THE BRAIN

PREDICTION AND INTELLIGENCE

MONDAY, MARCH 21, 2005

JEFF HAWKINS

CHAIRMAN & EXECUTIVE DIRECTOR
REDWOOD NEUROSCIENCE INSTITUTE

MARCH 20 TO 22, 2005 | FAIRMONT SCOTTSDALE PRINCESS | SCOTTSDALE, ARIZONA

PC FORUM is the IT industry's premier executive conference for business decisionmakers, investors, entrepreneurs, thinkers and policymakers. In its 28th year, the Forum provides a stimulating and constructive environment to do business and to exchange ideas with peers. There are no canned presentations, sponsors, exhibition booths or hype at this event. Instead, it provides a serene setting in which business partnerships are established and strengthened, debate and discussion flourishes, and strategies are reset for the year ahead. The next PC Forum will be held on March 12 to 14, 2006, at the La Costa Resort in Carlsbad, CA. For more information, visit www.pcforum2006.com or e-mail pcforum@release1-0.com.

THE BRAIN

PREDICTION AND INTELLIGENCE

JEFF HAWKINS
CHAIRMAN & EXECUTIVE DIRECTOR
REDWOOD NEUROSCIENCE INSTITUTE

ESTHER DYSON, CNET NETWORKS One thing John Thompson mentioned during the last panel is the issue of what you recognize. Recognizing stuff and predicting patterns is what Jeff Hawkins is now going to talk about. He spoke here, much to my amazement, 12 years ago. Many of you who were here remember him and, of course, have run into him in the industry elsewhere. Some of his wonderful, memorable ideas are close to where he can commercialize them. You might find them useful in your products, such as for detecting security threats or whether a user is ready to buy your product so you can make him a better offer.

JEFF HAWKINS, REDWOOD NEUROSCIENCE INSTITUTE It's a pleasure to be back here, and I appreciate you inviting me to speak. I also appreciate you giving a copy of my book to everyone here. When I checked in yesterday, I opened my little pink-red bag and said to the lady behind the counter, "I probably don't need this." She was a bit disappointed and looked at me like, "Oh, you don't want our gift?" I said, "Well, I'll take it." I have been designing a lot of technologies, and many of them have reached the market. Many people know me for my work at Palm and Handspring and now PalmOne, but I have another side of my life, which I'm going to talk about today. When I was a teenager, I made a list of questions that were interesting to me. The No. 1 question was, Why does the universe exist? It seemed like nothingness was more likely than somethingness, but here we are. The second question was, If it exists, why does it behave the way it does? Why are the laws of physics what they are? Why does $E=MC^2$ and not MC^3? The third question was, Given the laws of physics, what is life, and how did it come about? And the last question was, Given there is life, what's intelligence, and how do brains work? I stopped there because I felt that problem was perhaps the most interesting and the most fundamental of them all. To even ask the question of what it means to understand something requires having a brain and having intelligence. I decided early in my life that I was going to work on understanding how the brain works, because it gets to the nature of what it means to be human. Why we make mistakes, why there are bad guys, why we have wars and disagreements about fundamental beliefs in the world are all a product of our brains. I also thought that if you could figure out how the brain works and you could build machines that work on the same principles, those machines would be a great tool for solving the other three problems. And I figured this could be done in my lifetime—a typical young hubris that we could solve how the brain works and build intelligent machines—and I set out to do it.

I've always been a part-time entrepreneur. The other part of my time has been spent working on brain research. It was slow in the beginning, but I've been making steady progress. When I talked here in 1993, I had one key insight, which is that brains are all about prediction. Three years ago, I started Redwood Neuroscience Institute, which is a nonprofit scientific institute that brings people together on a particular problem. In those three years, we've made rapid progress. My book, "On Intelligence," which came out in October of last year, talks about what the neocortex does and how it does it. It's fairly readable if you haven't opened it yet; don't be intimidated. There's only one chapter that's hard: chapter 6. You can skip it, but don't miss chapters 7 and 8. The last chapter talks about an industry that could be created around machines that work the same way, but I don't tell you how to build such a machine because I didn't know. But now I do. In the last year, Dileep George, a graduate student at Stanford University, made progress in expressing the neocortex in mathematical terms. Once you've done that, you can create it in software and hardware, and we started doing that, too. I believe we're on the verge of creating a major new technology, and there's a lot of ramifications. So that's what my talk is about, which is a lot to do in the 15 minutes I have remaining, but I'm just going to give you the flavor for it. I'm going to introduce you to the neocortex and tell you what it does and how it does it. Then I'm going to tell you how we can build machines that work the same way. I brought some props with me. I brought my plastic model of a human brain. It's about the size of a young adult's brain. You could roughly say that there are two parts to it: the old part, which is your brain stem, and this big pink thing around the top here, which is your neocortex. This is where everything's happening. My neocortex is talking; yours is listening, if you're awake. If I were to take your neocortex out of your head and iron it flat, it would look like this. [*Holds up a cloth dinner napkin.*]

This is a dinner napkin. It's about a thousand square centimeters. Your neocortex is about this size. The napkin is about two millimeters thick. Your neocortex is just a little bit thicker than this. We have about 30 billion cells in this dinner napkin and about 300 trillion connections. No one really knows the exact number. All of your knowledge, everything you've learned in life, is stored in those cells and connections, and we want to understand how it works.

There are two things that are really remarkable about the neocortex. One is that different regions do different things; it's very functionally organized. There are areas that deal with vision, hearing, touch, language, mathematics, music, motor planning and so on. The other fact is that each region looks the same physically. Vernon Mountcastle speculated in 1979 that the way the neocortex does vision is the way it does hearing, which is the way it does language and math. It sounds hard to believe, but it's true. What makes the vision area vision is how it's connected to the other parts of the neocortex. So the eyes come into the visual area, the ears come into the auditory area and so on. But there's a common cortical algorithm, so if we understand how the neocortex works solving a specific problem, we're going to understand how it works solving all the problems. A solution to vision is going to be a solution to language, motor planning, mathematics, science, business and so on. So that's my introduction to the neocortex. Now onto what it does. It is a memory system. It is not a computer; it does not have a CPU or a clock or anything like that. Memories store patterns and recall patterns. It's a memory system that's different than any one you've ever used before. Through exposure—like growing up in the world, looking around, hearing things, touching things—it builds a model of the world. When you come across new things, which is every moment of your waking life, it says, "Let me recall from my memory model what things like this happened in the past," and it makes predictions about the future. Most behavior—like my speech—is controlled by your neocortex and is a byproduct of prediction. If you understand how the brain stores patterns and makes predictions, you'll understand how it makes motor behavior, which I'm not going to get into any more detail today. I'm just going to talk about prediction.

Basically, when I'm talking about prediction, it's something you're doing subconsciously all the time. I've never stood on this stage before, and I've never touched this lectern like this, but I'm going to put my hand down right now, and I felt it hit the lectern. I had expectations about it. My brain was predicting what it was going to feel before it felt it. I know that because if my hand went right through the lectern, I would have been very surprised. I would say, Something is wrong here! If it felt like cloth or metal or water or Jello, I'd be very surprised. So the point is that my brain kind of knew, even though it's never felt this thing before. It says, "It's going to feel like wood because I saw it, and it better be right about here." You're making predictions every second of your waking life. As you step, as you look around the room, your brain is saying, "I know what's going to happen. At least, I'm trying to know what's going to happen." Because you're listening to me and I'm speaking in English, your brain is predicting what word will be at the end of this… [*Pauses*]

…sentence. [*Laughter*]

You're doing this all the time. It's not something you're necessarily conscious of as you're doing it. Just a brief understanding of how this fits into the evolutionary scheme. [*Shows a slide of a reptilian brain.*]

Here's a simple picture of a reptile brain. A reptile has senses; it has sophisticated behavior. It does not have a neocortex. You don't have to have a neocortex to have an interesting life. [*Laughter*] But all mammals have a neocortex, and only mammals do. The neocortex is literally this new memory system that sits on top of the old brain. It remembers what was going into the old brain, and it feeds back predictions into the old brain. It lets you see and imagine the future and lets you anticipate things. You can anticipate things better than a crocodile or an alligator. You can imagine what you're going to see when you go through this room or the first time you enter a men's room or ladies' room or whatever. What happened with humans is that our cortex got very large, so we have a very big model of the world. Compared with other animals with large brains, ours has uniquely taken over much of our motor control, so the same algorithm is making me speak. That's why we have language and why we are a different species than other animals.

Now on to how it does it. In the next picture of a brain, we see the different regions of the neocortex. You see that region A might be connected to region B by the big bundles of fibers. When region A connects to region B, B connects to A. So they're connected both ways, but they're asymmetric, meaning the connections one way look different than the connections the other way. If you spend several years of your life mapping it out, you get a very complex picture that looks like this. [*Shows slide of the hierarchical connectivity of a monkey's neocortex.*]

This is a map of a monkey's neocortex. The little boxes are regions of the brain. The lines are showing which way the regions are connected. The only point I want to make about this is that it's in a hierarchy. In this picture, the patterns come in from the bottom, from the eyes, the body. The information moves up from region to region to region, and flows back down from region to region to region. Here is a very simplified view of this for a human. Each of these little boxes is a region of the cortex that is doing something different. At the bottom, you have a whole bunch of little visual, auditory, motor and sensory regions, and they're connected in this converging hierarchy. It's like a big inverted tree. As patterns come in, they flow back down again and recall things, and they make predictions about what happens next. So if I see something, it'll lead to predictions about what I'm going to see, which may lead to predictions about what I'm going to hear. Let's give an example. Watch my hands. [*He claps.*]

I just clapped my hands together. Your brain saw my hands come together, and it had a visual prediction of what was going to happen. It expected my hands to either stop or bounce apart. It did not expect them to pass through each other. You'd have been surprised if they had turned into potatoes or something else. At the same time, that pattern comes up the visual hierarchy and down the auditory hierarchy, leading you to a prediction about what you're going to hear. You expected to hear a clap. If you did not hear a clap or if it was delayed, you would have been surprised. If it sounded like a pig squeal, you would have been surprised. Your brain had very precise predictions. You're doing this all the time in your waking life, and you're not aware of it.

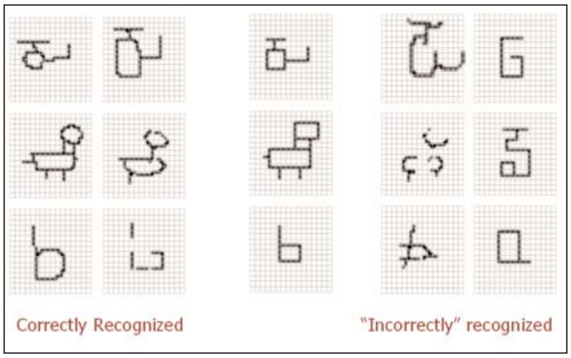
If you can understand what one region of the cortex does, you can understand what all of it does. Here is a simplified version of what it does; the more complex version is in the book. Basically, every region stores sequences of patterns. The memory is memories of sequences. A sequence is like a melody. So you have a series of patterns coming in. A region says, "Oh, I recognize this sequence," if it can, and it says, "If I recognize it, I'm going to pass the name of the sequence up"— not the details, but like the name of the song. So the region above is now looking at a sequence of names, and it forms sequences of those and so on up the hierarchy. At the top of the hierarchy, as the pattern moves down, it can invoke a series of patterns. Those invoke another series, then another, and so on down the hierarchy. So I can say, "Give 20-minute brain talk," and that makes me talk for 20 minutes, basically with a very high-level concept coming down. I can do this over and over again if you'd like to hear it. [*Laughter*]

Then, similarly, when you have very complex patterns coming in at the bottom, they leave the more stable patterns at the top. That is what we do when we are trying to figure out the causes of the world. What causes security problems? What is language about? This is the basic architecture of the brain. So if I'm a region in the cortex and I know the sequence of patterns, and I'm listening to my little melody, I know what's going to happen next, like the word at the end of the sentence or the next note in a melody. I pass that down the cortex; so I coalesce time going up, and I unfold time going back down. This hierarchical structure of the cortex actually captures the hierarchical structure of the world. This is how the world is organized, and this is how your perception of the world is organized. We basically have structure within structure within structure and sequences of time within sequences of time. It takes a bit longer than this to really get a sense for it, but let's move on. This theory explains lots of things about psychology, learning and behavior. You can really understand what thinking is. You can explain what creativity, intuition and prejudice are. Those are all the same thing, and they're unavoidable, so that's an unfortunate part of life. We can really understand a lot about what consciousness is, how we learn, and why it takes 18 years. I don't have time to talk about all that stuff, but it does play into that field quite a bit. Chapter 7 of the book talks about some of these topics.

> "[THE BRAIN] LEARNS SEQUENCES AND STORES CONDITIONAL PROBABILITIES, MEANING IT COMPARES A SERIES OF LEARNED SEQUENCES TO A SET OF CAUSES. IT STORES THIS IN A CONDITIONAL PROBABILITY MEMORY TABLE, WHICH IS BASICALLY JUST A TABLE OF NUMBERS. WHEN YOU PUT THESE TOGETHER IN A HIERARCHY, WHICH IS WHAT THE CORTEX IS, YOU HAVE A TREE OF CONDITIONAL PROBABILITY FUNCTIONS."
>
> –JEFF HAWKINS

I now want to talk about building intelligent machines. This is the new stuff that you're not going to find in the book. This is a little technical, but just bear with me. You'll get the gist of it. Each region of the cortex talks in the language of probability theory. A region of the cortex doesn't say, "Oh, this is song A." It says, "I think it's song A, but it could be B or C or D, and here's my probability distribution." Every part of the brain is constantly saying, "Here's my distribution of probability of what I think is going on and what sequence I'm listening to." What it's storing is sequences. Patterns come in from the bottom, it learns sequences and stores—here's the most technical part of the talk—conditional probabilities, meaning it compares a series of learned sequences to a set of causes. The pattern coming in might be associated with a set of causes. It stores this in a conditional probability memory table, which is basically just a table of numbers. When you put these together in a hierarchy, which is what the cortex is, you have a tree of conditional probability functions. Many years ago Judea Pearl developed a technique to understand trees of conditional probability functions—believe it or not, people spend a lot of time on this. He was not thinking about the neocortex at that time nor did he understand how the neocortex works when he did this work. So you can't just take this technique and plop it down; you have to modify it a bit. But if you take this thing called belief propagation and you match it with sequences and a few other tricks that we throw in, you actually can get the whole thing to work. I'm going to give you an example of that.

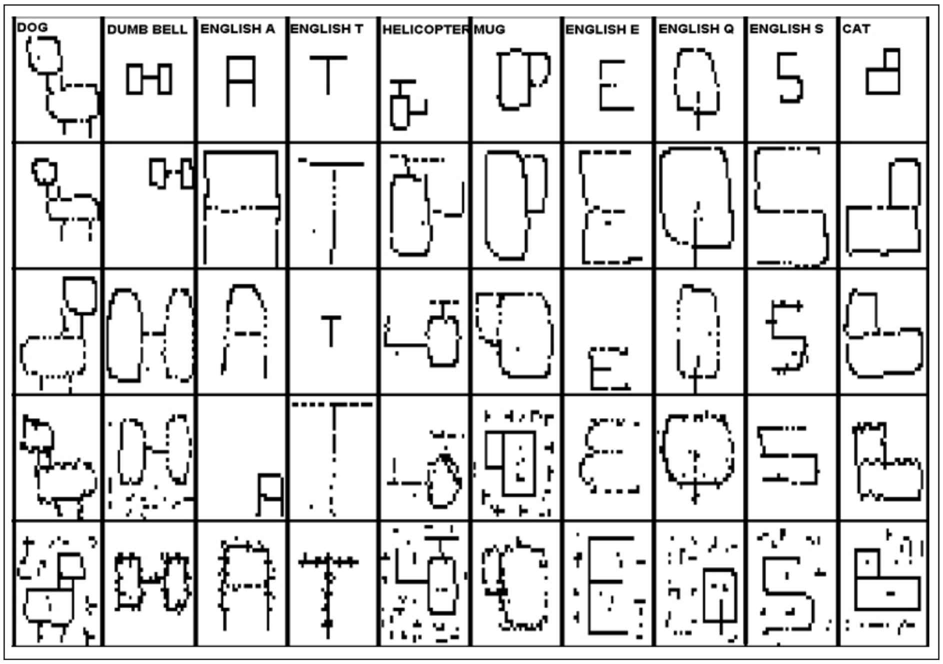
There is no software today that you can show a picture to and have it say what it is. There's nothing that can say, "This is a cat," or a dog or a microphone or a coffee cup or piece of candy. It's a very hard problem. On the other hand, humans do it very easily. So we decided to do a visual pattern recognition experiment. We made a visual hierarchy like the visual hierarchy of the cortex. We only used three levels to the hierarchy. There's not a lot of memory involved in this one. These are the 90 symbols that we wanted our software to recognize. They are simple black-and-white line drawings, no color, no gray scale. So we simplified things, but this is still a very hard problem. The trick is, we were going to train it on some examples, but also going to show it all kinds of weird stuff it's never seen before. We want to see if it can figure out what it's looking at.



Recognition: Examples

This center column [*see slide, left*] shows the symbols we trained it on. We call this the helicopter, the dog and the letter "b." The way we train it is that we move it through time. Basically, it's like your child holding an object and moving it through space. Time is absolutely essential to this whole system. You can't get it to work without it. Later we can show it characters which it didn't see.

Here you see variations of the dog and the letter "b." It recognizes them all, and I'll tell you what it means to recognize it. Basically, it forms a belief about what it's seeing. In the fourth column are samples that it did not recognize as those things, but it recognized the things in the fifth column. This slide shows you symbols that were recognized correctly. [*See slide, below.*] Here in the first column is our little dog. You notice the dog can be looking left or right; it can be distorted in all kinds of ways—different sizes, different locations. We didn't have to program this into the system. It just learns to do this. It learns in the same way that you learn to do this stuff. It uses hierarchical memory; we show it some patterns, and it learns to make these beliefs about it. In the bottom of the sixth column you see this little coffee cup. You can see how noisy this can get. It's hard to know that's a cup, but this system does it. Obviously, the question you might ask is, can this be scaled? I believe it can. I'm certain of it. But you might be skeptical, and I would believe that.



Correctly recognized test cases

I needed to come up with a name for this technology. Since it's a technology, it doesn't have to have a very consumer-friendly name, so I picked something very descriptive. I call it hierarchical temporal memory or HTM. I'm going to impress upon you that I believe it is a fundamental technology. It's as fundamental as in the 1950s when they were starting to understand programmable computers. We're just starting to understand how this thing works. Today, we can build this in software. Tomorrow, we'll build it in hardware. It's applicable to many, many problems. It's applicable to standard AI problems, such as vision, language and robotics. It can be applied to non-typical AI problems, such as weather, markets, math and science. It's anywhere you have structured data where you're trying to find the causes of some sensory input, like El Niño is a cause for weather. It's a very high-level structure for a cause for weather, and you want to discover those things. That's what this system does. We can build these things to be faster than humans; we can build them with higher capacity than humans; we can build them with a deeper sense of thought or deeper ability to understand. You can build very exotic senses into it—in addition to eyes and ears and skin it can have sonar and radar. It can have special senses like weather systems or things that look at molecules. The last thing I want to point out is that it's not threatening. I'm not talking about building replicating robots. This is not Stepford Wives or C-3Po. These are tools that don't feel. They're not emotionally laden, and they don't feel imprisoned or anything like that. But they're going to think, and they're going to understand their world.

If you want to learn more, here's what you should do. You've got the book; read it. If you like it, give it to someone else and ask them to read it. If you don't like it, then that's fine. I'm in the process of starting a business to promote the technology and to maximize the impact of the technology. I've been somewhat inundated with people who have read my book and want to work with me. These are scientists, academics and researchers. I'm trying to figure out a structure in which to go forward with this. The business hasn't been announced yet, but it will be shortly. If you want to find out when it gets announced, go to RNI.org and sign up for the e-mail notice. You will only get one e-mail and one e-mail only to tell you about the new business. If you want to delve into some other details, you can browse the book's website, OnIntelligence.org. It has a discussion forum, the software for the simulation is there and some technical papers and things like that. And that's it. I'm done. Thank you. [*Applause*]

DYSON Thanks. We've carefully arranged a break because Jeff is going to stick around. I'm sure lots of you have questions for him. To me this is tremendously exciting, and I hope you share the excitement.

ABOUT THE SPEAKER

JEFF HAWKINS CHAIRMAN & EXECUTIVE DIRECTOR, REDWOOD NEUROSCIENCE INSTITUTE

Jeff Hawkins is the chairman and executive director of the Redwood Neuroscience Institute, a non-profit scientific research institute focused on understanding how the human neocortex works. He is also the architect of many computer products including the PalmPilot, Visor and Treo families of handheld computers and communicators. Hawkins co-founded Palm Computing in 1992 and Handspring in 1998, and is currently CTO at palmOne. In 1980 Hawkins dedicated his career to understanding how our brains perform high-level cognitive tasks, but concluded that the time was not right for his theoretical approach to understanding the brain, so he returned to industry with the intent of resuming his goal some years later. In 2002 Hawkins created the Redwood Neuroscience Institute as a place where research into cognitive theory could flourish. His theories on the inner workings of the brain are described in his book *On Intelligence*.