

Bob

Behn's Performance Leadership Report

An occasional (and maybe even insightful) examination of the issues, dilemmas, challenges, and opportunities for improving performance and producing real results in public agencies.



Vol. 10, No. 8, April 2013
Copyright © 2013 by Robert D. Behn

On the analytical challenge of

Distinguishing Between Outliers & Harbingers

At the beginning of September 2011, the record of the Boston Red Sox baseball team—83 wins and 52 losses—was the best in the American League. The statisticians at **CoolStandings** “calculated” that the team had a better than 99 percent chance of making the playoffs.

Yet, as all of New England remembers, the Sox crashed and burned. The team lost twenty of its last twenty-seven games. Very bad. In fact, Nate Silver, who writes the **FiveThirtyEight** political calculus blog for *The New York Times* suggested that the “September Collapse of the Red Sox Could Be Worst Ever.”

Yet, not only was September 2011 a disaster for the Red Sox. April 2012 was only slightly better; the team won 11 and lost 11. In May, the team won 15 and lost 14. Then things really deteriorated. The Sox ended the 2012 season losing ten straight games for a record of 69 wins and 93 losses. Very, very bad.

Like Silver and the CoolStats guys, John Henry, the Sox's principal owner, is also a professional number cruncher. He built his fortune trading global commodity futures based on sophisticated statistical formulas.

Still, that doesn't mean that Henry—the smart and rich quant guy—knew how to interpret the data from that September 2011 disaster. Like everyone else in baseball, Henry operates with a lot of data—and not just data on wins and losses. He has data on every pitcher's individual pitch; on every batter's individual swing (or choice not to swing); on every ball hit to left field with two outs in the bottom of the eighth inning and a runner on first; on every . . .

From his abundance of data, John Henry and his management team—including more quant guys with a director of “Baseball Information Services”—could never figure out what to do to make the team better in 2012.

This should not be particularly surprising. Learning from data is not

necessarily easy. Interpreting the data is inevitably difficult. Do the data indicate some kind of significant trend? Or is what looks like a major change in direction with some real significance actually just one more brief blip in the data, little more than another random fluctuation? It's difficult to tell.

It's even hard for the big-time quant guys to tell. As the Red Sox owner explained a year after the September 2011 collapse. “What appeared to be an outlier month in September 2011 turned out to be a harbinger instead.”

How do you distinguish an outlier from a harbinger? The cute answer, of course, is “in retrospect.” Some time later—weeks, months, years, or

Learning from data is not necessarily easy. Interpreting data is inevitably difficult. Do the data indicate some kind of trend? Or is what looks like a big change with some real significance little more than another random fluctuation? It's difficult to tell.

maybe decades later—it will be obvious to all. And the randomness in the data, which will be reflected in the randomness of the early prognostications, means that some can assert that their prediction (a.k.a. guess) was prescient. Looking backwards, it is relatively easy to distinguish the blips from the trends.

That, however, is a little late. You want to act quickly on the trends, while you ignore the blips. First, however, you have to distinguish between them.

It may be a natural, human tendency to dismiss a weird data point as an outlier. After all, it is just one data point. How can you predict a trend from a single—and obviously weird—piece of data. And yet, every new trend has to start with a single

(usually small) change in the data.

So when confronted with that weird data point, the key analytical question is: What, besides the randomness in the world, might explain this datum? Does there exist the beginning of a small, yet important trend that this apparently random blip may explain?

Of course, the Boston Red Sox did not just have an outlier data point. They had an outlier month. An entire month of weird data. Yet, sometimes, even an accumulation of data does not look like a convincing harbinger.

Other times, a single data point is interpreted as precisely that. The September 11, 2001 attack on the World Trade Center was interpreted by many, wrote John Mueller of Ohio State University a year later, as just such a “harbinger.” These observers were suggesting, he continued, that “such extensive destruction to life and property will become common even routine.” Yet, he continued, the attack “may turn out to have been a statistical outlier”—an aberration.

Unusual data, unprecedented data, weird data—all suggest the need to stop and think: Is this an outlier or a harbinger?

Do the random fluctuations of the world—or, perhaps, an alignment of unusual circumstances and unique capabilities—suggest that this is a freakish, one-time outlier? Or are their ways—yet to be identified ways—in which these surprising data are so different from existing trends that they deserve to be investigated as a harbinger?

This is a difficult distinction to make. Just ask John Henry. **B**

Robert D. Behn, a lecturer at Harvard University's John F. Kennedy School of Government, chairs the executive-education program “**Driving Government Performance: Leadership Strategies that Produce Results**.” He is working on a book tentatively titled *The PerformanceStat Potential*.

To be sure you get next month's issue, subscribe yourself at: <http://www.ksg.harvard.edu/TheBehnReport>. It's free!

For the inside secrets about **Driving Government Performance**, go to: <http://hks.harvard.edu/EE/BehnReport>.