

September 7, 2012

# The Weatherman Is Not a Moron

By NATE SILVER

From the inside, the National Centers for Environmental Prediction looked like a cross between a submarine command center and a Goldman Sachs trading floor. Twenty minutes outside Washington, it consisted mainly of sleek workstations manned by meteorologists working an armada of flat-screen monitors with maps of every conceivable type of weather data for every corner of the country. The center is part of the [National Weather Service](#), which Ulysses S. Grant created under the War Department. Even now, it remains true to those roots. Many of its meteorologists have a background in the armed services, and virtually all speak with the precision of former officers.

They also seem to possess a high-frequency-trader's skill for managing risk. Expert meteorologists are forced to arbitrage a torrent of information to make their predictions as accurate as possible. After receiving weather forecasts generated by supercomputers, they interpret and parse them by, among other things, comparing them with various conflicting models or what their colleagues are seeing in the field or what they already know about certain weather patterns — or, often, all of the above. From station to station, I watched as meteorologists sifted through numbers and called other forecasters to compare notes, while trading instant messages about matters like whether the chance of rain in Tucson should be 10 or 20 percent. As the information continued to flow in, I watched them draw on their maps with light pens, painstakingly adjusting the contours of temperature gradients produced by the computers — 15 miles westward over the Mississippi Delta or 30 miles northward into Lake Erie — in order to bring them one step closer to accuracy.

These meteorologists are dealing with a small fraction of the 2.5 quintillion bytes of information that, I.B.M. estimates, we generate each day. That's the equivalent of the entire printed collection of the Library of Congress about three times per second. Google now accesses more than 20 billion Web pages a day; the processing speed of an iPad rivals that of last generation's most powerful supercomputers. All that information ought to help us plan our lives and profitably predict the world's course. In 2008, Chris Anderson, the editor of *Wired* magazine, wrote optimistically of the era of Big Data. So voluminous were our databases and so powerful were our computers, he claimed, that there was no longer much need for theory, or even the scientific method. At the time, it was hard to disagree.

But if prediction is the truest way to put our information to the test, we have not scored well. In November 2007, economists in the Survey of Professional Forecasters — examining some 45,000 economic-data series — foresaw less than a 1-in-500 chance of an economic meltdown as

severe as the one that would begin one month later. Attempts to predict earthquakes have continued to envisage disasters that never happened and failed to prepare us for those, like the 2011 disaster in Japan, that did.

The one area in which our predictions are making extraordinary progress, however, is perhaps the most unlikely field. Jim Hoke, a director with 32 years experience at the National Weather Service, has heard all the jokes about weather forecasting, like Larry David's jab on "Curb Your Enthusiasm" that weathermen merely forecast rain to keep everyone else off the golf course. And to be sure, these slick-haired and/or short-skirted local weather forecasters are sometimes wrong. A study of TV meteorologists in Kansas City found that when they said there was a 100 percent chance of rain, it failed to rain at all one-third of the time.

But watching the local news is not the best way to assess the growing accuracy of forecasting (more on this later). It's better to take the long view. In 1972, the service's high-temperature forecast missed by an average of six degrees when made three days in advance. Now it's down to three degrees. More stunning, in 1940, the chance of an American being killed by lightning was about 1 in 400,000. Today it's 1 in 11 million. This is partly because of changes in living patterns (more of our work is done indoors), but it's also because better weather forecasts have helped us prepare.

Perhaps the most impressive gains have been in hurricane forecasting. Just 25 years ago, when the National Hurricane Center tried to predict where a hurricane would hit three days in advance of landfall, it missed by an average of 350 miles. If [Hurricane Isaac](#), which made its unpredictable path through the Gulf of Mexico last month, had occurred in the late 1980s, the center might have projected landfall anywhere from Houston to Tallahassee, canceling untold thousands of business deals, flights and picnics in between — and damaging its reputation when the hurricane zeroed in hundreds of miles away. Now the average miss is only about 100 miles.

Why are weather forecasters succeeding when other predictors fail? It's because long ago they came to accept the imperfections in their knowledge. That helped them understand that even the most sophisticated computers, combing through seemingly limitless data, are painfully ill equipped to predict something as dynamic as weather all by themselves. So as fields like economics began relying more on Big Data, meteorologists recognized that data on its own isn't enough.

**The I.B.M. Bluefire** supercomputer in the basement of the [National Center for Atmospheric Research](#) in Boulder, Colo., is so large that it essentially creates its own weather. The 77 trillion calculations that Bluefire makes each second, in its mass of blinking lights and coaxial cable, generate so much radiant energy that it requires a liquid cooling system. The room where Bluefire resides is as drafty as a minor-league hockey rink, and it's loud enough that hearing protection is suggested.

The 11 cabinets that hold the supercomputer are long and narrow and look like space-age port-a-potties. When I mentioned this to Rich Loft, the director of technology development for NCAR, he was not amused. To him, this computer represents the front line in an age-old struggle to predict our environment. “You go back to Chaco Canyon or Stonehenge,” Loft said, “and people realized they could predict the shortest day of the year and the longest day — that the moon moved in predictable ways. But there are things an ancient man couldn’t predict: ambush from some kind of animal, a flash flood or a thunderstorm.”

For centuries, meteorologists relied on statistical tables based on historical averages — it rains about 45 percent of the time in London in March, for instance — to predict the weather. But these statistics are useless on a day-to-day level. Jan. 12, 1888, was a relatively warm day on the Great Plains until the temperature dropped almost 30 degrees in a matter of hours and a blinding snowstorm hit. More than a hundred children died of hypothermia on their way home from school that day. Knowing the average temperature for a January day in Topeka wouldn’t have helped much in a case like that.

The holy grail of meteorology, scientists realized, was dynamic weather prediction — programs that simulate the physical systems that produce clouds and cold fronts, windy days in Chicago and the morning fog over San Francisco as they occur. Theoretically, the laws that govern the physics of the weather are fairly simple. In 1814, the French mathematician Pierre-Simon Laplace postulated that the movement of every particle in the universe should be predictable as long as meteorologists could know the position of all those particles and how fast they are moving. Unfortunately, the number of molecules in the earth’s atmosphere is perhaps on the order of 100 tredecillion, which is a 1 followed by 44 zeros. To make perfect weather predictions, we would not only have to account for all of those molecules, but we would also need to solve equations for all 100 tredecillion of them at once.

The most intuitive way to simplify the problem was to break the atmosphere down into a finite series of boxes, or what meteorologists variously refer to as a matrix, a lattice or a grid. The earliest credible attempt at this, according to Loft, was made in 1916 by an English physicist named Lewis Fry Richardson, who wanted to determine the weather over northern Germany on May 20, 1910. This was not technically a prediction, because the date was some six years in the past, but Richardson treated it that way, and he had a lot of data: a series of observations of temperature, barometric pressures and wind speeds that had been gathered by the German government. And as a pacifist serving a volunteer ambulance unit in northern France, he also had a lot of time on his hands. So Richardson broke Germany down into a series of two-dimensional boxes, each measuring three degrees of latitude by three degrees of longitude. Then he went to work trying to solve the equations that governed the weather in each square and how they might affect weather in the adjacent ones.

Richardson’s experiment failed miserably. It “predicted” a dramatic rise in barometric pressure that hadn’t occurred and produced strange-looking weather patterns that didn’t resemble any seen in Germany before or since. Had he made a computational error? Were his equations

buggy? It was hard to say. Even the most devoted weather nerds weren't eager to solve differential equations for months on end to double-check his work for one day in one country six years in the past.

What Richardson needed, he thought, was more manpower. He envisioned a weather-forecasting center with some 64,000 meteorologists, all working simultaneously to have the computational speed to make accurate weather forecasts in real time. His dream came to fruition (sort of) in 1950, when the first computer weather forecast was tried by the mathematician John von Neumann and a team of scientists at the Institute for Advanced Study in Princeton, N.J. They used a machine that could make about 5,000 calculations a second, which was quite possibly as fast as 64,000 men. Alas, 5,000 calculations a second was no match for the weather. As it turned out, their forecast wasn't much better than a random guess.

**Our views about** predictability are inherently flawed. Take something that is often seen as the epitome of randomness, like a coin toss. While it may at first appear that there's no way to tell whether a coin is going to come up heads or tails, a group of mathematicians at Stanford is able to predict the outcome virtually 100 percent of the time, provided that they use a special machine to flip it. The machine does not cheat — it flips the coin the exact same way (the same height, with the same strength and torque) over and over again — and the coin is fair. Under those conditions, there is no randomness at all.

The reason that we view coin flips as unpredictable is because when we toss them, we're never able to reproduce the exact same motion. A similar phenomenon applies to the weather. In the late 1950s, the renowned M.I.T. mathematician Edward Lorenz was toiling away in his original profession as a meteorologist. Then, in the tradition of Alexander Fleming and penicillin or the New York Knicks and Jeremy Lin, he made a major discovery purely by accident. At the time, Lorenz and his team were trying to advance the use of computer models in weather prediction. They were getting somewhere, or so they thought, until the computer started spitting out contradictory results. Lorenz and his colleagues began with what they believed were exactly the same data and ran what they thought was exactly the same code; still, the program somehow forecast clear skies over Colorado in one run and a thunderstorm in the next.

After spending weeks double-checking their hardware and trying to debug their code, Lorenz and his team discovered that their data weren't *exactly* the same. The numbers had been rounded off in the third decimal place. Instead of having the barometric pressure in one corner of their grid read 29.5168, for example, it might instead read 29.517. This couldn't make that much of a difference, could it? Actually, Lorenz realized, it could, and he devoted the rest of his career to studying strange behaviors like these by developing a branch of mathematics called chaos theory, the most basic tenet of which is described in the title of his breakthrough 1972 paper, "Predictability: Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?" In other words, a small change in initial conditions can produce a large and unexpected divergence in outcomes.

Chaos theory does not imply that the behavior of the system is literally random. It just means that certain types of systems are very hard to predict. If you know the exact conditions of a coin as it leaves someone's hand, you can — with the right laboratory equipment — predict, almost perfectly, which side it will land on. And yet the slightest disturbance to that motion can change a coin toss from being almost wholly predictable to almost wholly unpredictable.

The problem with weather is that our knowledge of its initial conditions is highly imperfect, both in theory and practice. A meteorologist at the National Oceanic and Atmospheric Administration told me that it wasn't unheard-of for a careless forecaster to send in a 50-degree reading as 500 degrees. The more fundamental issue, though, is that we can observe our surroundings with only a certain degree of precision. No thermometer is perfect, and it isn't physically possible to stick one into every molecule in the atmosphere.

Weather also has two additional properties that make forecasting even more difficult. First, weather is nonlinear, meaning that it abides by exponential rather than by arithmetic relationships. Second, it's dynamic — its behavior at one point in time influences its behavior in the future. Imagine that we're supposed to be taking the sum of 5 and 5, but we keyed in the second number as 6 by mistake. That will give us an answer of 11 instead of 10. We'll be wrong, but not by much; addition, as a linear operation, is pretty forgiving. Exponential operations, however, extract a lot more punishment when there are inaccuracies in our data. If instead of taking  $5^5$  — which should be 3,125 — we instead take  $5^6$ , we wind up with an answer of 15,625. This problem quickly compounds when the process is dynamic, because outputs at one stage of the process become our inputs in the next.

Given how daunting the challenge was, it must have been tempting to give up on the idea of building a dynamic weather model altogether. A thunderstorm might have remained roughly as unpredictable as an earthquake. But by embracing the uncertainty of the problem, their predictions started to make progress. "What may have distinguished [me] from those that proceeded," Lorenz later reflected in "The Essence of Chaos," his 1993 book, "was the idea that chaos was something to be sought rather than avoided."

**Perhaps because chaos** theory has been a part of meteorological thinking for nearly four decades, professional weather forecasters have become comfortable treating uncertainty the way a stock trader or poker player might. When [weather.gov](http://www.weather.gov) says that there's a 20 percent chance of rain in Central Park, it's because the National Weather Service recognizes that our capacity to measure and predict the weather is accurate only up to a point. "The forecasters look at lots of different models: Euro, Canadian, our model — there's models all over the place, and they don't tell the same story," Ben Kyger, a director of operations for the National Oceanic and Atmospheric Administration, told me. "Which means they're all basically wrong." The National Weather Service forecasters who adjusted temperature gradients with their light pens were merely interpreting what was coming out of those models and making adjustments themselves. "I've learned to live with it, and I know how to correct for it," Kyger said. "My whole career might be based on how to interpret what it's telling me."

Despite their astounding ability to crunch numbers in nanoseconds, there are still things that computers can't do, contends Hoke at the National Weather Service. They are especially bad at seeing the big picture when it comes to weather. They are also too literal, unable to recognize the pattern once it's subjected to even the slightest degree of manipulation. Supercomputers, for instance, aren't good at forecasting atmospheric details in the center of storms. One particular model, Hoke said, tends to forecast precipitation too far south by around 100 miles under certain weather conditions in the Eastern United States. So whenever forecasters see that situation, they know to forecast the precipitation farther north.

But there are literally countless other areas in which weather models fail in more subtle ways and rely on human correction. Perhaps the computer tends to be too conservative on forecasting nighttime rainfalls in Seattle when there's a low-pressure system in Puget Sound. Perhaps it doesn't know that the fog in Acadia National Park in Maine will clear up by sunrise if the wind is blowing in one direction but can linger until midmorning if it's coming from another. These are the sorts of distinctions that forecasters glean over time as they learn to work around potential flaws in the computer's forecasting model, in the way that a skilled pool player can adjust to the dead spots on the table at his local bar.

Among the National Weather Service's detailed records is a thorough comparison of how well the computers are doing by themselves alongside the value that humans are contributing. According to the agency's statistics, humans improve the accuracy of precipitation forecasts by about 25 percent over the computer guidance alone. They improve the temperature forecasts by about 10 percent. Humans are good enough, in fact, that when the organization's Cray supercomputer burned down, in 1999, their high-temperature forecasts remained remarkably accurate. "You almost can't have a meeting without someone mentioning the glory days of the Cray fire," Kyger said, pointing to a mangled, half-burnt piece of the computer that was proudly displayed in the office where I met him. "If you weren't here for that, you really weren't part of the brotherhood."

Still, most people take their forecasts for granted. Like a baseball umpire, a weather forecaster rarely gets credit for getting the call right. Last summer, meteorologists at the National Hurricane Center were tipped off to something serious when nearly all their computer models indicated that a fierce storm was going to be climbing the Northeast Corridor. The eerily similar results between models helped the center amplify its warning for [Hurricane Irene](#) well before it touched down on the Atlantic shore, prompting thousands to evacuate their homes. To many, particularly in New York, Irene was viewed as a media-manufactured nonevent, but that was largely because the Hurricane Center nailed its forecast. Six years earlier, the National Weather Service also made a nearly perfect forecast of [Hurricane Katrina](#), anticipating its exact landfall almost 60 hours in advance. If public officials hadn't bungled the evacuation of New Orleans, the death toll might have been remarkably low.

**In a time when** forecasters of all types make overconfident proclamations about political, economic or natural events, uncertainty is a tough sell. It's much easier to hawk overconfidence, no matter if it's any good. A long-term study of political forecasts conducted by Philip Tetlock, a

professor at the University of Pennsylvania, found that when political experts described an event as being absolutely certain, it failed to transpire an astonishing 25 percent of the time.

The Weather Service has struggled over the years with how much to let the public in on what it doesn't exactly know. In April 1997, Grand Forks, N.D., was threatened by the flooding Red River, which bisects the city. Snowfall had been especially heavy in the Great Plains that winter, and the service, anticipating runoff as the snow melted, predicted that the Red would crest to 49 feet, close to the record. Because the levees in Grand Forks were built to handle a flood of 52 feet, a small miss in the forecast could prove catastrophic. The margin of error on the Weather Service's forecast — based on how well its flood forecasts had done in the past — implied about a 35 percent chance of the levees' being topped.

The waters, in fact, crested to 54 feet. It was well within the forecast's margin of error, but enough to overcome the levees and spill more than two miles into the city. Cleanup costs ran into the billions of dollars, and more than 75 percent of the city's homes were damaged or destroyed. Unlike a hurricane or an earthquake, the Grand Forks flood may have been preventable. The city's flood walls could have been reinforced using sandbags. It might also have been possible to divert the overflow into depopulated areas. But the Weather Service had explicitly avoided communicating the uncertainty in its forecast to the public, emphasizing only the 49-foot prediction. The forecasters later told researchers that they were afraid the public might lose confidence in the forecast if they had conveyed any uncertainty.

Since then, the National Weather Service has come to recognize the importance of communicating the uncertainty in its forecasts as completely as possible. "Uncertainty is the fundamental component of weather prediction," said Max Mayfield, an Air Force veteran who ran the National Hurricane Center when Katrina hit. "No forecast is complete without some description of that uncertainty." Under Mayfield's guidance, the National Hurricane Center began to pay much more attention to how it presents its forecasts. Instead of just showing a single track line for a hurricane's predicted path, their charts prominently feature a cone of uncertainty, which many in the business call "the cone of chaos."

Unfortunately, this cautious message can be undercut by private-sector forecasters. Catering to the demands of viewers can mean intentionally running the risk of making forecasts less accurate. For many years, the Weather Channel avoided forecasting an exact 50 percent chance of rain, which might seem wishy-washy to consumers. Instead, it rounded up to 60 or down to 40. In what may be the worst-kept secret in the business, numerous commercial weather forecasts are also biased toward forecasting more precipitation than will actually occur. (In the business, this is known as the wet bias.) For years, when the Weather Channel said there was a 20 percent chance of rain, it actually rained only about 5 percent of the time.

People don't mind when a forecaster predicts rain and it turns out to be a nice day. But if it rains when it isn't supposed to, they curse the weatherman for ruining their picnic. "If the forecast was

objective, if it has zero bias in precipitation,” Bruce Rose, a former vice president for the Weather Channel, said, “we’d probably be in trouble.”

The National Weather Service, on the other hand, takes plenty of blame when its cautious forecasts seem retrospectively unwarranted. I was reminded of this when I arrived in Tampa for the Republican National Convention. The city was briefly in Hurricane Isaac’s cone of chaos before the storm took a westward tack. The airport and roads were remarkably quiet, no doubt because some reporters and delegates (and thousands of tourists) heeded caution and stayed home. When I so much as mentioned the weather forecast, my taxi driver turned and launched into a series of obscenities.

*This article is adapted from “The Signal and the Noise: Why So Many Predictions Fail — but Some Don’t,” to be published this month by Penguin Press.*

*Kevin Erskine is a professional pseudonym used by the Dutch photographer Erik Hijweege.*

*Nate Silver runs the [FiveThirtyEight](#) blog. His book on forecasting and prediction, “The Signal and the Noise,” will be published this month.*

*Editor: [Jon Kelly](#)*

*This article has been revised to reflect the following correction:*

**Correction: September 23, 2012**

*An article on Sept. 9 about weather forecasting gave the wrong solution for 55. It is 3,125 — not 3,215.*