

From one of the editors of the renowned book Beyond Calculation comes a new collection of equal brilliance. Focusing on the impact of computers on humans, Talking Back to the Machine features nineteen world-leading experts who tell us about how computers will affect the ways we live, learn, think, teach, communicate, and relate to each other in the coming decades. The essays describe the myriad ways, both good and bad, in which our lives will be altered by information technology, and what we can do to influence these changes.

With Contributions By

Gordon Bell Ioel Birnbaum **James Burke** Vinton Cerf Peter J. Denning Bran Ferren Fernando Flores Murray Gell-Mann Reed Hundt David J. Kasik Brenda Laurel Pattie Maes Carver Mead Nathan Myhrvold William Perry Raj Reddy **Elliot Soloway** Bruce Sterling Maurice Wilkes

PRAISE FOR Beyond Calculation

"... a compelling book.... I think everyone will take away from this book a particular vision that captures their imagination...."

-Nature

"... a useful, interdisciplinary benchmark of the everevolving state of computer capability...."

-Computers in Physics

An Imprint of Springer-Verlag New York, Inc. 175 Fifth Avenue New York, NY 10010

ISBN 0-387**-98413-**5

www.copernicus-ny.com



GORDON BELL

The Folly of Prediction

James Burke

Our first speaker is known throughout the industry as the Father of the Minicomputer, although I'm sure he has mixed feelings about this. And I guess he is the embodiment of the saying, "You only know where you're going if you know where you've been." He's been a very significant part of the industry since the very early days, so when he talks about the matter of predicting, his views are grounded in considerable experience of doing just that in his 23 years, for instance, as vice president of R & D at Digital Equipment Corporation. He was educated at MIT. From 1966 to 1972, he was professor of Computer Science and Engineering at Carnegie Mellon, and from 1986 to 1987 was the first assistant director of the National Science Foundation Computing Directorate. He led the National Research Network Panel that became the NIIGII and authored the first high-performance computer-and-communications initiative. Today, he's an industry consultant-at-large and senior researcher at Microsoft, concerned principally with telepresence. So you could say he's been around

the block. He's written numerous books and papers on computer structure and startup companies, and in 1991 he published High Tech Ventures The Guide to Entrepreneurial Success, so you may want to take notes. Currently, he's on the technical advisory boards of Ambert Adaptive Solutions, Cyrus Logic, DES, Fake Space, University Video Communications and others. He's also director of the Bell Mason Group, supplying expert systems for venture development to startups, investors, governments and entrepreneurial initiatives. His many awards include the Institute of Electrical and Electronics Engineers Von Neumann Award, and the National Medal of Technology for his "continuing intellectual and industrial achievements in the field of computer design, and for his leading role in establishing computers that serve as a significant tool for engineering, science and industry." I suppose what that all adds up to is you're about to hear it from the horse's mouth on a subject we are all desperate to get right: predicting the future. The title of his talk-and this is where you can tell he has a handle on the matter-is "The Folly of Prediction." Ladies and gentlemen, please welcome Gordon Bell.

Gordon Bell

Why Do We Predict?

What is this urge we have to predict? Where does it come from? The desire to foretell has always been with us and is perhaps a selective advantage for our species, since our lives and fortunes have frequently depended on guessing correctly. With respect to social Darwinism, we might say that the ability to predict accurately affords a society a selective advantage. In ancient times, the astronomer priests looked to the heavens to predict the time for the spring planting. Joseph's Biblical prediction of seven lean years after seven fat years led to seven years of saving for that proverbial rainy day. Prediction is also a source of individual power. One who could predict a solar eclipse in an age when such phenomena were feared was an important person indeed. The one who could predict the course of an illness, the next rainfall or the return of the caribou acquired power and prestige.

Here we are concerned with the history and science of prediction as it applies to computing. While such prediction is not usually a matter of life and death, the motivations are the same: profit, prestige, power, safety, privacy and security. At times we engage in prediction merely for amusement or out of curiosity. It enriches our lives by providing a vision of what we can become as individuals and as a society, and by offering us the challenge to achieve our vision. Moore's law, which describes the exponential rate of increase of chip

The Folly of Prediction

technology, is one such challenge and vision. Another was presented to the scientific community in an article by Vannevar Bush in the July 1945 *Atlantic*. There, he outlined a vision—prescient conceptually if not in the specific technologies—of how knowledge would be organized in the future.

On a more commercial level, it is necessary for corporations to plan-at least that's what business school teaches—in order to avoid surprises. No manager wants to be taken unawares. So they start planning, and the next thing they know, they're given a budget to realize their plans, and now they have to maintain this budget, which requires further planning...

Market predictions have become an integral part of the corporate culture. In order to raise capital, there has to be a business plan complete with all sorts of predictions. I find it amusing that this has given rise to a professional class of predictors—a sort of latter-day coterie of sibyls or Delphic oracles. People actually pay other people to predict a new market they're entering, be it penbased computing or video on demand. They have no idea whether there will be a market for their product, but the oracle has spoken because they paid it to speak that way, and on the basis of that they go out and raise venture capital to make the vision come true.

And sometimes it does. Our predictions can help us extract large sums from the federal government or other funding sources. For example, now that we've got a teraflop, we could predict that if we had a petaflop, we could build a better bomb by next year, or journey to Jupiter in 2001 or perhaps create a virtual Jupiter right here on Earth. If we are creating grand challenges for ourselves, perhaps it is merely to perpetuate our own funding so we can keep on doing exactly what we're doing.

Finally, let's not forget predictions for the sake of idle curiosity. It's fun to make predictions, to guess at what the future will bring. Perhaps it's a way of transcending time, of grabbing a piece of tomorrow—or maybe it's just because I like to make bets. I've never been to Las Vegas, but I get my gambler's frisson by betting on the future of the digital universe. I have a little fun and perhaps earn a little money on the side. In 1996, I bet my friend Jim Gray that there would be videophones in wide use by 2001, and I expect to lose! The bet was designed to encourage me to work to make this come true, or in the words of Alan Kay: "The best way to predict the future is to invent it."

Learning From Our Past Predictions

Since the birth of the computer, many predictions have been made about the future of digital technology, some of which have become well-known. Probably the most famous post-Babbage, post-Turing and pre-computer-industry prediction about computing was made in 1943 by Thomas Watson, chairman of IBM, who said, "I think there is a world market for maybe five computers." Fifty years later we can all laugh at how wrong Watson was, but if we look at the first large-scale calculator that IBM built for Harvard—a 50-foot-long behemoth—we might learn that predictions require some history. Watson had no history of computers on which to base his prediction. In fact, we didn't even call them computers then—that's what we called the folks who did the computing! But his was considered a great prediction because it held for about 10 years. All predictions are at least implicitly time-limited—if nothing else, the principle of entropy will sooner or later put an end to everything. In general, the less historical precedent we have to go on, the shorter the time period of a prediction's validity. With no history at all, we might as well be reading tea leaves or entrails. We need history in order to predict!

Another set of predictions from the early days comes from a report I read for amusement from time to time: the 1969 report of the Naval Supply Command, which employed a panel of experts dubbed "Delphi" to forecast the state of computing over the next 15 years. It is interesting to look back and see what people were envisioning almost 30 years ago. Most of their predictions were wide of the mark, though a few were on target. One of the accurate ones had the use of punch-card readers declining after reaching a peak of 1500 cards per minute in 1974. Frequently, the panel's predictions were hurt by their ignorance of the market-some of the products they were predicting were already out there. So, while their punch-card prediction was correct, they were apparently unaware that in 1969 there were already readers available that exceeded their 1500-card-per-minute prediction. The Delphi panel also predicted that advances in computer memory would give us large memories on the order of four megabytes. At the time, memory was based on magnetic cores. Based on the state of technology, it seemed likely that magnetic cores would still be produced at the end of the next decade. But in the 1970s, metal-oxide semiconductor (MOS) memories appeared from nowhere and, like the comet that did in the dinosaurs, wiped out magnetic core.

In 1972, I gave a future-of-computing talk at MIT. My first prediction was that future computers would be both cheaper and faster, with every decade seeing a new platform based on reduced cost. I also predicted that the semiconductor companies eventually would become the computer manufacturers. This was right after the introduction of the first microprocessor, and I foresaw a time when there would be an entire computer on a chip, including the processor, memory and I/O. This is finally happening, and we now have

The Folly of Prediction

system-on-a-chip computer companies. After that, I declared that we were badly in need of networks, because we were at the point where people were becoming the networks that transported information between the various computers. If I had carried the thought further, I would have concluded that all computers needed to be linked together, just as the Internet does. My final prediction—one that was considerably less prescient—concerned semiconductor evolution. As I saw it in 1972, semiconductors were going to evolve for six more years and then level off. In fact, the number of transistors per chip and performance are still doubling every 18 months and are projected to continue doing so for at least another decade. This misprediction is understandable because as an engineer, I couldn't see beyond the projects that we were creating for the next two, three-year product generations.

A number of our predictions were overly optimistic. I recall my own 1960 forecast that, without a doubt, speech recognition would hit the scene by 1980. A 1962 prediction by the head of RCA Labs on the commercialization of speech was even more fanciful: He described a speech typewriter that would take in data from a microphone and produce printed pages; then another machine would translate that into the foreign language of one's choice; and finally, a third machine would spew forth the translated version as speech. To put it bluntly, speech predictions have been wildly optimistic and consistently wrong. My prediction that speech recognition would occur in 1980 was made after I had spent a year working on computerized speech. At the time, I said to myself, "I don't want to be in this business; this is a 20-year problem." Well, I was wrong. It was a 40-year problem, which is about where I usually am on my predictions—optimistic by a factor of two.

Another Delphi panel prediction had parallel processing being online by 1975. In fact, the Cray 1 confirmed it. I recall a wager I made with Danny Hillis who was sure that in 1995, massively parallel computers would do most of the scientific processing. His prediction was based on a DARPA (Defense Advanced Research Projects Agency) initiative to produce a massively parallel machine, which might actually have come to pass had the company not gone out of business. A 1,000-node multiprocessor—which did not make it by 1989 as I had predicted or planned and which I was asked to build by 1992—was finally available in 1996. The lesson: Parallelism is more difficult to implement than anyone imagines.

New Overtakes Old, Forever Thwarting Predictors

While an established technology gives us more of the same over time, eventually a new technology, which at the outset may be more expensive and less productive, replaces the old one as its superior qualities become apparent. As the new technology is developed, it becomes more efficient, faster and cheaper through better production techniques and a declining cost curve. While constantly improved performance is taking place at the top, from the bottom come new, lower-priced computers. This model, which I came up with in 1972, still holds true today. Newer and cheaper wins; the old dies off. The mainframe would be dead if the market based its decisions solely on the cost of operations, yet it is still hanging on just to hold our legacy data and run legacy programs. Based on the high cost of converting programs and data, one can almost predict that it will be with us for the next century.

So what can we really predict about computers? A few years ago I predicted the network computer, and my friends at Microsoft and Intel assured me that no one wanted anything that cost less than \$2,000. In 1998, the challenge is to produce home computers costing less than \$1,000 and network computers for only a few hundred. Then there's the system on a chip around which an industry is now forming that includes processing, memory and interfaces to the real world. The home area network has arrived, and now we're talking about various forms of body area networks—computers that are actually attached to us. Having had two heart attacks, I'm particularly interested in one such item—a cardioplastic implant, whereby a piece of back muscle is wrapped around the heart and becomes part of it. Some experts recently predicted cochlea implants, which are already here. Bionic limbs, whatever those are, will arrive by 2013. Artificial vision will take much longer because all the complex fiber optics will have to be coupled in.

Law I: Expect the Expected

History can be a reliable guide to prediction. Knowing what has happened in the past guides us in predicting how things will progress in the future. There are some cases where the past is completely reliable. Even before Newton propounded his law of universal gravitation, predictions of a sunrise every day and the return of spring every year were made on the basis of their regular occurrence. With computing, as with any human enterprise, predictions are more subtle, but knowledge of past performance can help. Thomas Watson's prediction that there would be a global market for only five computers needs to be greatly modified on the basis of the history over the past 50 years. With historical knowledge, we can create a mathematical model and extrapolate into the future. Perhaps the most well-known extrapolation is Moore's law, which asserts that circuit densities of semiconductors will continue to double approximately every 18 months or, equivalently, increase at a rate of about 60 percent a year. We started around 1972 with 1K of memory, so if we run that out to 2010, that gets us up to 40 gigabytes for a single memory. Moore's law comes straight from the product-development cycle, whereby new semiconductor processes, materials and products yield a new generation of chips every three years. Indeed, the product gestation time for a minicomputer was three years. For a microprocessor it was also three years. A mainframe's was four years, while supercomputers continue to take at least four.

While we've got Moore's law working for us, in January 1997 a single electron was stored in a seven-by-seven-nanometer cell at the University of Minnesota. Based on this achievement, Nathan Myhrvold, chief technology officer at Microsoft, has noted that by 2010 we could get two and a half petabytes on a chip, which would accelerate Moore's law by 30 years. So is Moore's law too conservative, and can it be accelerated?

Law II: Expect the Unexpected

If the future continues along the lines laid out by the past, then we should be confident about our predictions. However, as we've seen over and over, just when we thought we knew what we were about, a new technology comes along to upset historical precedent. So another law of prediction is to expect the unexpected: Sooner or later some cataclysmic event will upset the virtual apple cart. We saw this happen with magnetic-core memory. Core seemed destined to stay for a while until MOS wiped it out. The same thing happened when CMOS wiped out the bipolar logic that was used in minicomputers, mainframes and supercomputers. Cray Research hung on to bipolar too long, and the Japanese built a range of CMOS supercomputers about six years ahead of them, offering faster, lower-cost products. (Cray solved the problem by getting the U.S. government to levy huge dumping tariffs against the Japanese.) The lesson is that over the long term, unexpected events will cause discontinuity and rapid change. Thus, in making predictions we must try to build the unexpected into our calculations.

Law III: For the Short Term, Bet Against the Optimist

We take Moore's law seriously because it has proved to be a good indicator over the last 30 years. Unlike natural laws, Moore's law is about the collective behavior of all those involved with the semiconductor industry—researchers, materials and process suppliers, chip manufacturers and users. A prediction gains credibility if it has historical parallels or has been made by someone with a successful track record. For short-term predictions, bet against the optimists. Their vision is clouded by desire, and what they see as just around the corner might still be a long way off. The further such visionaries are from the reality of actually having to produce what they are predicting, the more overly optimistic they are likely to be. Conversely, if the person doing the predicting is a scientist or an engineer, someone actually at work on the project in question and aware of all the hurdles, they are likely to be quite conservative. Six years is the horizon of the engineer who is on a three-year product cycle. Products under development typically take three years or so to reach the market, so one hears short-term predictions that, in some area, progress will continue for six more years and then flatten out—that is, after two generational cycles.

With respect to longer-term predictions, the development cycles of science and engineering are good indicators. Here we can cite Carver Mead's rule, which states that it takes 11 years from the first observation of some phenomenon to the point where it can be commercially successful. Some examples of this are the invention of the transistor in 1946 to its adoption in a computer in 1957; integrated circuits in 1956 to their adoption in 1967. The first Arcnet LANs were developed in 1972, and adoption of the Ethernet happened in 1981—only nine years later. The time period from invention to adoption is shrinking. From the inception of the Web and the publishing of documents using HTML browsers to adoption in PCs and workstations was only a couple of years. With the Web, software now spreads like wildfire.

Law IV: Don't Assess the Market Based on Your Own Personal Characteristics

We can learn about the laws of prediction by examining failed predictions. Furthermore, it's gratifying to see those people we thought should have known better landing so wide of the mark. A former boss of mine, Ken Olsen, the president of Digital Equipment Corporation, predicted in 1977 that there would be no market for home computers. At that time I had been studying home use of computers among colleagues at DEC for about 10 years. He more recently predicted that his company, Modular Computer Systems, would not be on the Internet, which was not a very good prediction, since at the time they already had a home page. Out of such predictions we may infer the following law: It is a mistake to equate yourself with the average user unless you happen to be the average user, which no one is. On the other hand, computing has progressed very rapidly because the designers have been developing systems that they want to use in their work—essentially becoming producer and consumer simultaneously. However, since these designers are decidedly not the average user, a frequent result has been the im-

plementation of user-unfriendly interfaces of the type Unix users know all too well.

Law V: Predict with Exponential Data when You Have a Few Data Points

When we have historical data on some event, we can fit it on a curve and extrapolate it out into the future. Of course, the data can be put on more than one curve, and the less data, the more freedom we have in deciding what curve fits best. This is particularly true with exponential growth. Exponentialgrowth curves appear almost flat for a long while when plotted on a linear scale, then the growth seems to become linear and suddenly takes off. Therefore, we have to employ the exponential plot with care, using it only when the change is truly exponential, which happens with observed growth history, underlying production capacity (learning curves) or consumption (demand curves). On the other hand, if we're too conservative, we'll miss the boat. Sometimes, what looks like no growth or linear growth may merely be the slow-growth part of an exponential curve that will eventually take off. The growth of Internet traffic is a good example of that. We didn't see the huge growth in the Internet, prompting Bob Lucky, vice president of Bellcore, to remark in 1995, "If we couldn't predict the Web, what good are we?" The growth of Internet use was exponential all the time, but the slow part of the curve looked linear, particularly since nobody anticipated the invention of the World Wide Web-addressing, HTML and the browser. The Web appeared just in time to maintain the Internet's exponential growth.

Internet growth has always been exponential—doubling every year. That allows us to project its growth based on its history. We can predict, then, that there will be a crossover point somewhere around 2003 when there will be more Internet users than there are people in the world. Obviously that won't happen, but it needn't necessarily dampen our prediction, because in figuring Internet growth we may, by then, have to factor in the auto market, the lightswitch market, the camera market and who knows, maybe even the dog and cat collar market. The year 2003 prediction was based on four data points: the number of current Internet users, the observation that this number is doubling every year, and the current world population and its rate of growth. Nicholas Negroponte, head of MIT's Media Lab, predicts that by 2000 there will be one billion people using the Web. I think his estimate is a little bit high. That's more people than there are PCs and it's about the number of TVs and phones, so the current infrastructure will not suffice. We'll need something like radio links to get there. So I don't think its a good prediction, especially since I've bet him \$1,000 even money that by December 31, 2000 there won't yet be a billion Web users—each with their own address—and five-to-one odds that there won't be a billion by the end of 2001.

I prefer to use just a couple of data points for predicting exponentials, otherwise unnecessary and usually unhelpful complications are introduced. So with something that you believe is growing exponentially on the basis of technological development, the economy and the marketplace, two data points will suffice. More points simply don't add more information-it's like computing more digits in a measurement that exceeds the accuracy of your measuring tools. For example, consider the growth in processing speed. We can begin with Maurice Wilkes's 1949 EDSAC computer which processed 700 instructions per second. In 1951, there was a jump to 50,000 ips when parallel memories using CRT storage tubes, and eventually core memories, replaced serial delay lines and drum memories. After that, growth continued exponentially, and depending on what we use as our second point, we can predict from 20 percent to 40 percent growth per year. Growth in memory and storage has been similar, and in fact, network capacity parallels all of these. The slowest growth component of the network is telephony at about 15 percent per year. It seems, therefore, that the limiting factor is going to be our ability to access the network from our homes.

Law VI: The More You Spend, the More You (Might) Get

An important factor in the growth of a technology is the amount of money spent on a computer because, in general, the more you spend, the more you would like to get. So, if someone is willing to spend the money, you might get better results sooner than later. In the 1960s, computer pioneer Herb Grosch declared that computing power grows as the square of what is spent for a single computer. This law has been refuted many times, with most observers believing instead that performance grows as the square root of what is spent due to the extraordinary growth in the power of inexpensive microprocessors. Triggered by DARPA's Strategic Computing Initiative in the early 1980s, there have been wonderful results in high-performance computing. A 2,000fold speedup in the most expensive supercomputers occurred between 1987 and 1997. We can retrospectively predict this by Moore's law (a factor of 100); by spending more (a factor of two), since instead of spending a mere \$30 million, we now spend \$100 million dollars for the largest supercomputers; and finally, by the switch from custom-designed, low-volume ECL circuitry to high-volume, high-performance CMOS (Complementary Metal Oxide Semiconductor) microprocessors (a factor of 10).

10

In the mid-1990s, the Department of Energy's Accelerated Strategic Computing Initiative set a target of 2010 for petaflops-1,015 floating-point instructions per second, which will require another 1,000-fold speedup. As we've seen, we expect to get a factor of a 100 with Moore's law. A factor of two comes from spending more for one system. Now, instead of capping spending at a mere \$100 million, we can spend \$500 million for a massively parallel, scalable supercomputer. If we centralize all three DOE centers into one location, that boosts it by another factor of three; or alternatively, the network can be made fast enough to achieve this speedup. Increasing competition would perhaps generate another factor of three, but this has to occur based on the use of high-volume components. So petaflops by 2010 seems possible in theory. But in a funny way, our quest for the ultimate parallel computer is like parallel lines-they never meet. Our reach eternally exceeds our grasp, and the goal is always a constant distance away. Bill Wulf, president of the National Academy of Engineering and professor of Computer Science at the University of Virginia, predicts-or rather has a vision-that "one can imagine millions of hosts in a loose confederation that to users will look like a massively parallel desktop computer." This is exactly what the Internet is. Harnessing it for a single problem is the trick!

Cyberization Is Our Quest and Our Fate

What can we predict about the future of computing based on our understanding of the history of computers? We have to consider three main factors: the platforms, the networks that connect them and cyberization, which is the entire process of putting information into a computer. It includes all the components of a user interface such as WIMP (windows, icons, mouse, pulldown menu). It includes implants that couple a computer to a person-sensors, effectors, cameras, digital money that encodes bits. It includes a digital representation of the physical world-books, money, newspapers, stocks, pictures and eventually television. Fundamentally, cyberization is the process of making the whole world digital instead of analog so it can be available anywhere, any time-the essence of cyberspace. Finally, it includes "state" information about all the networks-highway networks (traffic at any point), power grids and water. This is where we're heading. I make this prediction based on the belief that cyberization has become our human and scientific quest to understand and achieve. We see that it is possible, and so it becomes our goal and fate. So cyberization is the interlinking of all our experiences and artifacts into a fractal network of networks of networks that start with the universe and end at our cars, our homes and our bodies.

Networks: Structure and Content

What will be the nature of this network? Will it be one network for data, a second network for such things as telephony and then other networks? Or will they merge into a single dial tone? I think we'd all like to have the latter.

Bandwidth Is Free, but None Is Available at That Price

There are many factors that influence the evolution of the network. Chief among them are commitment, money and vision. Irwin Dorros, who was head of AT&T Long Lines in 1981 before the breakup, thought that integrated services digital network would be ubiquitous by 1985. He was wrong because at the time there was no need for such a network. If ISDN ever becomes the ubiquitous connection, it will clearly be by default. It's not something any of us are anxious to have, but in fact, it's going to be in the communications line. ISDN was an expensive investment that had no application and was never finished. When it should have been deployed for computing, the telephone companies didn't do it. Now that we have the computers to connect, the bandwidth is too low for video (by a factor of 10 to 20), and the presence of ISDN has inhibited investment in a faster network. Furthermore, conventional POTS (Plain Old Telephone Service) is running at half the speed of ISDN, so evolution is catching up with what was to be revolutionary simply because of economics and commitment of deployment. I have worked in this area a long time, and from my experience I can say that network bandwidth becomes available more slowly than anyone can ever predict. So, beware of predictions based on the notion that bandwidth is free. It's free alright—you can't get any of it at that price.

Vision and Faith in Science

Now for a prediction I'm happy to make. It goes back 50 years to Vannevar Bush's *Atlantic* article in which he wrote, "There will always be plenty of things to compute in the detailed affairs of millions of people doing complicated things." So, I feel I can safely predict that we won't run out of things to compute, which means that those of us in computing have chosen a wonderful area. On the basis of Bush's predictions, I conclude that faith in science and a vision of what can be useful are good predictors. In his article, Bush outlined an information storage and retrieval system with hyperlinks that he dubbed "memex," which many believe was the original vision for the World Wide Web.

Of course, in making predictions it helps to be a Vannevar Bush. It also helps to be really lucky, because none of the technology that Bush outlined

The Folly of Prediction

turned out the way he saw it. Bush had been responsible for technical manpower during World War II, so he had seen some amazing developments including radar, the jet engine and various automatic control systems. But all of these inventions pale in comparison to the integrated circuit and the computer. With the amazing things that have already been done with them in such a short time, and the even more amazing things—many as yet unimaginable—that will be done with them in the future, they will, I believe, turn out to be among mankind's greatest inventions.