The Inscrutable Lines of Cause and Effect.

© 2011 by Byron Reese

[[ NOTE: Most of this talk is from two chapters of a book I have written called Golden Age 2.0: *How the Internet will end ignorance, disease, hunger, poverty and war.* I have just completed the book and am looking for a publisher. Please email any suggestions or contacts to byronreese@gmail.com. ]]

There is a drug called Wellbutrin that is prescribed for depression. It has a long and successful clinical history.

Some number of years ago, a few people taking Wellbutrin reported that their cravings for cigarettes went down. GlaxoSmithKline, the makers, tested it and found out that the drug was a highly effective aid in smoking cessation. So they repackaged it under the name Zyban and it, the exact same drug, is prescribed to help people stop smoking.

Think about that. No one contemplated Wellbutrin as a smoking drug. It only became one because users of the medicine happened to notice it had this effect.

How many unseen connections like that are there in the universe? Causes and effects we cannot perceive because they are buried in the infinitude of data that populates the universe. I think there are millions.

Another example:

Children in schools with fluorescent lights get fewer cavities than those in schools with incandescent lighting. Why is this? Fluorescent lighting increases the body’s production of saliva which helps prevent cavities.

The subtle interplay of everything in the universe with everything else in the universe is dramatically more complex than we are able to comprehend. Our ability to find cause and effect within it is highly suspect. We stumble upon causal associations almost as by accident, like the scientific equivalent of “Your chocolate is in my peanut butter.”

However, this state of affairs will soon change due to the convergence of four trends; the reduction in cost to essentially zero of four things. Three of these are well known to everyone in this room: the cost of storing data, the cost of moving data, and the cost of performing calculations on data.

The fourth one of these is a bit a more subtle. It is the reduction to essentially zero of the cost of collecting and recording data. Presently, there is a miniscule amount of data that is actually in digital form. I am not talking about taking medical records that are in paper form and making them digital. I mean all the trillions of pieces of data created every second on this planet. Soon, all of this data will be collected passively due to the widespread proliferation of cheap sensors, whose cost is also rocketing towards essentially zero. This covering of the planet in sensors will not be done by any central organization, it is already happening because of the immense improvements data can provide to our lives. A camera mounted on my house that scans the faces of everyone knocking on my door and checking to see if they are criminals is a good unto itself, but also passively collects data. A skillet that detects botulism and remembers everything ever cooked in it so that it can be recreated is quite useful, and generates data. The cost of collecting this will fall to zero soon, it will become ubiquitous, and the impact of this, combined with the other three trends, will be profound.

Already, as you pass through modern life, you leave a Digital Echo, a picture of who you are and what you are doing. More and more of your everyday life leaves such an echo.

You can see this today. When you fill up your car with gas your location and activity are digitally recorded, presumably for all time. It comes in the form of your credit card statement, that gives an accurate, albeit extremely abbreviated, record of your comings and goings. Your cell phone logs everyone you call.

Take a moment with me and think about if this is pushed to its logical extreme. What if everything you did was logged? We will talk about privacy in a minute, but assume that this is passive collected and completely private.

First off, imagine every word you said was recorded by your personal recorder and automatically transcribed. This would be very useful. No more straining your mind to remember what you promised the client you would have by Friday – you just look up the transcript.

Then imagine GPS is layered in. Very accurate GPS that tracks your every move, even in your own home.

Next, imagine everything you do is remembered. Not just that you went to a certain address but that the address was a movie theater and based on where you sat you saw Episode VII of Star Wars.

But let’s get more detailed. You are watching TV and flipping through the channels. Every channel you pause on, every channel you watch, every channel you come back to - all perfectly logged.

In the kitchen, your pan logs everything you cook in it. Each piece of silverware you own knows you are the holder, and calculates the calories you consume. Your house is aware of every room you are in and what you did in it.

So think about everything being recorded. Everything you buy. Every meal you order. Every restaurant you visit. Every word you type. Every book you read. The time at which you have set your sprinkler to turn on. When you last went to the dentist. Everything you saw, that your eyeball tracked to, and how long you looked at it. Not just everything you ever looked at, but your physiological response. Did your eyes dilate? Your pulse increase? Your muscles tighten? Did you smile?

Imagine if it was all recorded. All your medical records. All your tax records. Every song you download and how many times you play it. Every person you meet. Everything your body does. Every heartbeat. Every bite you eat, every step you take. Every breath you breathe.

Imagine if it is all recorded. All of it. A real Digital Echo of your life.

Love it or hate it, this is the inevitable direction technology is heading. This logging is a bi-product of all kinds of things we want computers to help us with. It is profound because if we can passively collect all the data our lives generate, we will have created a collective memory for the planet. A record of every cause and every effect.

This would be a monumental accomplishment. This is not just “the end of forgetting.” It would mean finally that humanity has an extremely robust collective memory and we need not learn and forget, learn and forget for eons. We could learn and remember. Everyone’s life experience, every action and its outcome, becomes the data that makes everyone else’s life better.

Let’s talk a bit about privacy.

I am interested in things we have built the Internet to do that have no offline corollary. These things teach us about ourselves. I already knew people wanted to sell the stuff in their attic or send money to other people before eBay or PayPal came along, because the offline world already invented garage sales and Western Union.

But Twitter? Well, that tells us something about ourselves we didn’t know. In fact it tells us a lot of things about ourselves, the kind of information we want to share and the immediacy of it.

Overall, I am really proud of the kinds of things we are building the Internet to do. It reflects well on us. We are building the Internet to connect with each other better, to share information, to collaborate, to offer mutual support, and so on. I know the list of nefarious uses of the Internet, but on balance, we are building it for good purposes.

One pretty big surprise was how much we are willing to expend time and energy for strangers with no hope of anything in return. No money, no fame, no glory, no nothing. You see it all over the Internet. You post some personal problem you have and people will give you advice. They will take time to write a great big forum post just for you, a total stranger they will never meet.

It’s nice, isn’t it? We have a natural desire to want to help others.

So, if people feel their privacy is being safeguarded, they will be willing contribute their data anonymously to the greater good. When I buy a book on Amazon, I know they use that fact, anonymously, to suggest that book to someone else later.

Over time, people will see that they will be contributing to the construction of a knowledge base which really is the complete recording of every action of every member of humanity. And that this information will be used to better the lives of other people in very real ways, which we will soon see.

Let’s assume that most people do share, anonymously, their Digital Echoes, the complete experiential logging of their life. And assume we have this ability to ask questions that have never been asked and get instant answers. A giant Wolfram Alpha interface. You could start looking around for lines that connect things we didn’t previously think were connected.

The machine could get you started by surfacing correlations that would not be expected. What if it surfaced that people who drink a certain kind of coffee have, on average, higher incomes and further that people who switched to this coffee subsequently made more money? Or that a certain group of people who do a seemingly unrelated set of a dozen activities report levels of happiness higher than average?

Or to continue with fictional cases: Why is it that gasoline made from oil refined at some refinery burns more efficiently? Why is it the population of some endangered animal is declining everywhere but one place? Why is it that traffic jams are fewer in a certain city than any other city its size? That dropout rates in some schools are lower than demographically matched schools anywhere else in the world?

Let’s dive deeper into, say, the medical implications of all of this:

Imagine for instance that everyone with skin cancer voluntarily shared their Digital Echo file on an anonymous basis. Say they also included the mapping of their genome, as well as every single thing they did in their daily lives. Every food they ate. Every exercise they did. Every song they heard. Everything that had been passively logged about their life.

With skin cancer, like all diseases, over time, some people get better and some people get worse, and we really don’t know why. Imagine a computer culling through this massive amount of data, inconceivably large, and pulling out patterns. A finding might look like this: “People who eat radishes get better slightly more than people who don’t.”

Now, you don’t know if the radishes make the people get better or if something that makes people crave radishes also beats back skin cancer. So what do you do? We dig deeper into the data.

Then we see that it is only people in a certain parts of the country that are getting better. Why would that be the case? The data shows certain pockets where radish efficacy is substantially higher and others where it is nonexistent. The computers would then see that most of the people who got better bought their radishes in stores that were stocked from certain farms. And not just certain farms, but farms that used a certain pesticide. Not just a certain pesticide, but pesticides that contain a certain chemical.

The machines surface this information. We act on it. Brilliant doctors try to understand what is happening and drug companies commission studies to see if this particular chemical, frequently found in radish pesticides, known to be harmless to humans, might just cure skin cancer.

So they run trials, and they find that in majority of cases it does cure skin cancer; and yet, in some it has no effect. The machines again churn on the data and surface that it is effective when a person has a certain common, though not universal, genetic marker.

Wow. Now, all of a sudden, you have this magnificent piece of knowledge. That for people who have a certain genetic marker, this radish pesticide ingredient will cure their skin cancer.

Findings of this kind are impossible today. There is no mechanism that can decipher the complex interactions of everything on the planet with everything else on the planet. Today, we discover things like “Wellbutrin helps people quit smoking” through chance and dumb luck.

You can see where this is all headed. Instead of creating hypothesis and constructing experiments, then modifying hypothesis and retesting, we can make all of history a test we can control for and sift through the trillions of pieces of data for meaningful information.

The world is full of unexplained anomalies. Why do people who win Academy Awards outlive people who are nominated but do not win? Why do first basemen live longer than anyone else on the team? It is said that tall people live shorter lives than short people. Why is this? Are taller people heavier and thus they die sooner? One study suggested the entire difference could be accounted for simply by the fact that when tall people trip, they fall further and are thus hurt more. There are many things like this that probably have some underlying reason that we can learn from. I am sure they are not actionable, but these cases are not really how the science will be used. They are illustrative of the principle. In actuality, scientists will study far more subtle associations of diet, activity, genetics, behaviors, and so on that might give us an insight in how to live longer, happier, and better.

My guess is we won’t have to absorb all of this information. If the system (Conspirator will call it “The System”) determines people like me are happier and healthier with more Thiamin in their system, it will come up in my custom vitamin formulation which also includes whatever nutrients I didn’t get enough of that day. Since I would certainly trust my doctor or nutritionist to create a custom vitamin formula for me based on all of my medical records and health conditions, I would be massively more comfortable letting the system do it based on every bite I swallow, every bit of my genetic code, and every byte of data it has access to about a billion people. My body would be kept in perfect condition, constantly monitored and optimized. All safely because it is built on collective memory and experience of the entire planet. It is the end of the age an anecdotal evidence, where because “Aunt Martha takes a teaspoon of honey every morning and hasn’t been sick in years” therefore I should as well.

Then we will come to understand the outliers better. Why do some people live to 120? They inexplicably seem to be the ones who eat a stick of butter every night before bed. Why is this? What is it about them and their lives that made them live so long or so well? Is it all genetic? If so, which genes? Is it behavioral? Psychological? I am eager to know.

Why do some people keep their mental faculties so late in life? Folklore tells us that doing crossword puzzles helps keep the mind active? Is this an old wives tale? Well, we will know in the future. We will even know why. Perhaps mental agility is a result of their extensive exposure to a chemical in pencil lead and newsprint that they got by doing all those puzzles. We will know. We will know how to live our lives best to maximize any and all factors. We may not choose to – we may prefer cheesecake and bacon instead of an estimated extra 2.4 months of life. But the choice will be ours and will be made based on data.

This future will come gradually. More and more data will be passively collected. More people will participate. The passage of time will grow the repository. Successes will come, encouraging more data collection and more people to participate. Once this ball gets rolling, it will speed up quickly.

This new way to do science will not be limited to only scientists. Anyone could sit down at their desk, formulate a hypothesis, ask the machine a question, receive an answer, modify the question, until causalities are found. This method allows us to treat the entire world as a controlled experiment in retrospect.

The world will still need ever-smarter specialists doing every more complex work. I am not saying the research scientist loses out to the florist in Akron, Ohio. I am, however, saying that the florist will be one of the millions of minds accessing the data looking for associations and testing hypotheses. She will be the one who figures out that people living in apartments whose floors are coated with a certain sealant are three times as likely to be taking anti depresseants. We would never find that today. How much potential is there in millions of discoveries like that?

Finally, this system will not just solve for human illness, but all kinds of other problems as well. Patterns in crimes will be discovered. Why do some houses get broken into and others don’t? Can the system learn to predict crime targets? All kinds of social issues will be able to be examined. Why are some areas poorer than others? Why do people stay in school in certain areas more than others? And on and on and on. A record of all human activity, with anonymity safeguards in place, allows us all to become part of the solution. All of us to put our minds to work on the problems of the world.

I take great comfort in this. And I think the answers to life’s mysteries are buried there in the data waiting for the florist in Akron to coax them out.

The world can thusly be transformed. Most of our vexing problems as a civilization are technological in nature. Disease is a technological problem. Hunger is a technological problem. Poverty is a technological problem. And being that they are technological problems, they will have technological solutions.

And it all starts here. The ability to ask a complex question that has never been asked before and get an answer is the core distinctive of WolframAlpha. You will tell your grandchildren that you saw it all happen, that you were there the moment the world changed. And since it changed for the better, wondrously better, you can rightfully claim to have been part of its forming.