

## Black Swans

What if Gaussian engineering is clear, simple, and wrong?

**M**Y HOUSE is on a river, and I often see swans drifting by my backyard in their swanlike serenity. Over the years every swan to come by has been white. I have thus reached an inescapable statistical conclusion: All swans are white.

I have watched that river through many storms, and though it has washed high up my back lawn, it has never reached my house. Given all my observed data on water levels, I have reached another conclusion: My house will never be flooded.

A popular book by Nassim Nicholas Taleb, *The Black Swan: The Impact of the Highly Improbable* (Random House, 2007), has me reassessing these conclusions, as well as rethinking much of my experience and education. Taleb assures me that black swans do indeed exist; it's just that I haven't seen them. How much data do I actually have, and what is the probability of an event that I have not yet seen?

Taleb calls an event a "black swan" when it is rare, unexpected, and impactful. He claims that life is dominated by such events—the chance meeting one day of your future spouse, the random acquaintance that forges your career, the serendipitous observation that leads to the next penicillin.

Throughout his book Taleb rails against the traditional way of teaching statistics and probability theory. You would be better off, he says, taking a course on postcolonial African dance. The problem, he says, is that statistics as taught and practiced leads to an unwarranted belief in mathematical certainty, in predictable behavior, and in a world dominated by bell-shaped "normal" Gaussian curves.

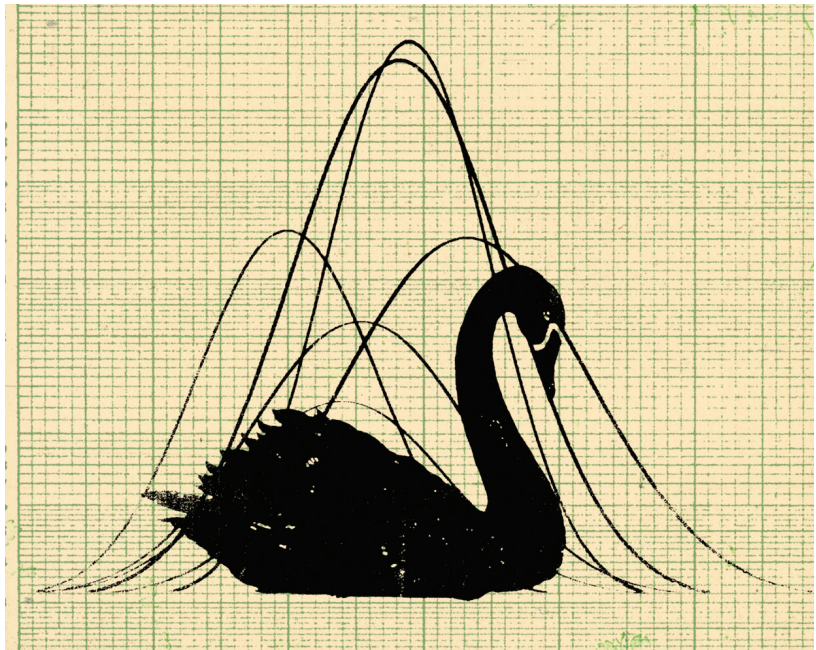
It isn't so, he says. Most of the world isn't Gaussian. Instead it's mostly ruled by the dreaded power law, with its heavy tail that presages regular occurrences of far-outlying events that are almost impossible in a Gaussian regime. Much of the Internet has this character: the popularity of Web sites, the frequency with which a book is sold online, the traffic rates of individual users. Such a distribution curve looks like a downhill ski run: It starts with a precipice of heavy users, swooping sharply downward into a long tail of rare instances that can't be ignored.

It seems to me there is a great deal of truth and wisdom in what Taleb says. We engineers do indeed relish the certainty of

mathematics. We dote on our models; over time we forget the assumptions that went into them. We cling to the central limit theorem and its promise of normality. Flip a random coin a few more times and the percentage of tails gets even closer to 50. Many of our designs depend on this kind of near certainty.

Most of communication theory, for example, is based on the "fiction" of additive white Gaussian noise. Look at received signals over a long enough interval and the multidimensional vector lies ever closer to the surface of a multidimensional sphere. If the spheres around other possible signal vectors don't intersect, then the probability of error can be made arbitrarily small. The math is beautiful and seductive.

Taleb, however, says that elegant mathematics appeals to mechanistic minds that don't want to deal with ambiguity,



and that to make it fit the real world you have to cheat somewhere in your assumptions. Like maybe the world isn't filled with additive white Gaussian noise after all.

I suppose we're guilty as charged, but I'd like to offer a mild defense. We do sort of know that the noise probably won't fit our nice model, but nonetheless all that elegant math does produce designs that are relatively robust against disturbances. Outliers resulting in errors still do occur, of course, but usually their effect is not catastrophic, as it can be in the financial realm, which is Taleb's *bête noire*. So I'm thinking that all of those old courses with bell curves and other Gaussian statistics weren't so bad after all.

I'm watching the river now, and here come the swans. I'm still convinced I will never see a black swan, but I am kind of worried about that flood thing. □