What Data Can't Do

By DAVID BROOKS
Published: February 18, 2013

Not long ago, I was at a dinner with the chief executive of a large bank. He had just had to decide whether to pull out of Italy, given the weak economy and the prospect of a future euro crisis.

The C.E.O. had his economists project out a series of downside scenarios and calculate what they would mean for his company. But, in the end, he made his decision on the basis of values.

His bank had been in Italy for decades. He didn't want Italians to think of the company as a fair-weather friend. He didn't want people inside the company thinking they would cut and run when times got hard. He decided to stay in Italy and ride out any potential crisis, even with the short-term costs.

He wasn't oblivious to data in making this decision, but ultimately, he was guided by a different way of thinking. And, of course, he was right to be. Commerce depends on trust. Trust is reciprocity coated by emotion. People and companies that behave well in tough times earn affection and self-respect that is extremely valuable, even if it is hard to capture in data.

I tell this story because it hints at the strengths and limitations of data analysis. The big novelty of this historic moment is that our lives are now mediated through data-collecting computers. In this world, data can be used to make sense of mind-bogglingly complex situations. Data can help compensate for our overconfidence in our own intuitions and can help reduce the extent to which our desires distort our perceptions.

But there are many things big data does poorly. Let's note a few in rapid-fire fashion:

Data struggles with the social. Your brain is pretty bad at math (quick, what's the square root of 437), but it's excellent at social cognition. People are really good at mirroring each other's emotional states, at detecting uncooperative behavior and at assigning value to things through emotion.

Computer-driven data analysis, on the other hand, excels at measuring the quantity of social interactions but not the quality. Network scientists can map

your interactions with the six co-workers you see during 76 percent of your days, but they can't capture your devotion to the childhood friends you see twice a year, let alone Dante's love for Beatrice, whom he met twice.

Therefore, when making decisions about social relationships, it's foolish to swap the amazing machine in your skull for the crude machine on your desk.

Data struggles with context. Human decisions are not discrete events. They are embedded in sequences and contexts. The human brain has evolved to account for this reality. People are really good at telling stories that weave together multiple causes and multiple contexts. Data analysis is pretty bad at narrative and emergent thinking, and it cannot match the explanatory suppleness of even a mediocre novel.

Data creates bigger haystacks. This is a point Nassim Taleb, the author of "Antifragile," has made. As we acquire more data, we have the ability to find many, many more statistically significant correlations. Most of these correlations are spurious and deceive us when we're trying to understand a situation. Falsity grows exponentially the more data we collect. The haystack gets bigger, but the needle we are looking for is still buried deep inside.

One of the features of the era of big data is the number of "significant" findings that don't replicate the expansion, as Nate Silver would say, of noise to signal.

Big data has trouble with big problems. If you are trying to figure out which e-mail produces the most campaign contributions, you can do a randomized control experiment. But let's say you are trying to stimulate an economy in a recession. You don't have an alternate society to use as a control group. For example, we've had huge debates over the best economic stimulus, with mountains of data, and as far as I know not a single major player in this debate has been persuaded by data to switch sides.

Data favors memes over masterpieces. Data analysis can detect when large numbers of people take an instant liking to some cultural product. But many important (and profitable) products are hated initially because they are unfamiliar.

Data obscures values. I recently saw an academic book with the excellent title, " 'Raw Data' Is an Oxymoron." One of the points was that data is never raw; it's always structured according to somebody's predispositions and values. The

end result looks disinterested, but, in reality, there are value choices all the way through, from construction to interpretation.

This is not to argue that big data isn't a great tool. It's just that, like any tool, it's good at some things and not at others. As the Yale professor Edward Tufte has said, "The world is much more interesting than any one discipline."