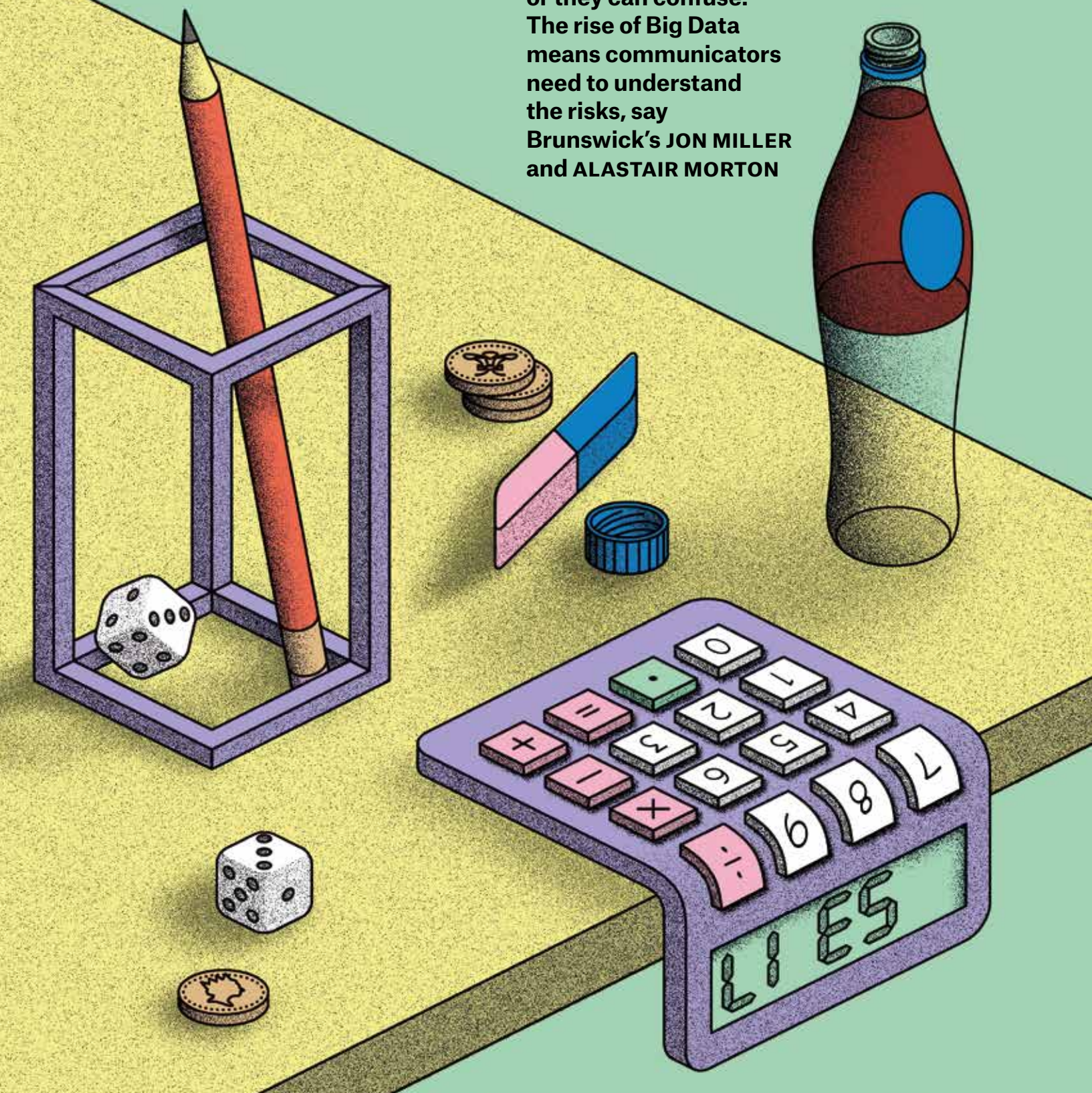
THE USE **AND ABUSE** OF NUMBERS

Numbers can clarify or they can confuse. The rise of Big Data means communicators need to understand the risks, say Brunswick's JON MILLER and ALASTAIR MORTON



A FEW YEARS AGO, University of California researchers spotted a curious phenomenon: companies tend to underperform after a CEO wins a prominent accolade in the business media – from *Forbes*, say, or *Bloomberg Businessweek*. Superstar CEOs, the researchers suggested, were likely to spend more time writing books and sitting on outside boards than their peers, thus neglecting their own businesses. It makes a great story. *The New York Times* went so far as to call the phenomenon "the curse of the business press."

However, a better explanation may be a simple case of **REVERSION TO THE MEAN**. From a high peak in performance, the most likely road is down. Above-average earnings reports tend to be followed by lower results. The so-called "curse" may be, in fact, just mathematics.

Numbers are a valuable tool in professional storytelling. Nothing drives a point home like a big, juicy statistic. But numbers also hold many traps and pitfalls that can create misrepresentations of the truth. Working at speed in a high-pressure environment, it is easy for those involved in corporate communications to let a number tell us what we want it to – a case of **CONFIRMATION BIAS**, interpreting information in a way that confirms our view of the world.

The use of **PERCENTAGES** is another area rife with misuse and abuse. A headline may read, "Profits up 25 percent." A big jump? Perhaps, or it could mean a steady increase in profitability from 4 percent to 5 percent. You could say that profits have increased by 25 percent, or that profitability has risen by a single point; both are correct. Clarity demands more explanation.

Of course, if the intention is to muddle or mislead, percentages are particularly useful. Customers seeing "50 percent extra free" on a bottle of cola will think half of it is free, but of course only a third (33 percent) of the bottle is free. If you hear that a specific risk has risen 100 percent, that may just mean the risk has increased from one in a million to two in a million.

One area of numbers, **PROBABILITY**, notoriously causes humans more grief than others because of the way we are cognitively wired. Imagine a coin that has been tossed and comes up heads several times in a row: most people would imagine it's more likely to come up tails on the next flip. But the coin has no idea how many times it has come up heads already. For each flip, the odds remain the same – 50 percent. This is known as the **GAMBLER'S FALLACY**, and our thinking about probability is riddled with similar misconceptions.

Our tendency to intuitively identify causes is responsible for another error, one that is commonly found in the mainstream media: the **BASE RATE FALLACY**. Consider the news story of a teenage male who has gone on a high-school shooting spree. Subsequent media coverage highlights the fact that he spent the preceding hours playing violent computer games, and soon there are outraged voices calling for the banning of these games. Of course, this ignores the base rate data that shows most teenage males play those same computer games and do not exhibit sociopathic behavior. Our hearts tell us there should be a connection. But hearts can't do math. If such a connection exists, it is likely to be far more complicated than simple cause and effect, and research has yet to prove it.

This tendency to jump to conclusions has been known for centuries, and is behind the oldest adage in data analysis: *cum hoc ergo propter hoc* (literally, "with this therefore because of this"), the principle that **CORRELATION DOES NOT IMPLY CAUSATION**. One satirical writer, for example, was able to show that global warming is being caused by the decline in the number of pirates, charting the statistical correlation as proof. It may sound ridiculous, but this kind of fallacious thinking frequently finds its way into the media: a study in 2012 found that internet users who suffer from depression check email more often and watch more video. That made the headlines because it seemed to show that heavy internet use is bad for your mental health. Without further study, the correlation itself proves no such thing.

Large numbers are another common source of confusion. Communications are peppered with millions and billions and even trillions, but many of us lack an intuitive grasp of **SCALE** – a sense of how big these numbers really are. Imagine counting the seconds as they pass: it would take about 11½ days to get →

SIMPSON'S PARADOX

As much as they can provide clarity, numbers also can easily obscure deeper truths, as noted by mathematician Edward Simpson in a widely cited 1951 paper. Simpson, a colleague of computing pioneer Alan Turing, showed how lumping data together can produce a false conclusion. A clear example of what is now known as **SIMPSON'S PARADOX** occurred in 1973 when the University of California, Berkeley was sued for sexual discrimination based on data showing only 35 percent of female applicants were admitted, compared with 44 percent of men. A closer look revealed women were applying to more competitive programs with lower rates of admission than men. Examined separately, individual departments actually had a small bias *in favor* of women. Simpson's Paradox highlights the danger of relying on numbers to tell the story without context.

ILLUSTRATIONS: ALEX WALKER

to a million – which sounds like a lot of counting until you realize that it would take almost 32 years to get to a billion. A trillion is more mind-bending magnitudes bigger: rewind a trillion seconds into the past, and our ancestors were about to invent the bow and arrow.

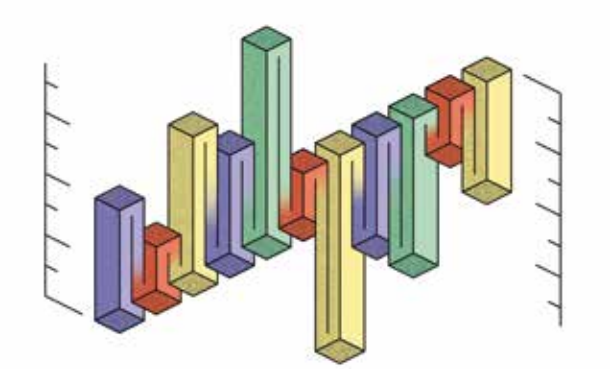
Numbers don't, in fact, speak for themselves. Knowing that this year's margin is 2 percent doesn't tell you anything, unless you know that last year the company made a loss. Customer satisfaction scores of 80 percent may sound high until you learn that competitors score 95 percent. Simply stating a statistic without benchmarks or comparators is an example of *ipse dixit* ("he, himself, said it"), otherwise known as **THE BARE ASSERTION FALLACY**. Behavioral economists show us that whether ordering a bottle of wine, buying a house or negotiating a salary, we need relative numbers to weigh what's on offer.

"Numbers are the masters of the weak, but the slaves of the strong," wrote Charles Babbage, whose steam-age Analytical Engine was the forerunner of the modern computer. "Whenever a man can get hold of numbers, they are invaluable: if correct, they assist in informing his own mind, but they are still more useful in deluding the minds of others." As **BIG DATA** grows more present in our daily lives, it's increasingly important to understand how numbers may be used and abused. Today's children are growing up in a world shaped and configured by data; if they are not data literate, they will be passive consumers, unable to fully engage with society. As early 20th century science fiction writer H.G. Wells foretold: "If we want to have an educated citizenship in a modern technological society, we need to teach them three things: reading, writing, and statistical thinking."

"A good decision is based on knowledge and not on numbers"

Plato, in *Laches*, 380 B.C.

The authors are no strangers to numbers; **JON MILLER** holds a Master's Degree in Artificial Intelligence, while **ALASTAIR MORTON** has a Doctorate in Pure Mathematics. Both are Partners in Brunswick's London office.



DOs AND DON'Ts

Numbers can buoy or sink any argument. This list can help keep your story on the right track

DO

TAKE CARE WHEN USING THE PAST TO PREDICT THE FUTURE

Guard against common errors when talking about probabilities, such as the *gambler's fallacy*

DON'T

JUMP TO CONCLUSIONS

Slow down. Does this number really show what you hope it shows, or is it your own *confirmation bias*?

DO

LOOK AT THE CONTEXT

Without looking at the background data and underlying trends, it's easy to wind up guilty of the *base rate fallacy*

DON'T

CONFUSE CAUSE WITH CORRELATION

Make sure you can explain how two numbers are connected, rather than just presenting data. Let the adage *cum hoc ergo propter hoc* serve as a reminder

DO

USE BOTH RELATIVE AND ABSOLUTE NUMBERS

Avoid the *bare assertion fallacy*, where numbers are baldly stated without any sense of scale. Answer the question, how big is this number?

DON'T

JUST FOCUS ON THE BIG NUMBERS

Chunking up can be misleading. As *Simpson's Paradox* shows, aggregated data distorts the real picture

DO

GET DATA LITERATE

Develop your statistical thinking to make sure that numbers are your servants and not your masters