Computational intelligence from AI to BI to NI

Paul J. Werbos

Center for Large-Scale Integration and Optimization of Networks (CLION)
Department of Mathematics
U. of Memphis
werbos@ieee.org

ABSTRACT

This paper gives highlights of the history of the neural network field, stressing the fundamental ideas which have been in play. Early neural network research was motivated mainly by the goals of artificial intelligence (AI) and of functional neuroscience (biological intelligence, BI), but the field almost died due to frustrations articulated in the famous book Perceptrons by Minsky and Papert. When I found a way to overcome the difficulties by 1974, the community mindset was very resistant to change; it was not until 1987/1988 that the field was reborn in a spectacular way, leading to the organized communities now in place. Even then, it took many more years to establish crossdisciplinary research in the types of mathematical neural networks needed to really understand the kind of intelligence we see in the brain, and to address the most demanding engineering applications. Only through a new (albeit short-lived) funding initiative, funding crossdisciplinary teams of systems engineers and neuroscientists, were we able to fund the critical empirical demonstrations which put our old basic principle of "deep learning" firmly on the map in computer science. Progress has rightly been inhibited at times by legitimate concerns about the "Terminator threat" and other possible abuses of technology. This year, at SPIE, in the quantum computing track, we outline the next stage ahead of us in breaking out of the box, again and again, and rising to fundamental challenges and opportunities still ahead of us.

Keywords: neural networks, computational intelligence, backpropagation, computational neuroscience, neural codes, history

1 INTRODUCTION

Harold Szu has asked me to review some of the main highlights of the history leading up to the modern field of neural networks (as in neural intelligence, NI), starting from the early work on artificial intelligence (AI) and discussing the ongoing relation to the study of biological intelligence (BI). Harold himself played an important role in this history; thus he knows that I have been involved in many aspects of the history, from the start. In engineering, neural networks are now classified as part of computational intelligence, which also includes fuzzy logic and evolutionary computing and related areas.

The full, three-dimensional history of the neural network field is beyond the scope of any brief conference paper. Late in 2014, I digitized many of the historical documents, which go into far more depth than anything which has appeared in the open literature as yet. There have been culture wars within culture wars, more lurid than the popular TV series House of Cards. At good times, our community has been a lot like a hall of mirrors, with ideas serving as metaideas for other ideas, and reflections bouncing up and down and across disciplines. At bad times, people acting like used car salesmen or worse have led to backwards progress, at least in their fragments of the field. There have been too many inaccurate statements made in print for any reasonable person to fully sort out, even if he or she had nothing else to do.

Because the full story is so complex, I will focus here on just a few major ideas and threads which I was most involved with, episode by episode.

2 NEURAL NETWORKS AND AI THROUGH REBIRTH IN 1987/1988

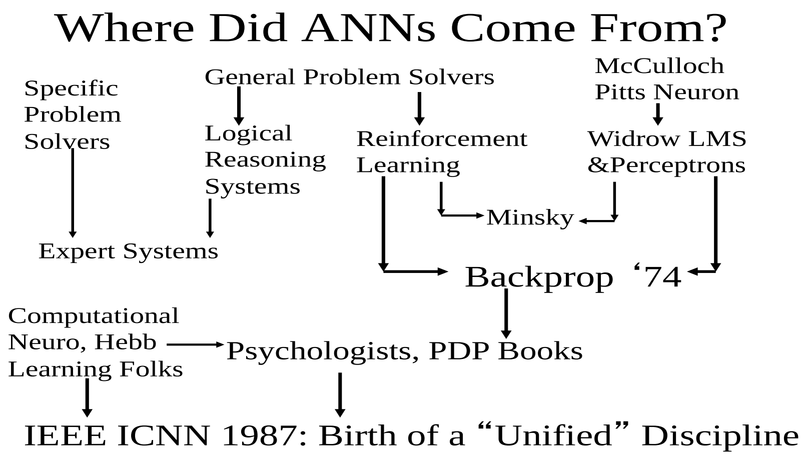

Figure 1. Overview of the early period

Figure 1 gives a condensed overview of what happened in this early period. There were two major streams of thought which had led to initial interest in mathematical neural networks – one stream which emerged from artificial intelligence (AI), the three sources at the top of this diagram, and another stream which emerged from people like Hebb. To simplify the story, I would say that Hebb was the grandfather of the neural network field, on its mother's side, and Von Neumann was the grandfather on its father's side.

Starting from the top – I remember an old cover story in Time Magazine, entitled "Thinking Machines?", back when computers themselves were new. There were three main schools of thought or strategies for how to achieve the goal of building truly intelligent machines in the 1960's.

One was a school which did not try to define or operationalize the word "intelligence," but tried instead to find solutions for specific tasks which seem to require intelligence. They would often say: "I cannot tell you what intelligence is, but I can recognize it when I see it. Only by building things can we begin to understand what it is." Some of the early work in this area, like Samuel's checker player, exhibited some important fundamental principles, which later helped us develop more general and powerful intelligent systems reflecting those principles [1-4]. But a narrow focus on specific applications also led to phenomena like the Big Blue chess system, which beat the human world chess champion, but relied on application-specific tricks and brute force which did not help us much towards the larger goal.

Another was the general problem-solver school led by Newell, Shaw and Simon [5,6]. They argued that intelligence is not a collection of specific algorithms for specific problems, but a generalized ability to solve any problem or learn any task. In many ways, they were the true founders of AI. They inspired a stream of crossdisciplinary research at Carnegie-Mellon University (CMU) which remains important to this day. However, from the viewpoint of a mathematician or engineer, the goal of designing a system which "solves all problems" is not well-defined enough to drive tangible progress. As part of its efforts, this school of thought put special efforts into two major classes of tasks, which are mathematically better defined but which still demand very general capabilities: (1) logical reasoning or theorem proving tasks; (2) "reinforcement learning," newly defined [1] as the task of learning how to maximize some kind of reward or utility signal over time, starting from no knowledge at all about the environment or plant which is sending out that evaluation signal. Marvin Minsky [5] was especially excited about the possibility of achieving truly brain-like intelligence through reinforcement learning.

John Von Neumann was a crucial source both for the second and the third stream of AI. He was the person who first understood and explained the modern concept of cardinal utility function [7], essential to a rigorous approach to reinforcement learning [2,3] and to modern practices of risk management and decision trees [8]. He also had historic discussions with Norbert Wiener and Warren McCulloch, which led to the initial vision of how to build intelligent systems by building networks of model neurons. The McCulloch-Pitts model of the neuron led directly to the early concepts of Multilayer Perceptron (MLP) and Adaline, pioneered by Rosenblatt and by Widrow, the two main leaders of neural networks in AI in those days. James Albus also developed an historic design, the CMAC design, intended to be a model of neural networks in the cerebellum.

In those days, many people simply assumed that any real computer had to be digital, had to be based on 0's and 1's. Thus the McCulloch-Pitts model assumed that every input x_k to a model neuron at any time would have to be 0 or 1. The model neuron, called a Threshold Logic Unit, would first calculate a linear combination of the inputs, which I would denote as:

$$v(t) = w_0 + w_1x_1(t) + \ldots + w_nx_n(t) \qquad (1)$$

and then output a "1" if this linear combination is positive and zero otherwise. Using one layer of linear neurons, Widrow was able to develop many very useful applications in signal processing , using Least Mean Square (LMS) training. LMS technology is still one of the main foundations of real-world signal processing today. However, the early neural network community had great frustration in trying to train networks of neurons with more than one layer, as required to solve even some very simple problems. In the 1960's, Amari wrote a paper which included a sentence suggesting that perhaps networks of neurons could be trained so as to minimize square error based on derivatives – followed by another sentence saying that this would not work. The frustration was expressed very forcibly and clearly in the classic book *Perceptrons*.

After the publication of that book, neural networks assumed the status of sheer heresy. If no one so far, including Minksy himself, could train neutral networks with more than one layer to address a general class of problems, then probably it would be impossible. Neural networks were not considered a frisky futuristic idea; rather they were considered to be an old discredited idea, like perpetual motion. If anyone even started to describe a way to overcome the problems, the immediate response would be: "It is absurd and presumptuous for you to claim that you have the ability to solve a problem which our great leader Minsky couldn't. Of course, we could not even consider such a possibility." And so, the actual technical details would generally not even be considered. It is frightening to me that many in Washington are now promoting changes in the proposal review process which would strengthen the already serious problems of inertia and entitlement versus technical substance.

I myself first arrived on this scene in 1964, as 17-year-old freshman at Harvard, having previously taken a wide variety of pure mathematics courses at Princeton and the University of Pennsylvania while in high school and middle school. I was basically a follower of Von Neumann's approach, but was equally inspired by AI [5] and by Hebb's approach to trying to understand the mind. I was also deeply curious about what to make of the theories of Freud, whom one of my classmates was always talking about; they seemed a bit weird to me, but I understood that I should not lightly discount what people learned after decades of very deep and thoughtful effort to understand the mind.

Hebb's beautiful book about learning and the brain [10] argued that we could replicate all the intelligence of the brain if we could somehow find and implement the correct "general neuron model," a model of how neurons in general learn. He suggested a learning rule in words, based on traditional ideas form psychology that connections which are often used get reinforced. For a special project for a summer computer class at the University of Pennsylvania, I tried to translate that idea into workable FORTRAN and demonstrate what it could do. However, no matter how I tried to adapt it, the Hebb rule basically ended up computing correlation coefficients. I knew enough about statistics at the time that I understood why one cannot learn even general linear relations just by using correlation coefficients; a fundamentally new approach would be needed. I never completed that course, but kept asking how to solve the problem. There were no courses at Harvard that really addressed the problem; thus I took the one and only neuroscience undergraduate course at that time, and courses in economics which addressed the issue of distributed intelligence and optimization. And an independent study with Marvin Minsky.

From thinking about the course in neuroscience, I felt more and more than Hebb's idea of a single general neuron model simply would not work, and would not match empirical reality anyway. If there are, say, three really major divisions of the brain, why not three different types of neuron learning rule, each very general, but each addressing different types of things that brains have to learn?

By the time I graduated in 1967, I was well aware of Minsky's disappointment that he and Selfridge never achieved the type of success they had hoped for with reinforcement learning. A design which "had to work" would not work, except on very small problems, not enough to qualify as a real model of intelligence. While others gave up on the reinforcement learning approach, I published a paper [11] arguing that it was still the right way to go to understand and replicate brain-like intelligence, and arguing that we could solve the scaling problems by designing learning systems more explicitly to approximate dynamic programming. Later we called this Adaptive Dynamic Programming (ADP).

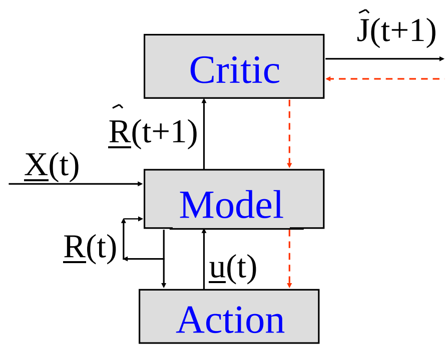

Figure 2. First ADP design proposed to Harvard 1971-71

For my Harvard PhD thesis, I proposed to study and implement the ADP design shown in figure 2. The idea was to achieve Hebb's goals, using three types of neuron, to implement three fundamental functions: (1) to learn to approximate the J function ("value function") of dynamic programming; (2) to learn a stochastic model of the environment, needed for correct training of the other two networks; and (3) to learn what actions to take. (I have recently digitized those thesis proposals, one of which I sent to an offset printer and distributed very widely in those days.) The dashed backwards lines represent derivative signals, to be computed by an algorithm which I then called "dynamic feedback," which is now called backpropagation. (As I understand it, that was the term invented by Rosenblatt for a totally different algorithm, but the community decided to reuse the term.) The pattern recognition problems described by Minsky [9] enter here as one part of how one could train these kinds of networks, such as the Model network, which should be able to predict classifications if need be.

A key part of making this work was to change the model neuron. To try to get acceptance, I proposed the minimal changes needed to make it work. I proposed the use of a "continuous logic unit (CLU)", whose inputs and outputs would all be in the range from 0 to 1. Equation 1 would still apply, but the output of a neutron would be set to equal v, except when v is outside the range (0,1), in which case it would be truncated to 0 or 1.

I remember vividly my discussion with Minsky, when I suggested we coauthor a paper reporting how this solves the problem of training MLPs. Minsky replied that no one would accept my change, because everyone knew that neurons input and output spikes, which are a 0/1 code. I then pulled out the book Sernsory Communication by Rosenbluth (one of the texts in my earlier Harvard neuroscience class), and showed him actual time-series recorded from cortical pyramid cells, the backbone of human higher intelligence. The time-series showed a regular pattern of "volleys," of piles of spikes, one pile after another, at regular time intervals, such that the actual size of the pile varied continuously between some minimum and maximum. The natural way to model such time-series is as continuous variables over sampled time, following the "clocks" which are known to govern the cortex [12,13]. Minsky said he simply could not get away with publishing the idea, empirically valid or not, because he felt that the force of traditional folklore was just too great in the world of computational neuroscience.

In the end, after more and more very weird experiences, the Harvard professor chosen to head my thesis committee explained that they were all very skeptical about whether my dynamic feedback method could actually give the correct derivatives for use in training any kind of network. (Ironically, that also included a person whom some computer scientists assert invented the method!) To address that, I provided a truly rigorous proof (using standards of logic I previously learned from a graduate course of Alonzo Church at Princeton), and the thesis was accepted. I proved and demonstrated that my new algorithm, the chain rule for ordered derivatives, can be used to give correct derivatives for a very wide variety of nonlinear systems, from neural networks to advanced statistical models to econometric models [14]. This 1974 thesis, the first account of the general backpropagation method, has since been reprinted as a book [15].

The Harvard faculty did require that I expunge the sections dealing in detail with neural networks as such. They argued that backpropagation and its applications in statistical forecasting were already enough for a thesis, and that I should save the other material for later. In 1980, I did have a chance to present some of that detail at an IFIP conference [16], and distributed that conference paper to a few select institutions, four of which announced soon after that they had invented a new method (though one of the four retracted the claim after priority become very clear.).

The solution to Minsky's problem was perhaps the most important single event leading up to the rebirth of the neural network field in 1987/1988, but it certainly was not the only important event, as Figure 1 suggests. When at MIT, Stephen Grossberg had shown that Hebbian learning actually can be used to construct a kind of associative memory – something which can play an important role as PART of a system like what is depicted in Figure 2 [3]. Kohonen and Hopfield published important work on generally similar lines. The cognitive scientists played a vital role in communicating and explaining this new way of thinking about intelligence.

The key event in 1987 was a wildly successful IEEE conference organized by Robert Hecht-Nielsen on neural networks, the first ICNN conference. Soon after came the formation of the International Neural Network Society, INNS, founded by Stephen Grossberg, and its first annual meeting in 1988. At that meeting, Barbara Yoon of CDARPA presented a major new funding initiative in neural networks, building on priori work by Jasper Lupo and Craig Fields.

3. EARLY NSF ERA 1988-1992

For me personally, the two most important events in 1988 were a new program in neuroengineering at NSF which I was invited to lead, and a chance to address the 1988 IEEE International Conference on Neural Networks (ICNN), which then implied an audience of thousands in San Diego, under the guidance of Bernie Widrow.

The program at NSF was an outgrowth of the existing program in photonic technology, which had begun to develop a major component in optical computing. Photonic engineers had told NSF more and more that photonics could give us a massive improvement in throughput in general-purpose computing, if we use photonics to implement truly massive parallelism exploiting neural networks kinds of architectures. As NSF checked out this claim, people in electronics said that they too could increase throughput thousands or millions of times, using parallel processing, but they said that it would only be useful for a few niche markets like matrix multiplication; it would not be relevant to general purpose computing. Then, Carver Mead of CalTech, one of the fathers of VLSI, told NSFF that the brain itself uses that kind of massively parallel architecture, and that it is not a niche market. The NSF program director in the area the concluded that a massive improvement in computing throughput is possible, but only if we make good on understanding how massively parallel processing can actually be made to work, in brain-like fashion, in general purpose computing. The neuroengineering program was set up to seek that kind of understanding.

For many years, neutral network chips and photonics were a major part of neural network conferences, but they never could quite keep up with Moore's Law, with the rapid growth in speed of general-purpose chips. Practical applications of ANNs grew in many areas, but implementation on digital chips was good enough until the last ten years or so (beyond the scope of this section). Nevertheless, looking ahead to the time when massive parallelism would be ever more important, the main focus of research in that program was to develop general purpose algorithms of dual use, either making the most of massive parallelism based on learning, and developing the kind of mathematics and algorithms which we will need in order to really understand intelligence in the brain in a functional way.

To replicate brain-like intelligence, I argued, we need to remember what the function of the brain as a whole system really is. The brain as a whole system inputs information form the senses, and ultimately uses them to decide on actions. Neuroscientists sometimes call these actions "squeezing and squirting." Neural networks for pattern recognition and memory are an important part of the brain and an important part of the field, but in the brain itself they are subsystems. To understand a subsystem, one must understand its function, and one must understand the larger function. The brain as a whole system is an intelligent controller, and we need to develop the mathematical understanding of that kind of intelligent control to have any hope of understanding and replicating brain-style intelligence. To follow through on this, I organized two workshops on neural networks for control, one in New Hampshire in 1988 [1], and another at the McDonnell-Douglas facilities in St. Louis in 1990 [17]. I announced a new emphasis area of "neurocontrol," an area of crossdisciplinary cooperation between neural networks and control theory, recruiting several of the top leaders in control theory to address the key issues in this field. Serious applications began to emerge in a wide variety of areas, such as chemical plant control, aerospace control and automotive control.

In 1991, the International Joint Conference on Neural Networks (a conference led alternately by IEEE and the International Neural Network Society) was held in Seattle, and sponsored to some degree by the Boeing corporation which had some of the early applications of neural networks. I will never forget the lunch in the Boeing executive dining room, where someone asked how one might identify emerging technologies that really work in the end. I mentioned the NSF work showing that the level of optimism in science fiction has turned out to be a good predictor, above and beyond more obvious predictors, and one of the Boeing people groaned very, very loudly. He said: "Have you seen the movie Terminator II?" I hadn't, but I promised him I would. That movie did get many, many people to think twice about what directions we were moving in, including me [18]. Thanks in part to help from the Ai community, and a few other gentle shifts in emphasis, we are not nearly in as much danger today of moving to that outcome as we were then, but I agree with Musk and Hawkings that we should not forget the risks.

Should we abandon this line of research altogether, based on these risks? Are humans really ready yet for the incredible new capabilities which still may lie before us? For many technologies, such as certain nuclear and longevity technologies, I have had to conclude, sadly, that we are not yet ready, and should wait. But better understanding of ourselves is something of a prerequisite to *becoming* ready to handle more, or even to cope with the huge risks we already face [13]. Thus I have tried as much as I can to resist the enormous, growing political pressures to deploy quick solutions even in risky areas, and put as much energy as possible into the things which are important as prerequisites to humans understanding themselves and maximizing their own natural capabilities. Unfortunately, the environment in Washington since 2013 or so has been shifting in a direction which makes that ever more difficult to sustain, and raised my concern about many downside risks, especially in neurotechnology as such.

Early in 1990, I still hoped that we could fulfill Hebb's basic vision, by filling in the boxes in Figure 2 with ever more well-designed components, as in [17]. I recognized that these designs required too much computation time to explain fine motor control in the brain; thus I postulated a "two-brain" design, a kind of master-slave system, combining a higher-level reinforcement learning system with good function approximation ability but slow computation, giving value directions to a faster feedforward subordinate system, like the cerebellum of the brain [19]. But I also had many discussions with Albus [20] and Meystel, who argued that the emergent capabilities we see in the brain require more than just the kind of structures I saw then. Even if we try to minimize the amount of apriori assumptions, and rely on emergence and learning to generate intelligence, we do need some additional capabilities [3]. The year 1992 was perhaps the high point in my focus on the limited type of intelligence which I now call "vector intelligence" [3].

Collaboration with McDonnell-Douglas was extremely positive at that time, as you can see from their chapters in the conference book [17]. For example, the breakthrough in low-cost continuous production of thermoplastic parts was crucial in establishing the feasibility of highly efficient aircraft like the Boeing Dreamliner design; however, after that successful breakthrough demonstration in St. Louis, new owners of Boeing outsourced the production of those parts to Japan, and reliability has become a concern. There are many, many equally interesting stories in the references.

Ford Motor company also developed two divisions, with expertise in neurocontrol and prediction with recurrent neural nets, which in my view was the world's top center by far in those fields for many years, and still a major player.

4. THE MIDDLE DECADE THROUGH ABOUT 2006

From 1992 to 2006, it becomes harder to summarize, because of the diversity of events and of communities going their own ways.

1992 was the year of the first IJCNN in China, a major milestone for the field. Already by 1992, the Institute of Automation of the Chinese Academy of Sciences had many interesting and innovative people who attended the IJCNN, but between 1992 and 2006 the Chinese presence has expanded tremendously. Each nation pursuing neural network research has its own strengths and weaknesses, but by some metrics one might argue that China is already ahead of the US. The emphasis on neural networks for engineering is part of the reason, but there may be other factors which help, such as Jiang Zemin's background as an electrical engineer and traditional Chinese culture which understands that intelligence is not limited to the formal manipulation of words.

The entropy of narrow disciplines not wanting to encourage crossdisciplinary understanding and collaboration has been a never-ending problem, requiring never-ending attention, especially when we believe on policy grounds that cross-cutting understanding should be the number one objective here. Within NSF, it was a major step forwards when Dr. Joe Bordogna, acting head of NSF, announced a director's competition for ideas for new cross-cutting initiatives. The cognitive science program was run at the time by Joe Young, who truly believed in the same fundamental cross-cutting directions I did, and we got together to put together a proposal for cross-disciplinary research addresses the core issues in learning which call for new mathematics and yet link to natural intelligence as well. I remember very vividly the meeting where we invited Howard Moraff from Computer Science to join us, and he basically said: "OK, let's agree on LEARNING as the theme." That was a huge step forward, at a time when most AI was still committed to expert systems and prior knowledge as the only allowable forms of "intelligence," and machine learning was a fairly small embattled enclave.

In my view, the LIS initiative had a major impact in steering a change in the culture. It was also a very enjoyable and entertaining experience, but I will spare you the details.

One of the many, many interesting events in this period was a workshop on neural networks for flight control, emphasizing the challenge of learning to control disabled aircraft fast enough to prevent them from crashing ("reconfigurable flight control," RFC), led by Charles Jorgensen of NASA Ames in 1994. White and Sofge of McDonnell-Douglas had reported how a real-time ADP controller could learn to restabilize disable F-15s, assuming heavy damage and using McDonnell's internal simulator, in 2 seconds, well enough to save about half of the damaged aircraft (versus only 2% which would survive with the usual technology). Jorgensen later went on to demonstrate success in autolanding a large, physical MD-11 with all its control surfaces locked up – after he also solved the problem of how to do verification and validation for this kind of learning technology. At this workshop, a young person from Washington University, Dr. Massoud Amin, presented impressive results on the use of time-lagged recurrent neural networks for system identification. Based on the depth of his understanding, we later recommended him to people we knew at the Electric Power Research Institute, who hired him to lead new activities on the electric power grid, which became more and more an area of important possible application.

In the year 2000, NSF and EFRI jointly sponsored a workshop in Playa del Carmen in Mexico, held in the same place and the same week as our ADP workshop of that year, to address the use of new more powerful global optimization methods to try to optimize the grid as one integrated system. One of the key talks was given by Ganesh (Kumar) Venayagamoorthy, then a graduate student under joint supervision of two of my PIs (Ron Harley and Don Wunsch), who has since become a leader in the effort to develop a true intelligent power grid, and held major annual conferences of his own out of Clemson.

Connections between fuzzy logic, neural networks and evolutionary computing also became ever stronger in this period.

For the electronics industry as a whole, the most important event towards the end of this period was the end of Moore's Law for speed. Basically, our PCs reach 3GHz speed then, and they are still there now. This contributed to a major growth of interest in Cellular Neural Networks (CNN), which had large conferences of its own (a few coordinated with IJCNNs), led particularly by Chua and Roska. The effort to shift towards better use of massively parallel processing [21,22] has become more and more intense since then.

5 MORE RECENT DEVELOPMENTS

In my view, the most important single development in the neural network field was the COPN initiative from NSF [23] and its outcomes. COPN (Cognitive Optimization and Prediction) was only a one-time funding effort, but the impacts are still visible. The program announcement for COPN was the result of intensive substantive discussions of program directors all across NXF engineering, and beyond. Those discussions were themselves a major crossdisciplinary activity. In a new paper in press [24], I explain at some length why the specific approach in COPN is essential to making real progress towards understanding the brain, unlike some other approaches to new initiatives which are less likely to produce positive breakthroughs (but do entail some serious risks).

One of the four large awards from COPN was to Andrew Ng and Yann LeCun. Like many of us in the neural network field, they had already known that "deep learning" is one of the core pillars of all true neural network design, since the 1980's or earlier. But the COPN award gave them the funds needed to actually do large-scale demonstrations on competitive benchmarks which classic AI people had guarded closely as their own fiefdoms for decades. During the term of the grant, they reported numerous breaking of world records in highly competitive benchmarks in image recognition, speech recognition and natural language, especially. The success in those efforts attracted further support, initially from Google and DARPA, and has very recently led to large new communities of people using these tools. LeCun's policy of making software available on an open-source basis has been another factor in the resulting breakthrough, which LeCun has called "the second rebirth of neural networks." Juergen Schmidhuber of Switzerland has also become a major force in the new developments, and builds on an understanding of many fundamental issues.

Another COPN award expanded the work of Harley and Venayagamoorthy towards a truly intelligent power grid. The Independent System Operators (ISOs) which run most of the power grid in the US today already make heavy use of powerful computers and algorithms to optimize flows of electricity, but they are running across fundamental limits in speed even as they have need to upgrade their algorithms a lot to fully account for stress form renewables and new loads. To achieve a breakthrough in speed requires massively parallel computing, and new types of neural network better able to cope with spatial complexity [3]. Venayagamoorthy has been moving ahead very forcibly to develop a new way to handle large-scale prediction and control issues in electric power, with links to industry and other nations, making use of physical cellular neural networks implementing new neural network topologies.

In my paper this year for the quantum computing track at SPIE, I look ahead to the future, to new possibilities to harness analog quantum computing to achieve levels of intelligence far beyond anything we have attempted in the past. But this year I have also retired from NSF; in this new era of sequestration and polarized politics, it is hard to predict how far we will go in living up to the incredible opportunities waiting for us, in these or other new technology areas.

References

[1] W.T.Miller, R.Sutton & P.Werbos (eds), *Neural Networks for Control*, MIT Press, 1990, now in paper
[2] Lewis, Frank L. and Liu, Derong, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, Wiley, 2012
[3] P. Werbos, From ADP to the Brain: Foundations, Roadmap, Challenges and Research Priorities, in Prof. Int'l Conf. Neural Networks, 2014. IEEE. http://arxiv.org/abs/1404.0554
[4] Fogel, David B., et al. "A self-learning evolutionary chess program."*Proceedings of the IEEE* 92.12 (2004): 1947-1954.
[5] E.A.Feigenbaum and J.Feldman, *Computers and Thought*, McGraw-Hill, 1963.
[6] Allen Newell and Herbert Simon, Human Problem Solving, Prentice-Hall, 1972

[7] J.Von Neumann and O.Morgenstern, *The Theory of Games and Economic Behavior*,Princeton NJ: Princeton U. Press, 1953.
[8] H.Raiffa, *Decision Analysis*, Addison-Wesley, 1968.
[9] M. Minsky & S. Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.
[10] D.O.Hebb, *The Organization of Behavior*, Wiley, 1949.
[11] P.Werbos, The elements of intelligence. Cybernetica (Namur), No.3, 1968.
[12] P. Werbos, P. Intelligence in the brain: A theory of how it works and how to build it, Neural Networks, Vol. 22, Issue 3, April 2009, Pages 200-212
[13] P. Werbos, Neural networks and the experience and cultivation of mind. *Neural Networks* 32 (2012): 86-95.
[14] P. Werbos, Backwards differentiation in AD and Neural Nets: Past Links and New Opportunities. In Martin Bucker, George Corliss, Paul Hovland, Uwe Naumann & Boyana Norris (eds), *Automatic Differentiation: Applications, Theory and Implementations*, Springer (LNCS), New York, 2005. Also posted at www.werbos.com/Mind.htm#mouse
[15] P.Werbos, *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*, Wiley, 1994
[16] P.Werbos, Applications of advances in nonlinear sensitivity analysis, in R.Drenick & F. Kozin (eds), *System Modeling and Optimization: Proc. IFIP Conf. (1981*), Springer 1992; reprinted in [15].
[17] White & D.Sofge, eds, *Handbook of Intelligent Control*, Van Nostrand, 1992
[18] http://drpauljohn.blogspot.com/2009/10/could-terminator-ii-come-true-true.html
[19] P.Werbos, Building and understanding adaptive systems: A statistical/numerical approach to factory automation and brain research, *IEEE Trans. SMC*, Jan./Feb. 1987.
[20] Albus, James S. "Outline for a theory of intelligence." *Systems, Man and Cybernetics, IEEE Transactions on* 21.3 (1991): 473-509.
[21] R. Kozma, R. Pino, and G. Pazienza (eds), *Advances in Neuromorphic Memristor Science and Applications*, Springer, 2012.
[22] Adamatzky, Andrew, Andrew Adamatzky, and Guanrong Chen. *Chaos, CNN, Memristors and Beyond: A Festschrift for Leon Chua*. World Scientific Publishing Co., Inc., 2013.
[23] National Science Foundation, *Emerging Frontiers in Research and Innovation 2008*, http://www.nsf.gov/pubs/2007/nsf07579/nsf07579.pdf
[24] P. Werbos, "How Can We Ever Understand How the Brain Works?", in Brain Dynamics (approximate title), Freeman and Kozma eds, forthcoming
[25] Andrew Ng, Y. Dan, E. Boyden and Y, LeCun, EFRI-COPN Deep Learning in the Mammalian Visual Cortex, http://www.nsf.gov/awardsearch/showAward?AWD_ID=0835878&HistoricalAwards=false
[26] Werbos, Paul J. "Computational intelligence for the smart grid-history, challenges, and opportunities." *Computational Intelligence Magazine, IEEE* 6.3 (2011): 14-21.