling Unipolar Depression Recovery, in [5], 1996, 469-483.

[27] Li J., Wang J and Drzewiecki G. Computer Modeling of Non-Adjacent Regional Ischemic Zones on Ventricular Function, *Comput. Biol. Med.*, 26, 1996, 371-383.