Talking Nets: An Oral History of Neural Networks

edited by James A. Anderson and Edward Rosenfeld

The MIT Press

From The MIT Press

MITCogNet

First MIT Press paperback edition, 2000

This book was set in Palatino on the Monotype "Prism Plus" PostScript Imagesetter by Asco Trade Typesetting Ltd., Hong Kong, and was printed and bound in the United States of America.

Photographs of Gail Carpenter and Stephen Grossberg by Deborah Grossberg. Photographs of James A. Anderson by Philip Lieberman. All other photographs by James A. Anderson.

Library of Congress Cataloging-in-Publication Data

Talking nets : an oral history of neural networks / edited by James A. Anderson and Edward Rosenfeld.
p. cm.
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-01167-0 (hardcover : alk. paper), 0-262-51111-8 (pb)
1. Neural computers. 2. Neural networks (Computer science)
3. Scientists—Interviews. I. Anderson, James A. II. Rosenfeld, Edward.
QA76.87.T37 1998
006.3'2'0922—dc21 97-23868
CIP

8 Stephen Grossberg

Stephen Grossberg is Wang Professor of Cognitive and Neural Systems; Professor of Mathematics, Psychology, and Biomedical Engineering; Chairman, Department of Cognitive and Neural Systems and Director, Center for Adaptive Systems, Boston University, Boston, Massachusetts. A good place to read further about Professor Grossberg's work is a 1995 paper, "The Attentive Brain," American Scientist *83: 438–449.*

July 1993, Portland, Oregon

ER: What is your date of birth and where were you born?

SG: December 31, 1939, in New York City.

ER: Tell us something about your parents and what your early childhood was like.

SG: My grandparents were from Hungary, around Budapest, so I'm a second-generation American. My mother was a school teacher. My biological father died when I was one. My mother remarried when I was five, and I was adopted.

My new father was an accountant. My mother was a devoted teacher, and she got her Ph.D. equivalency at a time when it was hard for a woman even to go to college. She very much influenced my religious attitude toward learning. We grew up first in Woodside, and then when my mother remarried, we moved to Jackson Heights, Queens, to a lower-middle-class neighborhood filled with upwardly mobile Jewish boys who were fiercely competitive.

I was always very interested in art and music.

ER: Was that a natural inclination, or was that something that was fostered at home?

SG: It came from within. I was drawing from a very early age. I used to win a lot of art prizes, including study at the Museum of Modern Art when I was in high school. As for music, I went to the neighborhood library and discovered they had racks of records. I discovered all the major composers there. That made me want to play piano, so my parents started to save and eventually bought a piano. I learned very quickly. I actually did a lot of things well and was always first in my class.

ER: Before you were first in your class, what were your early childhood experiences like? Did you have a brother and sister?

SG: You'll have to prime me on this sort of thing because I don't usually talk about myself. I usually talk about work or other people. I'm a middle child. My older brother is two years older. My younger brother is the child of the second marriage. He's six years younger than me. This is a difficult position to be in if you want to be a scientist, apparently—to be the middle child and also not the child of the living father. How this worked out, I don't know. I guess it worked out just because I have certain talents, and I worked incredibly hard.

I knew I didn't want the life that seemed imminent. I looked around, and I saw a lot of very nice people who seemed unhappy with their lives. I wanted to find a higher form of life. I used to think about it almost in religious terms, although I'm not what you'd call a traditional religionist if only because I'm too much of a loner. I don't like believing things just because other people believe them. I try to find a path toward some higher form of existence. This is really fundamental to my whole point of view.

I was very aware of the fact that living things are either growing or they're dying. I had a strong sense of the dynamics of life—you know, blooming and decaying. It was already clear to me when I was very young that we have a short time on earth. It was also clear that society creates barriers to choice and that I had to find a way to keep my options open broadly so that I could eventually figure out how I could touch something that was more enduring. This, to me, was a deeply religious feeling: how to be in touch with the enduring beauty of the world, even though you can only personally be here for a very short time. It seemed the only way to do that at the time, given my limited options because my parents had no money, was to be incredibly good in school.

ER: What were your early childhood experiences like?

SG: How early do you want?

ER: As early as you can remember.

SG: Oh, I can remember when I was one.

ER: You can remember when you were one?

SG: Well, I have one memory, and that was when they took my dying father from the house to the hospital. The big black bag of the doctor was right in front of my face. That's my only memory from so early. Later, I was lonesome. I was very shy. In fact, one of the hallmarks of oh, the first twenty-odd years of my life was extreme shyness. I also didn't have much experience with how to be a man because I never knew my first father, and my second father was very distant.

My mother was marvelous—is still a remarkable woman—but being Hungarian was not good at showing or responding to affection.... I don't

know if you know about Hungarians. They had the highest suicide rate in the world for many years, and one reason is that they wean you early from any show of affection. This isn't true of all Hungarians; for example, if you read the life of John von Neumann, you see that he had quite a different life because he formed a strong attachment to his father and was pampered by all the women in his extended family. One thing that made this possible was they had a great deal of money, and they led a privileged life.

We had little money. Although my parents were totally committed to their children and deeply loved us, there was a lot of stress, and little overt affection. My older brother was much more affected by my father's death than I was. That made him insecure and aggressive, and he used to beat me up regularly. That was frightening and made me feel vulnerable.

So I became shy and withdrawn, and—like a lot of people who are this way—became very creative, fantasized a lot, and tried to find another more appealing world. My world was the world of art and music at first and a world of trying to find approval. I found approval by trying to do very well in school, which also gave me great satisfaction because I was learning about things that often described other, more perfect, worlds.

In fact, many years later, when I read some of Einstein's essays, the phrase "the painful crudity and hopeless dreariness of daily life" stuck with me. Since that time, I've built a life with my family and friends that is happy, fulfilling, and full of meaning for me. But in my early childhood there was more painful crudity and hopeless dreariness because there were no examples of lives around me that I wanted to live. There were good people who were doing their best, but I viewed their lives as painfully crude and hopelessly dreary, so I had to find something that wouldn't feel that way to me, and from an early age I found it in learning more about the world. I also realized that I wanted to better understand why so many really nice people seemed unhappy, so I got very interested in people.

Also, you know, if you grow up in New York, there are only two major forms of life: people and dogs. You can't even see the sky half the time, so it wasn't as if you were in Nature and looking upon wonderful seasons and constellations. I therefore got very interested in the most interesting thing in New York, which was human life, and how people get on, how we come to know things about the world, and so on. From an early age I had a yearning to understand people, and I figured I would become a psychiatrist, as soon as I figured out what a psychiatrist was.

ER: And how early in your life did you start to draw?

SG: Oh, from the earliest years. I was drawing, oh, goodness, certainly well before I was five. First, I had all the usual coloring books, but then I started more active drawing, and I drew at a high level for my age. In fact, it became ridiculous when I was in public school and high school. I used to generate these large illustrated books that shocked my teachers because they were at a professional level.

The problem was that, even though I was in gifted classes, there was no one there to really help me build my confidence or move as fast as I could. Instead, it was a highly competitive environment. My public school, which was PS 69 in Queens, turned out to have kids with an unusually high overall cumulative IQ.

For example, we took standardized tests in the eighth grade, on which the highest you could go was the equivalent of having graduated from high school. It was called 12.0 plus, the twelfth grade plus, when you were in the eighth grade. The teacher foolishly read our scores out loud. It was 12.0 plus, 12.0 plus, 12.0 plus, until there was one poor kid with an 11.9—that kid was crushed. It was a sick environment. There was really little opportunity to enjoy being smart, apart from the fact that the satisfactions in learning were great, but the competitiveness was horrendous.

ER: You said you were also very involved with music, and I was wondering what form that took.

SG: Well, basically, my parents knew how much I wanted to play, and they were able to manage buying a little piano when I was in seventh grade. Within the year I was playing pretty advanced things—like Bach partitas, Gershwin's "Rhapsody in Blue," and lots of Chopin. My teacher called me a "genius," but I guess every music teacher tells parents that their child is a genius! One reason that I didn't go into music was I realized that, although I could play pretty well, I didn't have great hands; I also didn't have absolute pitch. I also tried composing some pieces for piano, and enjoyed this a lot but this still did not satisfy my yearning to contact enduring truths.

I wanted to do something where I could touch the eternal. I had this feeling that we're only here for a moment, and when we're not growing and helping others to grow, then we're dying. My hope was that the fruits of my mind might live longer than my body, and whatever understanding I could achieve would endure even as my body collapsed. So I very much needed to find something more enduring, more universal. These was this religious sense of needing to be in touch and commune with the world at an early age. This was my way of seeking a better future: to find a level of reality in life that could endure.

ER: Where did you go to high school?

SG: I went to Stuyvesant.

ER: Which is one of the New York schools for gifted children.

SG: It was either Stuyvesant or Bronx Science. Bronx Science was about an hour, an hour and a half away, and Stuyvesant was forty-five minutes, but in those days you could take the subway and feel safe. When I first attended Stuyvesant, it was a wonderful experience. I had some very good teachers, and I flourished in many ways there. But I was also aware of the terrifying statistics of the place. What do I mean by that? We had a class of maybe seven hundred kids. This was a time when there was still prejudice against

Jews in schools. There was a quota system. And only the top, a small segment of a place like Stuyvesant, would even get into college. Of course, you could go to CCNY [City College of New York] which brought out generations of great scientists. But I didn't even know about CCNY then.

Let me just give you an anecdote to set the stage. I remember going to a party after I graduated from Stuyvesant for kids from Bronx Science and Stuyvesant. One kid came up to me and said, "You're Grossberg, aren't you?" and I said, "Yeah," and he said, "I've hated you for four years."

I said, "But have we met?" and he said, "No, but I wished you would die."

I said, "Why?" and he said, "Because if you'd died, I'd be one higher in rank at the school."

I felt that all the time. There were several hundred kids who all had grade-point averages of around 92 percent—that's several hundred kids within fractions of a point from each other, which determined whether they got into college. I also knew kids who got three 800s on their college boards, but didn't get into any college to which they applied on the first round.

There were quite a few of us who had three 800s. I had three 800s on my boards, too, but I was also, fortunately, first in my class with 98-point-something average. So I succeeded within this system. I realized, though, that I couldn't stand this relentless competition much more. I needed out. I wasn't getting a chance to pursue my own goals, my own aspirations. My whole life was being spent on competing to escape, and I realized I had to find a way to get free from this relentless rat race fast. I didn't yet know what freedom meant, but I knew that I needed it to find out what I was going to do with my life.

Unfortunately, my family had no money with which to visit colleges. It was also a conceit of the time that, if you wanted to escape, you should try to go to an Ivy League school; that's where smart Jewish kids went. And so I started looking at Ivy League schools, and applied to several. Dartmouth had a senior fellowship program, which meant that if you were good enough in your classes, then in your senior year, you didn't have to take courses anymore. You'd have a free year to do research, whatever that was. That was one reason I applied to Dartmouth. Anyway, I got into a number of these schools with fellowships, including Harvard and Yale, but I got a bigger fellowship from Dartmouth, which was important because my parents needed the money.

In Dartmouth, my goal was to try to do so well that maybe I'd get a senior fellowship. I worked so hard at Dartmouth that many professors said that I was the best student that they ever had.

I was so highly motivated to find my way that, when I took Psychology 1, it unexpectedly created a storm of ideas in my mind. I got immensely engaged by human verbal learning data, animal discrimination learning data, and human attitude change learning. I was entranced by the implications of these data for how things are going on moment by moment in our minds—the kind of things that I still talk about: the real-time dynamics of individual

minds. I could see that studying mind brought together several of my yearnings.

First, it was a good way to better understand people. Second, it gave me a way to better understand the processes of adaptive growth and development that were so much a part of my view of the world.

In fact, just anecdotally, I don't know if you know who Stuart Kauffman is? [A MacArthur Fellow, now at the Sante Fe Institute.] Well, Stu and I were classmates at Dartmouth, and we met just before school started at an overnight hike where new freshmen got a chance to know each other. Stu and I found each other that first day, and got into this long philosophical debate having to do with the mind: how do you know and how do you see, etc.

I can't remember the details, but I do remember being up in a loft one night, and we were still talking away while other kids who were trying to sleep were saying, "Shhh, shhh."

Even then, what would happen in our debates was, we would be discussing some topic during which I would say something, and Stu would say, "But that's not philosophy" because, you see, both of us were deeply interested in philosophy; we were high school philosophers! I had always thought of myself as being interested in philosophy and trying to define large issues and how to understand them.

Stu went on to become a philosophy major in college, and then he went to England on a Marshall scholarship in philosophy. It was only later that he came around to my view that philosophy doesn't have the methods that we need, and then made a big switch to medicine and from there to his present research in evolutionary biology.

But already, as freshmen we were having this battle. I'd say, "But I don't care if it's philosophy or not; this is what I want to know, and I want to find the right tools to know it." I was already searching for tools to understand better how our minds know the world, so when I read classical psychological learning data—the data of Hull, Guthrie, Pavlov, and all these other people —they really changed my life.

That year (1957–58) after Psychology 1, I went through a major intellectual struggle trying to figure out how to represent the real-time processes underlying these learning data. That is when I introduced the so-called Additive Model, which later in 1984 was called the Hopfield model by various people who didn't know the literature of our field. By then I had published at least fifty papers on it.

This misattribution. You know, when I introduced this model, indeed this modeling framework, it was really original, because there was nothing like it in the field. AI [artificial intelligence] was itself barely formed in 1957. There was just nothing to turn to for guidance. One had to find one's own way. I derived a lot of guidance from the bowed serial position curve of human verbal learning. The bow reflects the fact that the middle of a list is often harder to learn than its ends. Why does it bow? Why is learning asymmetric between a list's beginning and end? When you have rest periods between

learning trials, why does the whole bowed distribution of errors change? Why do errors occur in the forward direction at the beginning of the list and the backward direction at the end of the list? To me, these data seemed extraordinary: first, that learning could go forward and backward in time; next, that silence between successive list presentations—the nonoccurrence of items in the future—could retroactively reorganize the entire distribution of learning. Events going backward in time excited me a lot and made me think about how to represent events in time.

I loved these data, and it was through studying them that I derived the Additive Model neural network with its short-term memory traces at network nodes, or cell populations, and its long-term memory traces in neural connections going forward and backward between these nodes, with the long-term memory traces at the synapses.

I think it's an interesting fact that I didn't know any neurophysiology when this model was derived. It was through quantitatively trying to understand the real-time dynamics of the serial position curve that I realized that there were short-term memory and long-term memory traces and competition among these distributed traces. I was also talking to my premed friends who told me about what they were learning about nerve cells, axons, synapses, transmitters, and the like when I realized that my model already had all of these properties.

I can hardly recapitulate my excitement when I realized this. It was such a passionate time. When it dawned on me that by trying to represent the real-time dynamics of behavior, you could derive brain mechanisms, I started reading neurophysiology with a vengeance. This first experience captures the story of my life as a thinker: To first try to understand behavior in a top-down way, always focusing on how behavior unfolds in real time, moment by moment, and trying to keep all homunculi out of the explanation. The model has to do it by itself, whatever its explanatory range. Such analyses have always made a link to neuroscience, and then computational and mathematical analysis showed how interactions among many neurons led to emergent properties that linked to behavior. Given the neural link, I'd then work bottom-up and top-down to further close the gap, pushing on both ends, between brain and behavior.

At that time, doing this work involved pretty extreme feelings of passion, terror, joy, and love. I was quite alone and pretty young—only 17 or 18—to be trying such a difficult path.

Anyway, to make a long story short, I did get a Senior Fellowship, and I spent my senior year continuing my research, including human verbal learning experiments. I knew that I had to make a difficult decision about what sort of career to follow. I loved psychology, and I view myself primarily as a psychologist and neuroscientist even today, rather than as a mathematician. I realized, though, that there were already many wonderful experimentalists but very few theorists. And I realized that, to be a good theorist, I needed mathematical techniques I didn't have, because from the first equations I

wrote down as a Freshman, when I was the deriving the Additive Model, I needed systems of nonlinear differential equations. I hardly knew any appropriate math for analysing these equations.

Before I derived the Additive Model, I was stimulated by Bill Estes's papers on learning models that were just coming out then. He used Markov models to describe his Stimulus Sampling Theory. In my analyses of serial learning, I remember trying to express some of the distributions of learned traces and errors by using Stimulus Sampling Theory. I finally managed to compute a formula that went on for pages. I then realized that this couldn't be the correct method. The results were uninterpretable and meaningless. After struggling very hard, I began to understand that there were both fast rates and slow rates hidden in the data. In this way, I was able to start teasing out short-term memory and long-term memory traces, network nodes, and directed paths between them.

The dynamics of these short-term memory and long-term memory made me start to use differential equations. This was all exciting, but also terrifying because, at first, I couldn't prove anything about these equations. After going through the model derivation phenomenologically and being very clear about the steps that led to the equations and qualitatively being able to argue why they should be able to explain the data, I couldn't prove it. Computers weren't there to help, either. I can jump ahead and say that when I went to Stanford to do graduate work, one of the first things I tried to do was to work with one of the top programmers there to help me program the model so that I could compute the distribution of errors. He wasn't able to do it, for one reason or another. That created a major problem and source of anxiety, because how do you convince people of something that you can't prove mathematically and for which there aren't any other computational tools?

By this time, I had qualitatively derived a lot of results about human verbal learning and about animal discrimination learning. I also had related ideas about the dynamics of attitude change. I had replaced statistical psychological models with neural network models, and was aware of the importance of competitive normalization and contrast gain control to link the two types of description together.

As this was happening, I became the first joint major in psychology and mathematics at Dartmouth. It was also made clear to me that I couldn't hope for a career in a psychology department at that time as a full-time theorist. One had to function primarily as an experimentalist. Even Bill Estes, I was told, had a lot of trouble getting his modeling papers published at first, even though he was already a distinguished experimentalist.

My equations for short-term memory and long-term memory were nonlinear, many-body, fast-slow systems of differential equations. This was challenging mathematics. I needed a way to make it look simple. Although I was, at first, more interested in human verbal learning and animal discrimination learning, I then saw how to derive the equations from simple ideas about

classical conditioning. That was exciting because both human and animal learning laws then had a similar form. These laws illustrated the type of universality that I was seeking.

Of course, none of these activities had anything to do with getting good grades. I became first in my class at Dartmouth for doing well in the standard curriculum. My research activities, in contrast, were not about getting grades; this had to do with how to be spiritually alive in the world. On the other hand, no one else was working to link brain to behavior with nonlinear neural networks. My intellectual work gave me a sense of purpose, but it also isolated me from my colleagues. Social acceptance and survival became a major issue, despite my intellectual success.

It was clear that I had to develop strong mathematical techniques in order to survive. I mean, how else could I prove anything? The computers weren't there. How else was I going to escape being considered a nut? At Dartmouth I was not considered a nut because I handed in one brilliant final exam after another, but I was still too shy to approach my professors personally. My own struggles to overcome my shyness have motivated me to set up an educational framework in our department that is designed to help students to be open and comfortable in their interactions with faculty.

It was not easy, while I was at Dartmouth, to figure out what to do with my life. One possibility was to become a mathematician because all science eventually becomes mathematics. If I could prove theorems about my neural models, then perhaps in that way I could continue my work.

But to become a mathematician when you really wanted to be a psychologist was no easy thing. I psyched myself into it with the following kinds of considerations. First, mathematics is a form of thinking, of cognitive processing. I tried to think of it as just one of the highest forms of cognition. This approach also helped me to better teach mathematics later on. Second, mathematics provided a way for me to learn large amounts of science fast, and I knew that I needed to learn a lot of science as part of my interdisciplinary training. I realized that if I opened a physics book on quantum mechanics, I'd either get stuck on trying to figure out how to read the equations, or I'd feel so comfortable with the language of mathematics that I could read the equations fluently and then be free to think about what the equations physically mean. Finally, I realized that I needed a virtuoso mathematical technique to express my own physical intuitions in an appropriate formalism, and then analyse the behavioral consequences of this formalism.

With these kind of intellectual rationalizations in mind, I decided to try to get a Ph.D. in mathematics. As you can imagine, I was pretty anxious about how all this would work out. Then the question arose as to where to go to graduate school. An advisor recommended that I go to Stanford because, at that time, Stanford had the strongest group in the world in mathematical psychology: Bill Estes was there, as were Gordon Bower, Dick Atkinson and Pat Suppes, among others. Stanford also had a strong department of applied mathematics.

So I figured I'd apply in mathematics at Stanford so I could also be close to the psychologists. Even if I got a degree in math, I wouldn't be out of touch with why I'm going into science, which was to understand the mind. And that's what I did. I went to Stanford.

Throughout all this, I can't overemphasize my sense of loneliness. I had a few wonderful professors, notably John Kemeny and Albert Hastorf, who were really very supportive, but there was always great anxiety because no one seemed to really understand what I was doing. I think they had the sense that because I was so "brilliant," unquote, I couldn't be a nut. I was doing what I as a young person was supposed to be doing: breaking new ground; and they tried to help me get to the people who could really evaluate what I was doing.

While this was going on, I wrote my senior fellowship undergraduate thesis at Dartmouth in 1960–61. It introduced the Additive Model and used it to analyse a lot of data about verbal learning. Because of this background, I don't believe that this model should be named after Hopfield. He simply didn't invent it. I did it when it was really a radical thing to do. My goal, to jump years later, was not to have any of these models named after anybody. I felt that models should have functional names—like Additive Model. Various power cliques do not seem to see things that way. They seek to aggrandize themselves even if, in so doing, they do violence to history.

When I went to Stanford I was sustained by my passion and love for science. My results enabled me to feel a little closer to the enduring beauty of the world, and gave my life a growing sense of focus and purpose. This was balanced against widespread indifference or skepticism about what I was doing. Without strong enough computational or mathematical tools, I realized that I had a limited amount of time to continue in this mode, because I was living off people's largess. I paid my dues by taking ninety credits of graduate mathematics, but there was no particular reason for established faculty to let me continue surviving as a scientist. Everyone else was planning to get a job in a well-established field, but there was no field that represented what I wanted to do.

Pat Suppes had been particularly active in getting me to come to Stanford. In fact, I was accepted in psychology and sociology in addition to mathematics. He was, however, incredibly busy. After I got there, I would hand him paper after paper that I was doing while I was taking my math courses —on human verbal learning, on animal discrimination learning, on competition, and so on. He never read any of them. When I would get up the courage to visit his office intermittently, he would ask what I was doing, and I'd give him a manuscript to read. I'd say, "I'd really appreciate if you'd look at it or maybe tell me what's wrong with it." I'd go back six months later, but he didn't have time to look at anything.

I greatly admired Bill Estes. I was unfortunately very shy, and Bill Estes was not exactly talkative. Whenever I visited him, I was always amazed by the fact that he was so quiet. He would sit there without changing his facial

expression or saying a word and wait for you to talk. I very much wanted to communicate with him because I could see why Stimulus Sampling Theory worked when it did. I had stimulus sampling operations in my neural model. I could see how, in the neural model, if you changed variables, you'd get ratios of long-term memory traces that were just like stimulus-sampling probabilities. I could see why Estes' model worked and why it would fail. But I found it almost impossible to talk with him. I've never resented him for it because he's a marvelous man, and that's just the way he is. But it would have made my life much easier if he would have been able to draw me out a little more.

I realized later that Estes and his Stanford colleagues had a real struggle of their own to get Stimulus Sampling Theory accepted by experimental psychologists and to make it work. Then here comes this kid with neural networks. Well, what are they? Nonlinear differential equations, emergent properties. They didn't understand it well enough to want any of it. And I was too young to have the social skills with which to try to change their paradigm. It was also too soon—this was in 1961 to 1964. The failures of Stimulus Sampling Theory were not yet obvious enough for that paradigm to be abandoned. My experiences at Stanford were, by and large, a great disappointment because I only went there to try to get in touch with these people.

Since I couldn't sell any of my work, all I could do was to work even harder to try to understand more and to get closer to the communion with Nature that I desired. The good things that Stanford offered me were that I took ninety credits of graduate mathematics and read lots of physics, psychology, and neurophysiology, so I kept growing intellectually. I worked hard as a graduate student although I was a very unhappy one as I took course after course.

For the first year and a half or so, at Stanford, I didn't do any of the research that I did as an undergraduate. I didn't work on neural networks because I was trying to cope with the very real challenges of being a mathematics graduate student. I did love studying mathematics. I found the mathematics to be really beautiful. And I was able then to read a lot of physics quickly because I learned all the relevant mathematics. Socially, though, the first year at Stanford was so disappointing that I thought the second year had to be better, because I'm an incurable optimist. I figured, "This is so bad that next year has to be better." Well, the next year was equally bad. So then I tried to get out. An unfortunate accident then occurred. I was on an NSF [National Science Foundation] fellowship, and I realized that I hadn't yet heard about my fellowship renewal. When I went to the office in the mathematics department, they said that I probably should have heard something by then. They inquired for me because renewal was supposed to be automatic. As it turned out, the NSF had mailed my renewal notice to the wrong address. I just had to fill it out and send it back, and it would have been renewed, but because of the delay it was just past the renewal deadline. So suddenly not only was I unhappy, but I also didn't have any money.

Then Stanford did a very nice thing. They gave me a fellowship that supported me for another year. After that, I knew I couldn't stand it there any more. I got my Master's degree, and thought I'd try to go to MIT, in part to study under people like Norbert Weiner, and also because my girlfriend was then a graduate student at Harvard.

By this time, I had been reading a lot of papers by people at the Rockefeller Institute—papers by the neurophysiologists there and also papers by people like Mark Kac on statistical mechanics. So I wrote a letter to Rockefeller asking for information about its program too. This was a period when Rockefeller had a lot of money. They responded to my letter by checking up on me. I don't really know how they did it to this day, but the next thing I knew, they invited me to visit there.

I figured that I'd visit Rockefeller and then I'd also visit MIT and see if I could get an interview there too. My visit to Rockefeller seemed unreal. It had a gorgeous campus right in the heart of Manhattan. There was a guard at the front gate whose name was Angel. It was really like going to Heaven. You could go from Heaven to Manhattan and back every day if you were a student at Rockefeller!

So I transferred to Rockefeller instead of MIT, and that was a wonderful experience for me in many ways. On the other hand, I still had the usual problems there. My primary mentors were Mark Kac and Gian-Carlo Rota, who had just come from MIT. Rota had a sense of what I was doing because, in addition to being a mathematician of great breadth, he also was a professor of philosophy. He kindly became my "protector." At Rockefeller you really needed a protector! Rockefeller was then set up as a set of laboratories, and there were no required courses at the time. There were a number of lecture series. Still, various students went to Columbia or to NYU to take other courses.

Because Rockefeller was so unstructured, if you didn't affiliate yourself almost immediately with a laboratory and get a lab chief to claim you, you were vulnerable to the fluctuating winds of political change. My protector was Gian-Carlo Rota. During my years there (1964–1967) I continued to make lots of discoveries. Then I also had to write a Ph.D. thesis. What I did for a thesis was to develop methods to prove global limit and oscillation theorems for the Additive Model, treated as a content addressable memory [CAM]. They were, I think, the first global CAM theories.

Even then there were problems because—I don't want to go through the sordid details—there were some professors who did not believe the theorems. I had struggled very hard to find a way to demonstrate that my models worked as I claimed they did. And what could be more secure than a theorem? The shock was that they didn't believe the theorems! They thought that there must be a mistake. These people called me crazy before I proved them. Then they said that the theorems were crazy!

Fortunately, by that time Los Alamos had a big enough computer to run the equations, and this was done by Stan Ulam's group. They were

interested in the theorems because they described a nonlinear collective phenomenon. At first they didn't believe the theorems either, but then they ran the networks on the computer, and the simulations did exactly what the theorems said they should. These CAM theorems analysed associative pattern learning in several critical cases: fully connected autoassociators, feedforward networks, and partially connected feedback nets. Given the difficulties I had in getting good scientists to believe my CAM theorems for Additive Model autoassociators in 1966, you can see why I am so annoyed that various people credit Hopfield for this model based on his work in 1984. I'll say more about this later.

Each graduate student at Rockefeller wrote up a first-year project. My first-year project in 1964–65 turned out to be a monograph of around five hundred pages, which synthesized my main results of the past ten years. It was called "The Theory of Embedding Fields With Applications to Psychology and Neurophysiology." It took me a long time to write it, and then the question was what to do with it. Several professors realized that students don't write five-hundred-page monographs every day. They wanted to get someone to evaluate it, so they arranged with me to mail it with a cover letter to 125 of the main psychology and neuroscience labs throughout the world. It went to David Hubel. It went to Steve Kuffler. It went to Eric Kandel. It went, actually, to most of the major neuroscientists and cognitive scientists in the world at that time. Unfortunately, no one seemed ready to understand it. But that monograph had the main results of my work of the past ten years and the seeds of my work for the next ten years. My published papers in the '60s and early '70s either published or worked out results that were in the monograph. It had a lot of results in it about reinforcement learning and human verbal learning. It also, among many other things, introduced a cerebellar learning model, which predicted that you'd have learning at the parallel fiber–Purkinje cell synapse. That was in 1964. David Marr made a similar prediction in 1969; Jim Albus in 1971. I published this model formally in 1969. Despite this background, the model is today often called the Marr-Albus model.

This has, all too often, been the story of my life. It's tragic really, and it's almost broken my heart several times. The problem is that, although I would often have an idea first, I usually had it too far ahead of its time. Or I would develop it too mathematically for most readers. Most of all, I've had too many ideas for me to be identified with all of them.

Please don't misunderstand my concerns about the so-called Marr-Albus or Hopfield models. My goal wasn't to get priority. Please understand that, first, shy people don't name things after themselves, and, second, I'm nothing. God is everything. I can't name after me something that is God's creation or God's proof. That's why I would try to give things functional names. But then many things that I discovered started getting named after other people! And I was not the only victim. Paul Werbos, David Parker, and, Shun-Ichi Amari should have gotten credit for the backpropagation model,

instead of Rumelhart, Hinton, and Williams. Christoph von der Malsburg and I developed competitive learning and self-organizing feature maps between 1972 and 1976. In fact, Teuvo Kohonen's first version in 1982 wasn't the version that he used in 1984 and thereafter. At the meeting in Kyoto where he presented the first version, I was the chairman of the session. After his talk, I went through his model's properties as part of the general discussion, and I noted that my 1976–78 version of the model had certain advantages. That is the version that was used two years later in his 1984 book. And now the model is often named after Kohonen. Well, if it's named after anyone, the name should include Christoph and me. To leave out Christoph, who had a key 1973 *Kybernetik* paper, which adapted aspects of my 1972 *Kybernetik* paper, or me for my 1976 *Biological Cybernetics* papers which put the theory in its modern form, that's just historically wrong.

If you're doing a reputable history, you have to get right who really invented things. For example, for Amari and Werbos and Parker not to be given primary credit for backpropagation is wrong. How did this happen? In the early 1980s, a type of social autocatalytic wave broke that led to renewed acceptance of neural networks. This wave had been building since the 1970s; I could feel it building then. Some people who were in the mainstream of various related disciplines rode this wave, stoked the wave, and marketed the wave, and they deserve credit for that. Rumelhart has done a great service to cognitive science by promoting neural models, but he, Hinton, and Williams didn't invent backpropagation; he and Zipser didn't invent competitive learning; and all you have to do is to read the published literature in order to see that what I say is true.

Parts of my 1964 monograph were broken up and developed into ten research papers. While I was a graduate student at Rockefeller, I submitted all ten papers to *The Journal of Theoretical Biology*, including my verbal learning model and my derivation of the Additive Model. When the journal got these ten papers from this unknown scientist, they didn't know what to do with them. Bob Rosen, with whom I became friendly years later, was one of the receivers. He said, "If you had sent us one article we would probably have accepted it, but we didn't know how to handle ten." So they rejected them all. That was in 1964–65. This was, of course, a major disappointment for me.

Despite these problems, I got a job at MIT because my advisors at Rockefeller wrote strong letters on the strength of my Ph.D. thesis. When I visited MIT, I was interviewed by both the electrical engineering department and the applied mathematics department. Both departments offered me an assistant professorship because my thesis was considered to be very original. I had introduced a new class of models, these nonlinear short-term and long-term memory models, and I had proved a kind of theorem that was unfamiliar, these global CAM theorems. Then I did something which I think in retrospect was a mistake: I accepted the job in applied mathematics rather than in EE. I didn't realize that MIT was, at that time, really a big engineering department and that the power and influence of that department was

overwhelming. I was influenced by the fact that Rota was returning to the MIT mathematics department when I arrived there.

Several of the math faculty were very kind to me. Norman Levinson was Norbert Wiener's greatest student, and he was one of the great mathematical analysts of our time. He and his wife Fagi took me under their wing as a kind of scientific godchild. They had two daughters about my age, but no sons. Fagi is, in fact, the Godgrandmother of our daughter. I wrote a large number of papers after I came to MIT in 1967, and Levinson, being a member of the National Academy, submitted a series of my notes in *PNAS* [*Proceedings of the National Academy of Sciences*]. There were three notes about neurobiological and mathematical properties of the Additive Model and the more general Shunting (membrane equation) Model. Then I got a series of mathematical papers in the *Bulletin of the American Mathematical Society* and the *Journal of Differential Equations*.

MIT was a good experience in many ways. First, there was the challenge of teaching math to kids. I'd never taught before. My first assignments were to teach math courses in things I'd never even studied! I met the challenge by being so overprepared that I had the whole course totally polished before the first lecture began. Having never lectured, I would go into classrooms to practice my lectures to empty rooms. No one at MIT gave us any advice or help with our teaching. Now such skills are taught in our department to students in a one-on-one faculty-student setting as part of their Ph.D. training.

I remember the first day that I went into a classroom at MIT; the kids saw me, and they audibly groaned because I looked very young at the time. They figured they were getting yet another graduate student teacher. Anyway, I was so overprepared that it went OK, and my teaching was effective. My research also went very well. I published forty-odd papers on an ever expanding set of topics during my first few years there. Because of the range of this work, I'll have to skip it in this summary.

As a result, I was promoted after my first year at MIT from assistant professor to associate professor. I also won a Sloan faculty fellowship. Everything was finally going really well. I was verbally promised a professorship, but then when the time came, there was a major recession. I don't know if you remember in the early to mid-seventies there was a deep recession. A lot of schools got scared. It was the first one for a very long time in the postwar era. Essentially everyone who got a job at MIT after World War II got tenure, but then things crashed, and they started dumping us all. Traditionally at MIT, after you were an assistant professor, there was a critical decision point when you'd either be asked to leave or you were honored by being made an associate professor without tenure. The idea was that everyone who was chosen associate professor would eventually be a professor with tenure. You didn't have to worry about tenure, because you'd already gotten a verbal assurance of your future at MIT.

When the recession hit, I was not helped by the fact that some people considered me a "controversial" case. And no one advised me as to how you

go about getting tenure. I was simply asked to give the department a representative list of people to write to for recommendation letters. I naively gave them a list of about fifty names of distinguished people across the fields of psychology, neuroscience, and mathematics. I got a very wide range of letters. A number of letters said I deserved a Nobel prize and I am a genius. I also had other letters that said, in effect: "Who the hell does he think he is trying to model the mind?"

At this point the department tried to break this deadlock by asking somebody whom everyone in mathematics would respect and who really knew what was going on. I don't know if you know who James Lighthill is? He was the Lucasian professor at that time in Cambridge University. That was Isaac Newton's chair. Lighthill had just written a scathing attack on AI. So they figured, first, he's a very substantial mathematician; he can understand all the math; and second, he has very strong attitudes about AI. Maybe some of them also thought that he'd therefore nail me and get it over with.

Anyway, he wrote a glowing three- or four-page letter which basically said that I was doing exactly what AI should have done. I've seen all these letters. I wasn't supposed to, but there were some people who were so upset that MIT didn't keep me that they wanted me to realize that the letters were, by and large, quite wonderful. They presented the type of case that an experienced reader expects to find when someone is doing something highly original, interdisciplinary, and technical.

I stayed an extra year at MIT. That's the year that Gail Carpenter came to the applied mathematics department at MIT. She's the best thing MIT ever did for me. It was because we overlapped that we could get together. Now we are very happily married and best friends. We have also done a lot of science together.

Then I went to BU [Boston University], and faced my next problem. I am telling you about these problems in order to reassure young people that you can hope to survive a lot of problems if you are true to your craft. A professorship of mathematics was created for me at BU through the President's office. I was told that my letters were the strongest that they had ever seen, but since President Silber was away in Europe, they couldn't offer me tenure until he read my case in the fall. A new dean then came in that fall, and said that he opposed creating professorships through the president's office. He said he would oppose me coming up for tenure in the fall, as promised by the previous dean, but would support me if I waited two years before going through the entire tenure process. Unfortunately, this man was no longer dean when I came up for tenure two years later. Who was? The man who had supported my tenure in his role as a Vice President two years earlier. No problem, right? Wrong.

This new Dean fired me. Why did he fire me? Because he claimed that the mathematics department was too big already. (It has since grown to more than twice its size then.) So I was fired from BU a few years after I got there, and I again lost the tenure that was verbally promised to me.

To make a long story short, the next six months were a very unhappy time. I appealed this decision to President Silber. It took quite a while before Silber considered the case. When he did, he also called up a lot of people to ask about me. Finally, he called me in to his office and said, "You're exactly what we want. I'm really sorry for the inconvenience." So I finally got tenure in 1975 after having been twice rejected for tenure. After 1975, for the first time I had some stability. It took me a few years to adjust to that. Now I hold an endowed Chair at BU and am one of its most respected faculty. My advice is: Never give up and don't hold grudges.

ER: You wanted to talk about Paul Werbos ...

SG: When I was at MIT, Dan Levine was one of my graduate students, and Dan was a friend of Paul Werbos who was then at Harvard. Dan told me about his very bright friend who was trying to do some work in neural networks, and he was having a lot of trouble with his Ph.D. thesis committee. So Paul came over and talked. The main thing I remember was that he seemed very bright and enthusiastic, but also talked a lot about all his troubles in getting people to understand and support what he was trying to do. This was a recurrent theme—that people who were making important discoveries about neural networks were hitting political brick walls.

The advice I gave him, which was the only thing I could do, was that he work out examples for people so they could see how his model worked. This he did, and he eventually got his Ph.D. thesis approved. You see, it was a period when I wasn't the only person experiencing brick walls right and left. I met a lot of very smart people who just vanished from the field. They just couldn't find a way to survive. Paul found a way. He deserves immense credit for that. The fact that, more than a decade later, people like Rumelhart, Hinton, and Williams were able to run with his ideas and further apply them when the scientific market was ready to receive them shouldn't deny the originators the credit that they deserve for introducing the ideas. I believe this both because it's the right thing to do, and also because that's why the field developed so fast in the 1980s. The foundations were already there; a lot of the main models were known. One can't believe that in 1982 suddenly everything was discovered. This just isn't the history of our or any other scientific field.

Around 1980, the Sloan Foundation started to give out grants in cognitive science. In college, I had gotten a Sloan predoctoral fellowship and at MIT I won a Sloan postdoctoral fellowship, so I figured I might get lucky again. I therefore called them up, and I asked, "If I submit a grant in cognitive science, would you consider it?" They said, "Well, you can't because you're not a center. We only make grants to centers." It was at that point that the concept of forming a center, a new administrative unit that could support people from many disciplines, firmly took hold in my mind. If I could only become a center, then I could work with people from many different disciplines without having to change departments.

I was at that time already working a lot with Gail Carpenter. I was also working with Michael Cohen, and more and more with Michael Kuperstein. Based on these and other projects, I was able to get a center grant, which allowed the Center for Adaptive Systems to get started in 1981.

The Center enabled me to start building up an interdisciplinary community of people interested in real-time modeling of mind and brain. I also wanted to help smart young scientists to have an easier time than I did. I almost didn't make it at multiple points, and I felt a commitment to making it easier for others to do so. After the Center succeeded, in 1988 I was able to get the university to start a graduate Ph.D. and M.A. granting program in Cognitive and Neural Systems. This program became a department two years later. It has developed an interdisciplinary curriculum so that graduate students can learn the field in a more systematic way. I also introduced the *Neural Networks* journal, and while introducing the journal, founded the International Neural Network Society [INNS], which began to bring together people from a lot of different disciplines.

One of the unfortunate facts about our field was that it was broken up into cliques that didn't cooperate. Physicists don't all love each other; in fact, I think they're probably one of the most competitive groups of scientists in the world. But they've learned how to cooperate to get more resources for all physicists. I hoped that INNS would help to fix this problem. So far, it has only achieved partial success because clique activities still tend to divide the field.

ER: Maybe you could say a little bit more about the more recent scientific work you've been doing.

SG: There've been a lot of streams of work in my life. The most pervasive stream has to do with parallel information processing and learning—the interactions between short-term and long-term memory. The earliest work was on human verbal learning—the problem of serial order in behavior and how you can get distributed patterns of errors that would evolve in a given context, like the bowed serial position curve, and why paired associate learning and serial learning were different.

I also did a long series of papers about global CAM and associative pattern learning. The problem was, how do you know it works as you would like? I spent years on proving that about what I call the Generalized Additive Model, which includes the so-called "Hopfield model," that I hope will not be called that for much longer.

I also did a lot of work about animal learning. If you think about conditioning—operant (or instrumental) and classical (or Pavlovian)—it forces you to also think about decision making, and associative learning between cognitive and emotional representations. Putting these concepts together leads you to think about the feedback between cognitive and emotional representations and how it focuses attention upon salient events. So, in thinking about this sort of decision making, I realized that I needed short-

term memory nets that were self-normalizing. I had this insight first in my 1964 monograph and developed it for conditioning around 1969. My first paper on operant conditioning, per se, was in 1971. It has supported a lot of subsequent work.

My early learning theorems included outstar theorems, in which single cells can sample distributed patterns. I then realized that, once you have stimulus sampling, you need to ensure the selectivity of sampling in response to the proper combination of environmental cues. I then introduced instar theorems to ensure selective sampling to trigger outstar learning.

My 1970 paper on neural pattern discrimination used Additive Models with thresholding of signals to show how you could construct selective instar pattern discriminators. I realized around this time that you have to match what you can learn and what you can discriminate through information processing. This insight led to instars in 1970. These discriminators needed two layers of inhibition. In a 1972 article, I pointed out that these layers were reminiscent of retinas, where the first layer was like the horizontal cell layer and the second like an amacrine cell layer. This article also showed how an instar could adaptively change its selectivity to input patterns. This 1972 paper influenced Christoph von der Malsburg, who used the Additive Model but also introduced the key idea of tuning the adaptive filter with the weights in the filter, whereas I was using adaptive thresholds. His article came out in 1973.

In the interim, because of my interest in how short-term memory works during discrimination learning, I had mathematically attacked the problem of how you design short-term memory networks. It was a thrill to prove mathematically that properly designed networks had self-normalization and limited capacity properties as emergent properties of the net. Then I started classifying signal feedback functions, and I proved that sigmoid signals had very good properties; they suppressed noise, and also had had partial contrast enhancing properties.

I also proved how to design a winner-take-all net. That caused a little debate between Jim Anderson and me because I liked using the membrane equation, or shunting net, wherein I could suppress noise and still get self-normalizing contrast enhancement. His Brain State in a Box got contrast enhancement at the price of also amplifying noise.

I summarized all these results in a 1973 article, wherein properties of shunting competitive-feedback nets for short-term memory were classified in terms of how different signal functions altered the pattern stored in memory. I think that this was the first paper that mathematically proved why a sigmoid function is important. When I read Christoph's paper in 1973, which I thought was a remarkable paper, I was very gratified that he had used the Additive Model, but he also modified it. In my 1972 article, I had used a learning law that included both Hebbian and anti-Hebbian properties. I introduced that law in 1967 and 1968 in PNAS. It was also used in ART [Adaptive Resonance Theory] later on.

When I saw Christoph's 1973 article, I realized that it had several problems. One problem was that he had used a purely Hebbian learning law. Left to its own devices, this law would only allow adaptive weights, or long-term memory traces, to grow. To prevent this, he alternated learning intervals with trace normalization intervals. The model thus did not run in real time and it used nonlocal interactions. Based on some modeling and mathematical work that I'd done in the past few years, I saw how to design a real-time local model.

One step was to control the contrast enhancement and normalization of activity in the category node level of the network. The theorems from my 1973 article on recurrent on-center off-surround networks helped me here. In that paper, I described the first winner-take-all competitive network. More generally, I proved how a sigmoid feedback signal function could achieve self-normalizing, partial contrast enhancement, which Kohonen now calls "bubbles." I also realized that the input vector needed to be normalized, and discussed how to do this with an L1 norm in my 1976 article. Later, in my 1978 article on human memory, I generalized this to an arbitrary Lp norm, and singled out the L2 norm case for its unbiased properties. Kohonen used the unbiased L2 norm in his articles from 1982 onward.

With these innovations in place, I could then return to the use of the mixed Hebbian/anti-Hebbian learning law of my 1972 and earlier articles (it was introduced, actually, in 1958 when I started my work at Dartmouth). This learning law kept the adaptive weights bounded without violating real-time and locality constraints. I saw that this model was far more general than the application to which Malsburg had put it, which was the development of hypercolumns in striate visual cortex. For me, it became a general engine for classifying the widest possible range of input patterns. That is why I titled my 1976 *Biological Cybernetics* articles "Adaptive Pattern Classification and Universal Recoding."

The "Universal Recoding" part came from my observation that you could map the outputs from the classifier part of the network into the inputs of an outstar pattern learning network to learn an arbitrary map from m-dimensional to n-dimension space. This fact is of historical interest for two reasons. First, Hecht-Nielsen presented basically the same model again in the mid-1980s and called it counterpropagation. It has since achieved some popularity under that name. Second, when people popularized backpropagation in the mid-1980s, they often claimed that, whereas backpropagation could learn such a map, previous models could not. That, like so many other claims during that period, just wasn't so. In fact, my "universal recoding" map could learn such a map in an unsupervised way using purely local interactions, whereas backpropagation always required a teacher and used a nonlocal transport of adaptive weights.

Trying to live with so many false claims has been difficult for me, at times. If I try to get credit where it is due, then people who want the credit for themselves often mount a disinformation campaign in which they claim that

all that I think about is priority. Because I have been a very productive pioneer, who innovated quite a few ideas and models, that can create quite a chorus of disinformation! If I don't try to get credit for my discoveries, then I am left with the feeling that eventually most of my ideas may become attributed to other people, especially if I have them too far ahead of my time.

Anyway, that's not why a person who has been scientifically active for as long as I have—now 40 years—keeps working. So, after designing the first self-organizing feature map of the type that is now used, I proved a theorem in a 1976 *Biological Cybernetics* article which says that such learning is stable in a sparse input environment; that is, an environment in which there aren't too many inputs or input clusters relative to the number of coding nodes. In fact, this learning has Bayesian properties, and I showed that the model's adaptive weights are self-normalizing and track the density of inputs coded by each recognition category. These properties were later exploited in the 1980s and thereafter in many applications by people like Kohonen.

My own interest was, however, primarily in how to classify arbitrary input environments, because no one controls the sparseness of inputs in the real world. I therefore also described examples in the 1976 article in which you could cause new learning to catastrophically erase old memories if the inputs were dense and distributed through time in a nonstationary way. This raised the urgent question of how the system could learn stably in a general input environment. I thought of this as a *stability-plasticity dilemma*, or how could a system continue to learn quickly in an arbitrary input environment without also forgetting what it earlier learned? Said in another way: Why doesn't fast learning force fast forgetting?

At this time, something exciting happened. I had published an article in 1975 in the *International Review of Neurobiology* on a neural model of attention, reinforcement, and discrimination learning. This model culminated almost two decades of work on classical and instrumental conditioning, which —as noted above—is the name given to those animal and human learning situations wherein rewards and punishments operate. In this article, I developed a model of cognitive-emotional interactions to explain how attention gets drawn to motivationally salient events. These phenomena included what is called attentional blocking and unblocking; or how do we learn what events predict rewards or punishments and focus attention upon them, while learning to ignore irrelevant events? This paper included my first adaptive resonances, which were feedback interactions that matched cognitive with emotional representations to focus attention in the desired way. I also needed to analyse what happened when a mismatch occurred, and this led me to introduce an orienting system that would search for and unblock previously unattended, but correct, cognitive representations, that could reliably predict the types of rewards or punishments that might be expected to occur.

One of the most exciting moments in my life occurred when I realized that the same dynamics of match/mismatch, search and learning that were needed to focus attention during adult cognitive-emotional interactions were also

needed to stabilize the development and learning of purely cognitive representations, from childhood on, including the learning of visual object recognition, speech, and other cognitive codes. This brought the reinforcement and cognitive literatures together in a truly radical way. Before that time, there had been bitter controversies between reinforcement and cognitive approaches to psychology. People like Skinner on the reinforcement side and Chomsky on the cognitive side were at each other's throats. In like manner, cognitive models in artificial intelligence were attacked for not being able to incorporate intentionality or feelings.

Part two of my 1976 *Biological Cybernetics* article introduced Adaptive Resonance Theory, or ART, to unify all of these apparent antagonisms. The key was to understand the central role of the stability-plasticity problem, or how to learn in real time throughout life without experiencing catastrophic forgetting. I showed that self-stabilizing learning required, among other things, the learning of top-down expectations, which focused attention on expected aspects of events. In other words, the stability of learning implies the intentionality of cognition and the fact that we pay attention. The universal status of the stability-plasticity problem helped to clarify why it could bridge between the cognitive and emotional domains. I also suggested that *all conscious states are resonant states,* and still have read no experiments that have led me to abandon this view.

I then did a lot of work on cognitive information processing, and I began realizing that I could only go so far until I knew what the functional units were that were being processed. These cognitive resonances were able to provide an intermodal binding of information from different sensory streams. But each of the sensory streams had its own heuristics. If you didn't know what the sensory units were, then you could go only so far. So I started working more and more on vision and language. I guess the most fundamental paper of that period was my 1978 human memory paper that was published in *Progress in Theoretical Biology,* because in that paper I offered a unified analysis which generated a lot of insights about perception, cognition, and motor control. That paper became a launching pad for the next ten years of work—just as my 1964 Rockefeller monograph had been a launching pad for the previous ten years of work.

Another stream of work tried to understand how to design neural net content-addressable memories. These networks always converge to one of a possibly very large number of equilibria in response to a fixed input pattern. That work greatly generalized my 1973 analysis of winner-take-all nets, sigmoids, and the like. Through it, I gradually identified in the mid-1970s a class of models which generalized recurrent on-center off-surround networks with additive or shunting dynamics, and which always approached equilibrium points. To prove convergence in all of these models, I introduced a Lyapunov functional method that made precise the idea that you can understand a competitive process by keeping track of who is winning the competition at any time. These results led me to conjecture that networks with

symmetric coefficients always converge, as a special case of one of my general theorems. Mike Cohen and I then tried to prove this. We ultimately failed, but in the process, and because we were thinking about Lyapunov methods, we came up with a Lyapunov function in 1981 that helped us to prove the conjecture directly. We published this work in 1982 and 1983.

This work is now known as the Cohen-Grossberg model and theorem by a lot of people. We didn't name it that ourselves. The name came about because John Hopfield published a special case of our result—the case of the so-called Additive Model—in 1984, and it was called the Hopfield model by his colleagues. I had actually introduced that model almost 30 years before and Mike and I had published the Lyapunov function for it before, so quite a few people were not happy about naming it after Hopfield. They called our, more general, work the Cohen-Grossberg model to protect it from being misnamed later on. This sort of thing unfortunately happened all the time.

I also was very interested in understanding how a more complex form of content addressable memory was designed; namely, a working memory. This is the type of short-term memory whereby, say, you can remember a new telephone number for a short time after you first hear it, but can then forget it entirely if you are distracted before dialing it. A lot of data now points to the frontal cortex as a site of working memory. I realized that this problem of short-term memory was intimately linked to a problem of long-term memory, which is the type of more enduring memory whereby, say, you can remember you own name. (Presumably you don't forget your own name every time that you're distracted!) The main issue was that, as a novel list of items—like the numbers in a new telephone number, or letters in a new word—is presented to you, you don't want its storage in working memory to force you to forget familiar learned groupings of those items. For example, if you've never heard the word MYSELF, but have learned the words MY, SELF, and ELF that are subwords of MYSELF, you don't want the storage of MYSELF in working memory to erase the long-term memory that you have of its familiar subwords. If this were true, then we could never learn a language. This study thus identified a variant of the stability-plasticity dilemma that applies to temporally ordered memories.

I was happy to identify two postulates for such a working memory that would realize this goal. These postulates guaranteed that the familiar learned groupings would not be forgotten if they were coded by the type of bottom-up adaptive filter that occurs in a self-organizing map or an ART system. That enabled me to write down rules for generating all of the working memories of this type. Remarkably, these postulates could be realized by a specialized version of the recurrent on-center off-surround nets that I'd already studied! I was then able to prove something really surprising. Previously, it had often been thought that you could just have a recency gradient in working memory, in which more recent events were performed before earlier events. But just at around the time that I was working—in the mid 1970s—data began to appear suggesting that you could have an

inverted U of short-term memory activity across items, with the earliest and most recent items performed before items in the middle of a list. In my model, larger activity of an item's representation translated into earlier performance. It turned out that I could characterize the conditions under which you'd get recency, primacy, or bowed (inverted U) gradients in working memory activity. I used this result to explain, for example, data about human free recall of recently presented lists of items. Here, items at the beginning and end of a list are recalled earlier, and with higher probability, than items in the middle. The main paradoxical result was this: The need to be able to store items in short-term working memory without destabilizing previously learned groupings in long-term memory sometimes implies that the *order* of storage is not veridical. It was veridical for short lists—a result which clarified concepts like the immediate memory span—but not for longer ones.

These results, combined with my earlier work from the 1960s on the long-term memory of temporally ordered lists, were both included in my 1978 human memory article, along with a lot of other stuff. They provided the foundation for more recent work about speech perception and motor planning. One of the most interesting things to me about results like this is that they showed how an adaptive property—like the stability of learned groupings—could lead to a maladaptive property—like the wrong order of storage—in certain environments. Many of my results are of this type, including results about mental disorders like schizophrenia, juvenile hyperactivity, and Parkinsonism, or about maladaptive partial reinforcement effects like persistent avoidance behavior, gambling, or self-punitive behavior.

ER: I wanted to discuss some of the other things that have gone on. You have patents, and I know that you've served on the science advisory board for Robert Hecht-Nielsen's company, HNC Software, and I was wondering if you could talk a little bit about some of the applications of your work.

SG: We've gotten patents on several of the ARTs—ART 1, ART 2, ART 3, ARTMAP. We've also gotten a patent on the BCS [Boundary Contour System, a computer vision algorithm] and on the Masking Field multiple-scale short-term memory and coding network. Our goal was not to interfere with research and development. We want to encourage that in every way possible. But if a company uses ART, say, to make a lot of money, then we would like to get some of it back to further energize the research that led to it. BU has been very good in helping us get patents under its university individual investigator agreement. I would say a lot of people are using ART, and more and more people are using our other models, but so far, most of this activity is still in the research and development phase.

ER: Do you have a relationship with any other company besides HNC?

SG: Well, I don't even have a relationship now with HNC; I was its first chief scientist, but am not any longer. To talk about more history, did you know that Robert Hecht-Nielsen entered the field in part because of me?

ER: He said that in his interview.

SG: My understanding is that Robert was reading the *Journal of Differential Equations* in the late 1960s where my early CAM theorems on the Generalized Additive Model were proved, and he got really interested in them. Robert then started to call me up every couple of years or so when he'd be in town on business, and we'd have lunch and talk about neural nets. I'd known Robert during years when one could only dream that neural net theory would be turned into a technology. When HNC started, Robert invited me to be its first chief scientist. It wasn't entirely clear to me what that would mean given that we were three thousand miles away from each other.

At the same time, Frederico Faggin and Carver Mead invited Gail and me to join their company, Synaptics. That created a serious conflict for me because I knew Gary Lynch was involved, and he's a really good neuroscientist. Carver and Frederico are, of course, top chip designers, and they promised to put some of our key algorithms into chips.

But I felt an old, even romantic, debt to Robert, and so I said yes to his offer. Apart from periodic meetings at HNC, not much happened. As the company faced the realities of trying to survive in a market where there weren't yet any niches, they needed very near-term products and business plans. My sense is that they changed direction several times before the company became a big success. I was too far away to be a large contributor to these strategies. So after our initial agreement wore off, it wasn't renewed.

ER: Do you hare other commercial relationships with other companies?

SG: We [the Center for Adaptive Systems] had a relationship with Hughes Co. on a joint DARPA grant. Gail has consulted for Boeing. A lot of our students have been getting good jobs at high tech companies. Several of our students were hired by MIT Lincoln Lab.

I feel that one really has to try to train people in the many interdisciplinary tools that the market needs. The nice thing about backpropagation is that it's easy to learn, so a lot of people use it. But backpropagation has a limited range. It's good for certain stationary problems where the variability in the data isn't too great, where there aren't too many inputs, and where you can run learning slowly and off-line. Fortunately, there are a lot of problems where these constraints hold. Backpropagation was there to help energize interest in the field, but there are at least as many problems where you want to learn in real time, on-line, with fast learning. Our students know backpropagation and ART, among many other skills.

ER: How many students at the Department?

SG: At this point there are fifty Ph.D. students and up to thirty M.A. students. The M.A. students are very interesting; all of them have full-time jobs in the area; for example, one of our M.A. students who recently graduated is an M.D.; he's at the Eye Research Institute. He's a clinical eye researcher, and he came to take courses to learn about neural models. Others are at Raytheon, MITRE, MIT Lincoln Labs, and so on. They're all already working professionals. We've had people come from the National Security Agency to

get a masters, who now tell us that ART is being used to guard the nation's safety. The masters' students form a very interesting pool. They are one reason why we teach many of our courses once a week in the evening, so that qualified working people can set aside an evening each week to take a course each term, or possibly two courses.

ER: This leads me to my final question, which is to ask you to speculate about the future of neural nets. Where do you see the field going?

SG: I don't really feel comfortable talking about the future because I could never have predicted the present. I have hopes rather than predictions. I hope there will be more harmonious interaction among neural network colleagues.

ER: Well, I was going to give you a last opportunity to address non-organizational and nonpolitical issues about the future of neural networks.

SG: I'll build my answer on some thoughts about why the brain is special. The following anecdote may help to make my point. Richard Feynman came into the field because he was interested in vision. When he realized that the retina is inverted, with the photodetectors behind all the other retinal layers, so that light has to go through all those layers before reaching them, he got out of the field. He couldn't figure out what kind of rational heuristics could be consistent with such a strange fact.

So here we see one of the very greatest quantum mechanicians admitting that brain dynamics are not just an easy application of quantum mechanics. On the other hand, the brain is tuned to the quantum level. You can see with just a few photons. The sensitivity of hearing is adjusted just above the level of thermal noise. So the brain is a quantum-sensitive measuring device. Moreover, the brain is a universal measuring device. It takes data from all the senses—vision, sound, pressure, temperature, biochemical sensors—and builds them into unified moments of resonant consciousness. The black body radiation problem, which Planck used to introduce quantum theory, also had a universality property. But then why isn't the brain just another application of garden-variety quantum mechanics? What's different?

My claim is that what's different is the brain's self-organizing capabilities. The critical thing is that we develop and learn on a very fast time scale relative to the evolution of matter. The revolution is in understanding universal quantum-sensitive rapidly self-organizing measurement devices.

Let's look at the history of science from this perspective. In the Newtonian revolution, the universe was described in terms of fixed, absolute coordinates. Then Einstein taught us that the way in which we make physical measurements, including how fast light travels, can influence what we know about the world. Then quantum mechanics went a step further and taught us that the act of measurement can actively change the states that are being measured, as in the Heisenberg uncertainty principle. But still, in all of these theories, the theory of the measurement device itself was really outside of physics. Physics taught us how measurement could be changed by the mea-

suring device, but it did not provide a theory of the measurement device, in this case, the brain.

Theories of mind and brain, in contrast, are really theories of measurement devices which happen to be self-organizing in order to adapt to an evolving world. Understanding such measurement devices would be a very big step in science. So why has it taken so long for such theories to get born?

My answer is in the form of a story that has comforted me greatly when I was trying to figure out why our field is so crazy. I've written about it in several papers and books, so you may already know my view. I believe that we are living through part of a century-long process that has gradually led to the recent flowering of neural networks.

If you look at the greatest physicists of the middle to late nineteenth century, you'll see that they were often great psychologists or neuroscientists too. For example, [Hermann von] Helmholtz started early in life to test whether philosophical Idealists like his father were correct. Was it really true that you can act on an idea at the instant that you have it? Helmholtz tested this by measuring how long it took a nerve signal to travel along the arm. To do this accurately, he had to compensate for factors like muscle activity and heat generation. By making very careful measurements, he discovered the law of Conservation of Energy, which is one of the foundations of nineteenth-century physics. And he did this to settle a philosophical question by using methods of neuroscience. Helmholtz was as interested in the physics of vision and audition as he was in the psychophysics of how we perceive visual and auditory events.

The same was true for [Ernst] Mach, who studied the Mach bands in vision as well as the Mach numbers that are important in aeronautics. Mach's interest in space and time helped to inspire Einstein's general relativity theory.

[James Clerk] Maxwell developed the kinetic theory of gases and his theory of the electromagnetic field, but he also developed an important theory of color vision.

All of these physicists were interested in both external physical space and time, and internal psychological space and time. Remarkably, this was no longer true in the very next generation of physicists. You might say that this was just due to career specialization, but that is not convincing, because it happened too fast. I believe that there were deep intellectual reasons for this schism between physics and psychology. In particular, these interdisciplinary physicists were discovering facts about mind and brain that contemporary physics couldn't explain.

Consider Helmholtz's experiences, for example. White light in Newtonian color theory is light that has approximately equal energy in all the visible wavelengths. Helmholtz's experiments showed him, however, that our percepts tend to desaturate toward white the mean color of the scene. This property is related to our brain's ability to compensate for variable illumination —that is, to "discount the illuminant"—when perceiving a scene. Helmholtz

realized that this was a highly nonlinear, context-sensitive process. Let me call this property of context-sensitivity "nonlocal" in the good sense that it involves long-range interactions (not in the bad sense, as in the back-propagation model, that it cannot be plausibly realized by a local physical propagation of signals).

Helmholtz also thought deeply about what we perceive. He claimed that we perceive what we expect to perceive based upon past learning. This was, greatly simplified, his idea of "unconscious inference." Such a view implies that bottom-up inputs from our experiences are matched against top-down expectations through some sort of cooperative-competitive process, and that these top-down expectations had to be learned. Such learning is a non-stationary process. So one needed nonlinear, nonstationary, and nonlocal models and mathematics in order to understand vision. Helmholtz realized that the necessary concepts and mathematics were not available at that time.

Fortunately, in the early twentieth century, you didn't need a lot of new math in order to do great physics. All of the revolutions in twentieth century physics started using known nineteenth century mathematics. For example, special relativity just used algebra; general relativity used Riemannian geometry; and quantum mechanics used matrix theory and linear operator theory. The physicists' main job was to discover new intuitions with which to understand the world. Once the intuitions were translated into models, the mathematics for understanding these models was ready and waiting. In contrast, to do psychology or neuroscience, you needed to discover new intuitions as well as new types of mathematics with which to analyse these intuitions. As a result, physicists, by and large, stopped studying psychology and neuroscience because their mathematical concepts were not adequate to understand the new data from these fields. Psychologists returned the favor by not wanting to learn much mathematics anymore, because the mathematics used by physics to explain the world was often irrelevant for explaining their data. It was the wrong math. This led, I think, to a major split between physical theorists and experimental psychologists and neuroscientists around the turn of the century. There followed almost a century of great physicists who knew nothing about psychology, but enjoyed nonetheless analogizing the brain to whatever was hot in technology, whether telephones, telegraphs, hydraulic systems, holograms, digital computers, or spin glasses. On the other side, psychologists often had profound intuitions about their data, but they didn't have appropriate formalisms with which to turn these intuitions into precise theoretical science.

This led to a century of controversy during which psychologists (and neuroscientists, too) often became divided into opposing cliques or camps that mapped out the extreme positions of some dimension of the data. For example, some psychologists collected data showing that learning seemed to be gradual, whereas others collected data showing that it seemed to be all-or-none. They were, in a sense, both correct, because the learning rate is context-sensitive, but they didn't have the quantitative tools that were

needed to map out in which circumstances one or the other outcome could be predicted. Likewise, one had the Gestaltists, who believed in the action of unseen electrical brain fields, on the one hand, and the Behaviorists, who believed that everything has to be observable, on the other. Etc.

The net effect was that each group collected more and more data to bolster its position, thereby leading to one of the largest collective data bases in the history of science, but none of the controversies was ever fully settled, because the conceptual and mathematical tools that were needed to describe the underlying nonlinear, nonstationary, and nonlocal processes were nowhere to be found. My own life's work has been passionately devoted to discovering new intuitive and mathematical concepts and methods with which to overcome these apparent antagonisms and to thereby achieve a new synthesis of ideas.

Why is this hard to do? Why has it been such a controversial path? I claim that, in order to self-organize intelligent adaptive processes in real time, the brain needs nonlinear feedback processes that describe dynamical interactions among huge numbers of units acting on multiple spatial and temporal scales. Such processes are not easy to think about or to understand. One of the controversies that I experienced early on was whether we needed differential equations at all. Many people wanted discrete or symbolic models, but self-organizing systems need to be described dynamically in real time. Their symbols emerge from their dynamics. Another controversy involved the use of nonlinearity. Jim Anderson and I clashed about this matter. I felt that Jim wanted to keep things linear as long as he could. So did Kohonen for a number of years. My own derivations used linear interactions wherever possible, if only to point clearly to those interactions which really had to be nonlinear.

Feedback in nonlinear systems is particularly hard to understand. It's essential to achieve the real-time self-stabilization of memory and other properties. But mathematically, it's a large step. Many people have held off as long as they could to avoid closing the feedback loop. Backpropagation illustrates this tendency, since in that model, the feedback loop is never really closed. That is why I think of backpropagation as a neo-classical model that is holding on to the old paradigm as long as it can. In backpropagation, bottom-up activation is used to compute an error, which then slowly adapts the model's weights. What you really need, however, is to close the feedback loop to reorganize the fast dynamics of the activations themselves, which in turn will alter the adaptive weights. Adaptive Resonance Theory boldly took this step, and in so doing helped me to turn Helmholtz's intuitive concept of unconscious inference into rigorous science; in particular, rigorous science that is relevant to why we are conscious.

In summary, I see us ankle-deep in a major revolution that will play itself out during the next few generations. This revolution is about how biological measurement and control systems are designed to adapt quickly and stably in real time to a rapidly fluctuating world. It is about discovering new heuristics and mathematics with which to explain how nonlinear feedback

systems can accomplish this goal. Along the way, we will continue to discover a lush landscape of models for explaining how intelligent and adaptive minds emerge from brain dynamics through their interactions with the world. Even now, such models are linking the detailed architectures of brain systems to the emergent behaviors that they control. Neural models are already being used to solve difficult technological problems, and have suggested explanations of debilitating mental diseases. Ours is a great field that can be of use to many people and also for better understanding ourselves.