Deloitte Review

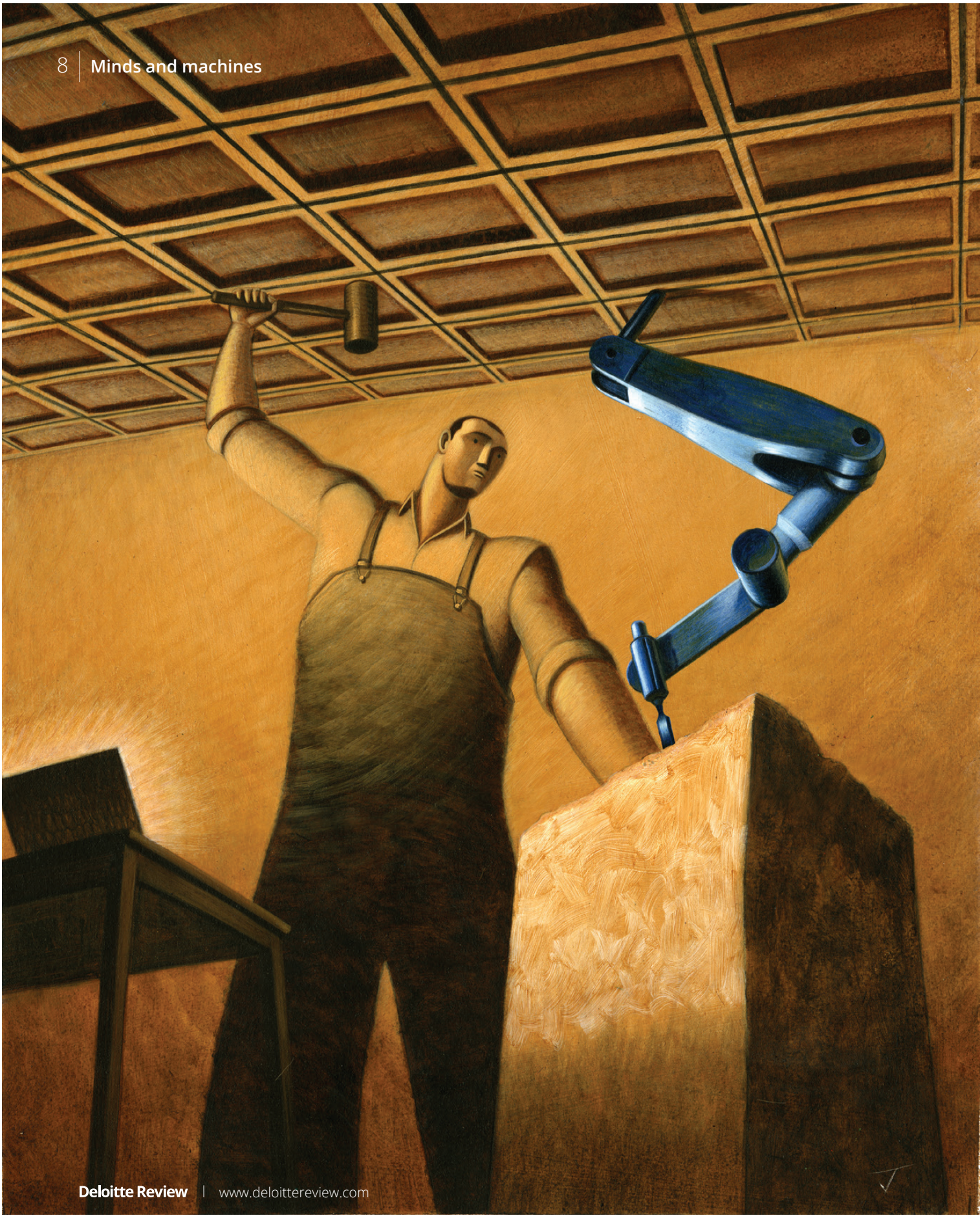
Issue 19 | 2016

Complimentary article reprint

MINDS and MACHINES

The art of forecasting in the age of artificial intelligence

By James Guszcza and Nikhil Maddirala



MINDS and MACHINES

The art of forecasting in the age of artificial intelligence

By James Guszcza and Nikhil Maddirala
Illustration by Jon Krause

HUMAN JUDGMENT IN THE AGE OF SMART MACHINES

TWO of today's major business and intellectual trends offer complementary insights about the challenge of making forecasts in a complex and rapidly changing world. Forty years of behavioral science research into the psychology of probabilistic reasoning have revealed the surprising extent to which people routinely base judgments and forecasts on systematically biased mental heuristics rather than careful assessments of evidence. These findings have fundamental implications for decision making, ranging from the quotidian (scouting baseball players and underwriting insurance contracts) to the strategic (estimating the time, expense, and likely success of a project or business initiative) to the existential (estimating security and terrorism risks).

The bottom line: Unaided judgment is an unreliable guide to action. Consider psychologist Philip Tetlock's celebrated multiyear study concluding that even top journalists, historians, and political experts do little better than random chance at forecasting such political events as revolutions and regime changes.[1]

The second trend is the increasing ubiquity of data-driven decision making and artificial intelligence applications. Once again, an important lesson comes from behavioral science: A body of research dating back to the 1950s has established that even simple predictive models outperform human experts' ability to make predictions and forecasts. This implies that judiciously constructed predictive models can augment human intelligence by helping humans avoid common cognitive traps. Today, predictive models are routinely consulted to hire baseball players (and other types of employees), underwrite bank loans and insurance contracts, triage emergency-room patients, deploy public-sector case workers, identify safety violations, and evaluate movie scripts. The list of "Moneyball for X" case studies continues to grow.

More recently, the emergence of big data and the renaissance of artificial intelligence (AI) have made comparisons of human and computer capabilities considerably more fraught. The availability of web-scale datasets enables engineers and data scientists to train machine learning algorithms capable of translating texts, winning at games of skill, discerning faces in photographs, recognizing words in speech, piloting drones, and driving cars. The economic and societal implications of such developments are massive. A recent World Economic Forum report predicted that the next four years will see more than 5 million jobs lost to AI-fueled automation and robotics.[2]

Let's dwell on that last statement for a moment: What about the art of forecasting itself? Could one imagine computer algorithms replacing the human experts who make such forecasts? Investigating this question will shed light on both the nature of forecasting—a domain involving an interplay of data science and human judgment—and the limits of machine intelligence. There is both bad news (depending on your perspective) and good news to report. The bad news is that algorithmic forecasting has limits that machine learning-based AI methods cannot surpass; human judgment will not be automated away anytime soon. The good news is that the fields of psychology and collective intelligence are offering new methods for improving and de-biasing human judgment. Algorithms can augment human judgment but not replace it altogether; at the same time, training people to be better forecasters and pooling the judgments and fragments of partial information of smartly assembled teams of experts can yield still-better accuracy.

We predict that you won't stop reading here.

WHEN ALGORITHMS OUTPERFORM EXPERTS

WHILE the topic has never been timelier, academic psychology has studied computer algorithms' ability to outperform subjective human judgments since the 1950s. The field known as "clinical vs. statistical prediction" was ushered in by psychologist Paul Meehl, who published a "disturbing little book"[3] (as he later called it) documenting 20 studies that compared the predictions of well-informed human experts with those of simple predictive algorithms. The studies ranged from predicting how well a schizophrenic patient would respond to electroshock therapy to how likely a student was to succeed at college. Meehl's study found that in each of the 20 cases, human experts were outperformed by simple algorithms based on observed data such as past test scores and records of past treatment. Subsequent research has decisively confirmed Meehl's findings: More than 200 studies have compared expert and algorithmic prediction, with statistical algorithms nearly always outperforming unaided human judgment. In the few cases in which algorithms didn't outperform experts, the results were usually a tie.[4] The cognitive scientists Richard Nisbett and Lee Ross are forthright in their assessment: "Human judges are not merely worse than optimal regression equations; they are worse than almost any regression equation."[5]

Subsequent research summarized by Daniel Kahneman in *Thinking, Fast and Slow* helps explain these surprising findings.[6] Kahneman's title alludes to the "dual process" theory of human reasoning, in which distinct cognitive systems underpin human judgment. System 1 ("thinking fast") is automatic and low-effort, tending to favor narratively coherent stories over careful assessments of evidence. System 2 ("thinking slow") is deliberate, effortful, and focused on logically and statistically coherent analysis of evidence. Most of our mental operations are System 1 in nature, and this generally serves us well, since each of us makes hundreds of daily decisions. Relying purely on time- and energy-consuming System 2-style deliberation would produce decision paralysis. But—and this is the non-obvious finding resulting from the work of Kahneman, Amos Tversky, and their followers—System 1 thinking turns out to be terrible at statistics.

The major discovery is that many of the mental rules of thumb ("heuristics") integral to System 1 thinking are systematically biased, and often in surprising ways. We overgeneralize from personal experience, act as if the evidence before us is the only information relevant to the decision at hand, base probability estimates on how easily the relevant scenarios leap to mind, downplay the risks of options to which we are emotionally predisposed, and generally overestimate our abilities and the accuracy of our judgments.[7]

Given that Michael Lewis's book was, in essence, about data-driven hiring decisions, it is perhaps ironic that hiring decisions at most organizations are still commonly influenced by subjective impressions formed in unstructured job interviews, despite well-documented evidence about the limitations of such interviews.

It is difficult to overstate the practical business implications of these findings. Decision making is central to all business, medical, and public-sector operations. The dominance and biased nature of System 1-style decision making accounts for the persistence of inefficient markets (even when the stakes are high) and implies that even imperfect predictive models and other types of data products can lead to material improvements in profitability, safety, and efficiency. A very practical takeaway is that perfect or "big" data is not a prerequisite for highly profitable business analytics initiatives. This logic, famously dramatized in the book and subsequent movie *Moneyball*, applies to virtually any domain in which human experts repeatedly make decisions in stable environments by subjectively weighing evidence that can be quantified and statistically analyzed. Because System 1-style decision making is so poor at statistics, often economically substantial benefits can result from using even limited or imperfect data to de-bias our decisions.[8]

While this logic has half-century-old roots in academic psychology and has been commonplace in the business world since the appearance of *Moneyball*, it is still not universally embraced. For example, given that Michael Lewis's book was, in essence, about data-driven hiring decisions, it is perhaps ironic that hiring decisions at most organizations are still commonly influenced by subjective impressions formed in unstructured job interviews, despite well-documented evidence about the limitations of such interviews.[9]

Though even simple algorithms commonly outperform unaided expert judgment, they do not "take humans out of the loop," for several reasons. First, the domain experts for whom the models are designed (hiring managers, bank loan or insurance underwriters, physicians, fraud investigators, public-sector case workers, and so on) are the best source of information on what factors should be included in predictive models. These data features generally don't spontaneously appear in databases that are used to train predictive algorithms. Rather, data scientists must hard-code them into the data being analyzed, typically at the suggestion of domain experts and end users. Second, expert judgment must be used to decide which historical cases in one's data are suitably representative of the future to be included in one's statistical analysis.[10]

The statistician Rob Hyndman expands on these points, offering four key predictability factors that the underlying phenomenon must satisfy to build a successful forecasting model:[11]

1. We understand and can measure the causal factors.

2. There is a lot of historical data available.

3. The forecasts do not affect the thing we are trying to forecast.

4. The future will somewhat resemble the past in a relevant way.

For example, standard electricity demand or weather forecasting problems satisfy all four criteria, whereas all but the second are violated in the problem of forecasting stock prices. Assessing these four principles in any particular setting requires human judgment and cannot be automated by any known techniques.

Finally, even after the model has been built and deployed, human judgment is typically required to assess the applicability of a model's prediction in any particular case. After all, models are not omniscient—they can do no more than combine the pieces of information presented to them. Consider Meehl's "broken leg" problem, which famously illustrates a crucial implication. Suppose a statistical model predicts that there is a 90 percent probability that Jim (a highly methodical person) will go to the movies tomorrow night. While such models are generally more accurate than human expert judgment, Nikhil knows that Jim broke his leg over the weekend. The model indication, therefore, does not apply, and the theater manager would be best advised to ignore—or at least down-weight—it when deciding whether or not to save Jim a seat. Such issues routinely arise in applied work and are a major reason why models can guide—but typically cannot replace—human experts. Figuratively speaking, the equation should be not "algorithms > experts" but instead, "experts + algorithms > experts."

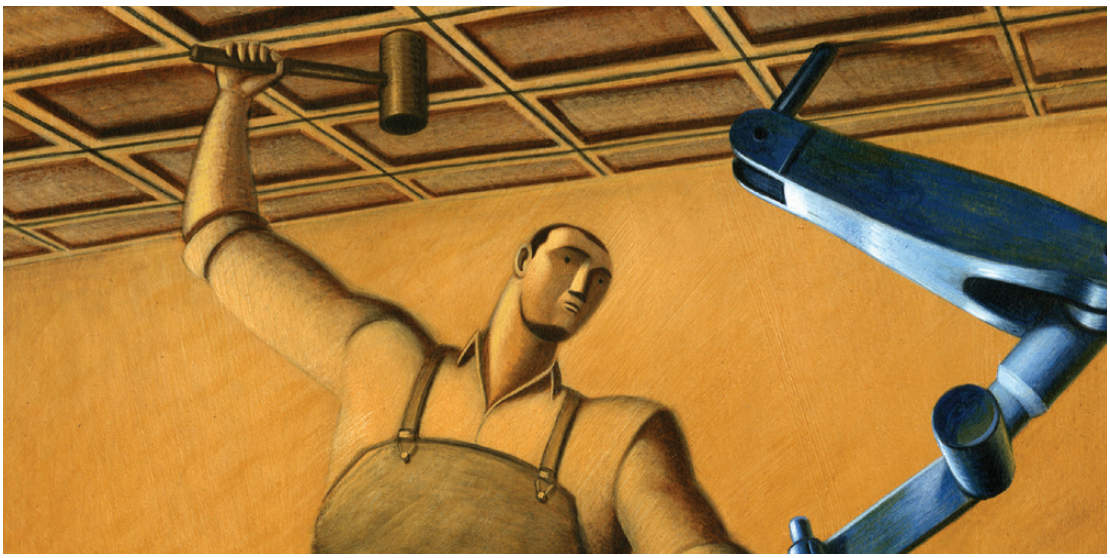
Of course, each of these principles predates the advent of big data and the ongoing renaissance of artificial intelligence. Will they soon become obsolete?

WHAT COMPUTERS STILL CAN'T DO

CONTINUALLY streaming data from Internet of Things sensors, cloud computing, and advances in machine learning techniques are giving rise to a renaissance in artificial intelligence that will likely reshape people's relationship with computers.[12] "Data is the new oil," as the saying goes, and computer scientist Jon Kleinberg reasonably comments that, "The term itself is vague, but it is getting at something that is real. . . . Big Data is a tagline for a process that has the potential to transform everything."[13]

A classic AI application based on big data and machine learning is Google Translate, a tool created not by laboriously encoding

Such issues routinely arise in applied work and are a major reason why models can guide—but typically cannot replace—human experts. Figuratively speaking, the equation should be not "algorithms > experts" but instead, "experts + algorithms > experts."



fundamental principles of language into computer algorithms but, rather, by extracting word associations in innumerable previously translated documents. The algorithm continually improves as the corpus of texts on which it is trained grows. In their influential essay "The unreasonable effectiveness of data," Google researchers Alon Halevy, Peter Norvig, and Fernando Pereira comment:

> [I]nvariably, simple models and a lot of data trump more elaborate models based on less data. . . . Currently, statistical translation models consist mostly of large memorized *phrase tables* that give candidate mappings between specific source- and target-language phrases.[14]

Their comment also pertains to the widely publicized AI breakthroughs in more recent years. Computer scientist Kris Hammond states:

> [T]he core technologies of AI have not changed drastically and today's AI engines are, in most ways, similar to years' past. The techniques of yesteryear fell short, not due to inadequate design, but because the required foundation and environment weren't built yet. In short, the biggest difference between AI then and now is that the necessary computational capacity, raw volumes of data, and processing speed

> are readily available so the technology can really shine.[15]

A common theme is applying pattern recognition techniques to massive databases of user-generated content. Spell-checkers are trained on massive databases of user self-corrections, "deep learning" algorithms capable of identifying faces in photographs are trained on millions of digitally stored photos,[16] and the computer system that beat the *Jeopardy* game show champions Ken Jennings and Brad Rutter incorporated a multitude of information retrieval algorithms applied to a massive body of digitally stored texts. The cognitive scientist Gary Marcus points out that the latter application was feasible because most of the knowledge needed to answer *Jeopardy* questions is electronically stored on, say, Wikipedia pages: "It's largely an exercise in data retrieval, to which Big Data is well-suited."[17]

The variety and rapid pace of these developments have led some to speculate that we are entering an age in which the capabilities of machine intelligence will exceed those of human intelligence.[18] While too large a topic to broach here, it's important to be clear about the nature of the "intelligence" that today's big data/machine learning AI paradigm enables. A standard definition of AI is "machines capable of performing tasks normally performed by humans."[19] Note that this definition applies to more familiar data science applications (such as scoring models capable of automatically underwriting loans or simple insurance contracts) as well as to algorithms capable of translating speech, labeling photographs, and driving cars.

Also salient is the fact that all of the AI technologies invented thus far—or are likely to appear in the foreseeable future—are forms of *narrow* AI. For example, an algorithm designed to translate documents will be unable to label photographs and vice versa, and neither will be able to drive cars. This differs from the original goals of such AI pioneers as Marvin Minsky and Herbert Simon, who wished to create *general* AI: computer systems that reason as humans do. Impressive as they are, today's AI technologies are closer in concept to credit-scoring algorithms than they are to *2001*'s disembodied HAL 9000[20] or the self-aware android Ava in the movie *Ex Machina*.[21] All we currently see are forms of narrow AI.

Returning to the opening question of this essay: What about forecasting? Do big data and AI fundamentally change the rules or threaten to render human judgment obsolete? Unlikely. As it happens, forecasting is at the heart of a story that prompted a major reevaluation of big data in early 2014. Some analysts had extolled Google Flu Trends (GFT) as a prime example of big data's ability to replace traditional forms of scientific methodology and data analysis. The idea was that Google could use digital exhaust from people's flu-related searches to track flu outbreaks in real time; this seemed to support the arguments of pundits such as Chris Anderson, Kenneth Cukier, and Viktor Mayer-

The nature of human collaboration with computers is likely to evolve. Tetlock cites the example of "freestyle chess" as a paradigm example of the type of human-computer collaboration we are likely to see more of in the future.

Schönberger, who had claimed that "correlation is enough" when the available data achieve sufficient volume, and that traditional forms of analysis could be replaced by computeralgorithms seeking correlations in massive databases.[22] However, during the 2013 flu season, GFT's predictions proved wildly inaccurate—roughly 140 percent off—and left analysts questioning their models. The computational social scientist David Lazer and his co-authors published a widely cited analysis of the episode, offering a twofold diagnosis[23] of the algorithm's ultimate failure:

Neglect of algorithm dynamics. Google continually tweaks its search engine to improve search results and user experience. GFT, however, assumed that the relation between search terms and external events was static; in other words, the GFT forecasting model was calibrated on data no longer representative of the model available to make forecasts. In Rob Hyndman's terms, this was a violation of the assumption that the future sufficiently resembles the past.

Big data hubris. Built from correlations between Centers for Disease Control and Prevention (CDC) data and millions of search terms, GFT violated the first and most important of Hyndman's four key predictability factors: understanding the causal factors underlying the data relationships. The result was a plethora of spurious correlations due to random chance (for instance, "seasonal search terms unrelated to the flu but strongly correlated to the CDC data, such as those regarding high school basketball").[24] As Lazer commented, "This should have been a warning that the big data were overfitting the small number of cases."[25] While this is a central concern in all branches of data science, the episode illustrates the seductive—and unreliable—nature of the tacit assumption that the sheer volume of "big" data obviates the need for traditional forms of data analysis.

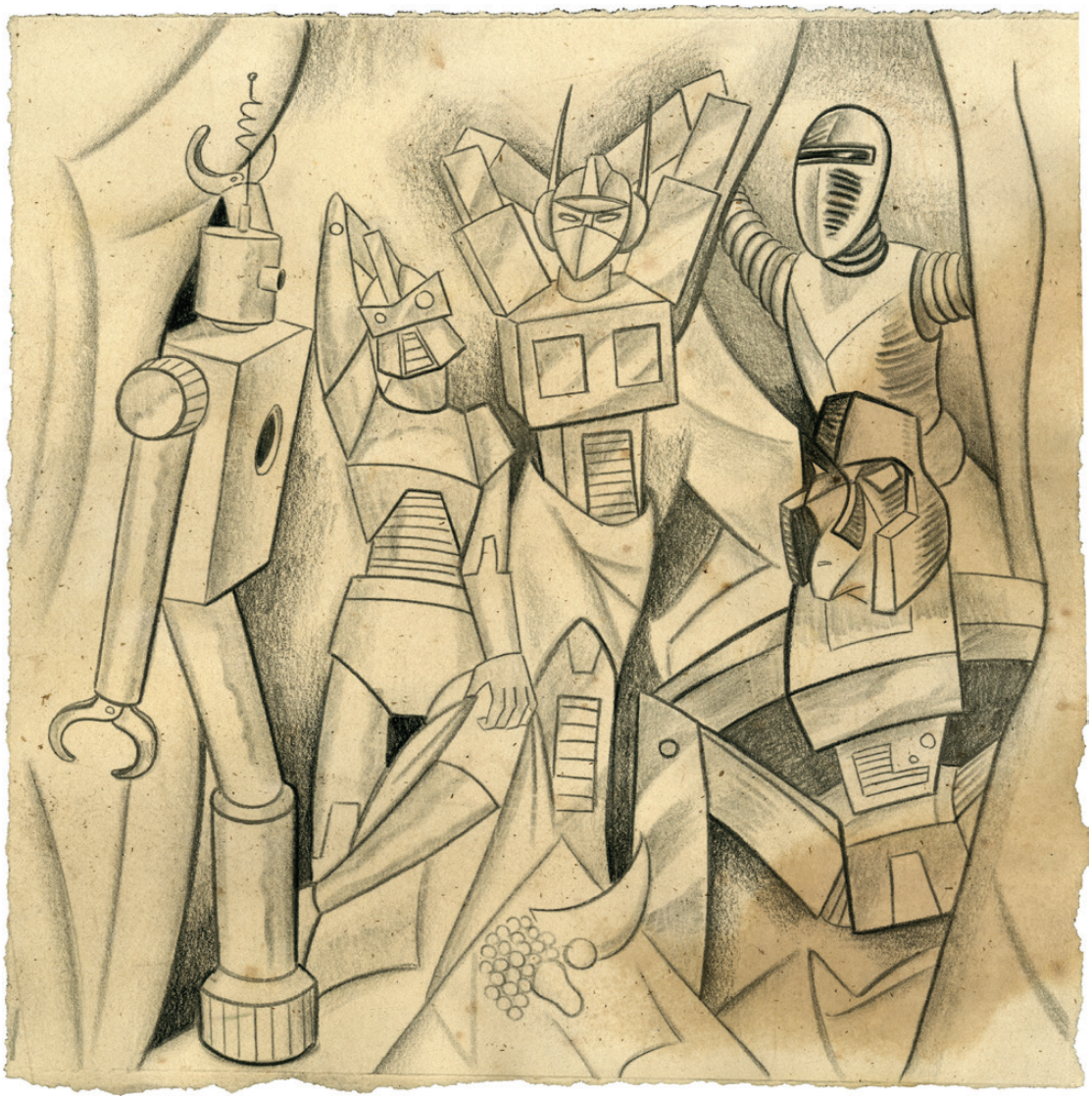
"When Google quietly euthanized the program," GFT quickly went from "the poster child of big data into the poster child of the foibles of big data."[26] The lesson of the Lazer team's analysis is not that social media data is useless for predicting disease outbreaks. (It can be highly useful.) Rather, the lesson is that generally speaking, big data and machine learning algorithms should be regarded as supplements to—not replacements for—human judgment and traditional forms of analysis.

In *Superforecasting: The Art and Science of Prediction*, Philip Tetlock (writing with Dan Gardner) discusses the inability of big data-based AI technologies to replace human

judgment. Tetlock reports a conversation he had with David Ferrucci, who led the engineering team that built the *Jeopardy*-winning Watson computer system. Tetlock contrasted two questions:

1. Which two Russian leaders traded jobs in the last 10 years?
2. Will two top Russian leaders trade jobs in the next 10 years?

Tetlock points out that the former question is a historical fact, electronically recorded in many online documents, which computer algorithms can identify using pattern-recognition techniques. The latter question requires an informed guess about the intentions of Vladimir Putin, the character of Dmitry



Medvedev, and the causal dynamics of Russian politics. Ferrucci expressed doubt that computer algorithms could ever automate this form of judgment in uncertain conditions. As data volumes grow and machine learning methods continue to improve, pattern recognition applications will better mimic human reasoning, but Ferrucci comments that "there's a difference between mimicking and reflecting meaning and originating meaning." That space, Tetlock notes, is reserved for human judgment.[27]

The data is bigger and the statistical methods have evolved, but the overall conclusion would likely not surprise Paul Meehl: It is true that computers can automate certain tasks traditionally performed only by humans. (Credit scores largely eliminating the role of bank loan officer is a half-century-old example.) But more generally, they can only assist—not supplant—the characteristically human ability to make judgments under uncertainty.

That said, the nature of human collaboration with computers is likely to evolve. Tetlock cites the example of "freestyle chess" as a paradigm example of the type of human-computer collaboration we are likely to see more of in the future. A discussion of a 2005 "freestyle" chess tournament by grandmaster Garry Kasparov (whom IBM Deep Blue famously defeated in 1996) nicely illustrates the synergistic possibilities of such collaborations. Kasparov comments:

> The surprise came at the conclusion of the event. The winner was revealed to be not a grandmaster with a state-of-the-art PC but a pair of amateur American chess players using three computers at the same time. Their skill at manipulating and "coaching" their computers to look very deeply into positions effectively counteracted the superior chess understanding of their grandmaster opponents and the greater computational power of other participants. Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process.[28]

MANY MINDS

HUMAN-COMPUTER collaboration is therefore a major avenue for improving our abilities to make forecasts and judgments under uncertainty. Another approach is to refine the process of making judgments itself. This is the subject of the increasingly prominent field of *collective intelligence*. Though the field is only recently emerging as an integrated field of study, notions of collective intelligence date back millennia.[29] For example, Aristotle wrote that when people "all come together . . . they may surpass—collectively and as a body, although not individually—the quality of the few best."[30] In short, groups are capable of pooling disparate bits of information from multiple individuals to arrive at a better judgment or forecast than any of the members of the group. Speaking figuratively,

a "smart" group can be smarter than the smartest person in the group.[31]

A famous early example of collective intelligence involved the inventor of regression analysis, Francis Galton.[32] At a Victorian-era English country fair, Galton encountered a contest involving hundreds of participants who were guessing the weight of an ox. He expected the guesses to be well off the mark, and indeed, they were—even the actual experts in the crowd failed to accurately estimate the weight of 1,198 lbs. But the *average* of the guesses, made by amateurs and professionals alike, was a near-perfect 1,197 lbs.[33]

Prediction markets are another device for combining forecasts. The logic of prediction markets mirrors economist Friedrich Hayek's view that a market mechanism's primary function is not simply to facilitate buying and selling but, rather, to collect and aggregate information from individuals.[34] The Hollywood Stock Exchange, for example, is an online prediction market in which people use simulated money to buy and sell "shares" of actors, directors, films, and film-related options; it predicts each year's Academy Award winners with a 92 percent reported accuracy rate. A more business-focused example is the Information Aggregation Mechanism (IAM), created by a joint Caltech/Hewlett-Packard research team. The goal was to forecast sales by aggregating "small bits and pieces of relevant information [existing] in the opinions and intuition of individuals." After several HP business divisions implemented IAM, the team reported that "the IAM market predictions consistently beat the official HP forecasts."[35] Of course, like financial markets, prediction markets are not infallible. For example, economist Justin Wolfers and two co-authors document a number of biases in Google's prediction market, finding that "optimistic biases are significantly more pronounced on days when Google stock is appreciating" and that predictions are highly correlated among employees "who sit within a few feet of one another."[36]

The Delphi method is a collective intelligence method that attempts to refine the process of group deliberation; it is designed to yield the benefits of combining individually held information while also supporting the type of learning characteristic of smart group deliberation.[37] Developed at the Cold War-era RAND Corp. to forecast military scenarios, the Delphi method is an iterative deliberation process that forces group members to converge on a single point estimate. The first round begins with each group member anonymously submitting her individual forecast. In each subsequent round, members must deliberate and then offer revised forecasts that fall within the interquartile range (25th to 75th percentile) of the previous round's forecasts; this process continues until all the group members converge on a single forecast. Industrial, political, and medical applications have all found value in the method.

In short, tapping into the "wisdom" of well-structured teams can result in improved

judgments and forecasts.[38] What about improving the individual forecasts being combined? The Good Judgment Project (GJP), co-led by Philip Tetlock, suggests that this is a valuable and practical option. The project, launched in 2011, was sponsored by the US intelligence community's Intelligence Advanced Research Projects Activity; the GJP's goal was to improve the accuracy of intelligence forecasts for medium-term contingent events such as, "Will Greece leave the Euro zone in 2016?"[39] Tetlock and his team found that: (a) Certain people demonstrate persistently better-than-average forecasting abilities; (b) such people are characterized by identifiable psychological traits; and (c) education and practice can improve people's forecasting ability. Regarding the last of these points, Tetlock reports that mastering the contents of the short GJP training booklet alone improved individuals' forecasting accuracy by roughly 10 percent.[40]

Each year, the GJP selects the consistently best 2 percent of the forecasters. These individuals—colloquially referred to as "superforecasters"—reportedly perform 30 percent better than intelligence officers with access to actual classified information. Perhaps the most important characteristic of superforecasters is their tendency to approach problems from the "outside view" before proceeding to the "inside view," whereas most novice forecasters tend to proceed in the opposite direction. For example, suppose we wish to forecast the duration of a particular consulting project. The inside view would approach this by reviewing the pending work streams and activities and summing up the total estimated time for each activity. By contrast, the outside view would begin by establishing a reference class of similar past projects and using their average duration as the base scenario; the forecast would then be further refined by comparing the specific features of this project to those of past projects.[41]

Beyond the tendency to form reference-class base rates based on hard data, Tetlock identifies several psychological traits that superforecasters share:

1. They are less likely than most to believe in fate or destiny and more likely to believe in probabilistic and chance events.

2. They are open-minded and willing to change their views in light of new evidence; they do not hold on to dogmatic or idealistic beliefs.

3. They possess above-average (but not necessarily extremely high) general intelligence and fluid intelligence.

4. They are humble about their forecasts and willing to revise them in light of new evidence.

5. While not necessarily highly mathematical, they are comfortable with numbers and the idea of assigning probability estimates to uncertain scenarios.

Although the US intelligence community sponsors the Good Judgment Project, the principles of (1) systematically identifying and training people to make accurate forecasts and (2) bringing together groups of such people to improve collective forecasting accuracy could be applied to such fields as hiring, mergers and acquisitions, strategic forecasting, risk management, and insurance underwriting. Advances in forecasting and collective intelligence methods such as the GJP are a useful reminder that in many situations, valuable information exists not just in data warehouses but also in the partial fragments of knowledge contained in the minds of groups of experts—or even informed laypeople.[42]

MIND THIS

ALTHOUGH predictive models and other AI applications can automate certain routine tasks, it is highly unlikely that human judgment will be outsourced to algorithms any time soon. More realistic is to use both data science and psychological science to de-bias and improve upon human judgments. When data is plentiful and the relevant aspects of the world aren't rapidly changing, it's appropriate to lean on statistical methods. When little or no data is available, collective intelligence and other psychological methods can be used to get the most out of expert judgment. For example, Google—a company founded on big data and AI—uses "wisdom of the crowd" and other statistical methods to improve hiring decisions, wherein the philosophy is to "complement human decision makers, not replace them."[43]

In an increasing number of cases involving web scale data, "smart" AI applications will automate the routine work, leaving human experts with more time to focus on aspects requiring expert judgment and/or such non-cognitive abilities as social perception and empathy. For example, deep learning models might automate certain aspects of medical imaging, which would offer teams of health care professionals more time and resources to focus on ambiguous medical issues, strategic issues surrounding treatment options, and providing empathetic counsel. Analogously, insurance companies might use deep learning models to automatically generate cost-of-repair estimates for damaged cars, providing claims adjusters with more time to focus on complex claims and insightful customer service.

Human judgment will continue to be realigned, augmented, and amplified by methods of psychology and the products of data science and artificial intelligence. But humans will remain "in the loop" for the foreseeable future. At least that's our forecast. **DR**

James Guszcza *is the US chief data scientist for Deloitte Consulting LLP.*

Nikhil Maddirala *is a business analyst with the Monitor Deloitte Strategy & Operations practice at Deloitte Touche Tohmatsu India LLP.*

Endnotes

1. Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton, NJ: Princeton University Press, 2005).

2. See *The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution*, World Economic Forum, January 2016, www3.weforum.org/docs/WEF_Future_of_Jobs.pdf.

3. Paul E. Meehl, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (Minneapolis, MN: University of Minnesota Press, 1954).

4. Philip E. Tetlock and Dan Gardner, *Superforecasting: The Art and Science of Prediction* (New York: Crown, 2015), pp. 344–49.

5. Richard E. Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Upper Saddle River, NJ: Prentice-Hall, 1980), p. 141.

6. Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011). In this book, Kahneman comments that in his younger days, Meehl was one of his heroes.

7. In the vernacular, these phenomena are called "The Law of Small Numbers," "What You See is All There Is" (WYSIATI), "The Availability Heuristic," "The Affect Heuristic," and "Overconfidence Bias." Since the pioneering work of Kahneman and Tversky, dozens of such cognitive biases have been documented and repeatedly replicated.

8. In a profile of Daniel Kahneman, *Moneyball* Author Michael Lewis commented that he was unaware of the behavioral economics implications of his story until he read a review of his book by the behavioral economics pioneers Richard Thaler and Cass Sunstein. See Richard Thaler and Cass Sunstein, "Who's on first," *New Republic*, August 2003, https://newrepublic.com/article/61123/whos-first, and Michael Lewis, "The king of human error," *Vanity Fair*, December 2011, www.vanityfair.com/news/2011/12/michael-lewis-201112.

9. In a recent interview, behavioral economist Richard Thaler commented on this irony, saying, "Take the fact that no one is willing to hire anybody without a job interview. There's lots of evidence that the usual job interviews are almost completely worthless in predicting any aspect of employee performance. . . . I predict that data-driven HR practices will become increasingly important in the coming years. And firms that really figure this out, either on their own or with the help of consulting firms, can have a big competitive advantage." See James Guszcza, "The importance of misbehaving: A conversation with Richard Thaler," *Deloitte Review* 18, January 25, 2016, http://dupress.com/articles/behavioral-economics-richard-thaler-interview/.

10. In essence, these are decisions about which columns (predictive variables) and rows (historical cases) should be used to build one's statistical model. These are expert judgments made by data scientists that, generally speaking, cannot be automated.

11. See Hyndman's lecture, "Exploring the boundaries of predictability: What can we forecast, and when should we give up?" http://robjhyndman.com/seminars/yahoo2015/.

12. For example, in *Superforecasting*, Tetlock comments, that "spectacular advances in information technology suggest we are approaching a historical discontinuity in humanity's relationship with machines."

13. Steve Lohr, "How big data became so big," *New York Times*, August 11, 2012, www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html.

14. Alon Halevy, Peter Norvig, and Fernando Pereira, "The unreasonable effectiveness of data," *IEEE Intelligent Systems*, March/April 2009, pp. 8–12, http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/35179.pdf.

15. Kris Hammond, "Why artificial intelligence is succeeding: Then and now," *Computerworld*, September 14, 2015, www.computerworld.com/article/2982482/emerging-technology/why-artificial-intelligence-is-succeeding-then-and-now.html.

16. The pixels in the digitized photos are the input variables fed into the models, and the algorithmic modeling technology is capable of automatically encoding data features corresponding to, for example, ears and sides of faces. The term "deep learning" denotes a suitably complex form of neural networks (where "deep" denotes extra layers in the network architecture, not metaphorical epistemological "depth"). Consistent with Hammond's comment, neural network algorithms have been under development since the 1980s. Their striking comeback owes to the availability of massive databases on which they can be trained.

17. Gary Marcus, "Steamrolled by big data," *New Yorker*, March 29, 2013, www.newyorker.com/tech/elements/steamrolled-by-big-data.

18. For example, in 2014, the eminent physicist Stephen Hawking famously declared that, "The development of full artificial intelligence could spell the end of the human race. . . . It would take off on its own, and redesign itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded." See Rory Cellan-Jones, "Stephen Hawking warns artificial intelligence could end mankind," *BBC News*, December 2, 2014, www.bbc.com/news/technology-30290540. Oxford philosopher Nick Bostrom's *Superintelligence* is a recent book-length treatment of this theme, often referred to as the "technological singularity." In *The Master Algorithm*, computer scientist Pedro Domingos, while not going so far as Hawking, argues for what he calls the "master algorithm hypothesis": that "All knowledge—past, present, and future—can be derived from data by a single, universal learning algorithm."

19. The AI pioneer John McCarthy introduced the common definition of AI as "getting a computer to do things which, when done by people, are said to involve intelligence." See Brian Harvey, "Artificial intelligence," 1997, www.cs.berkeley.edu/~bh/v3ch6/ai.html.

20. An interesting historical footnote: Marvin Minsky was the scientific adviser to *2001*'s director Stanley Kubrick and its screenwriter, Arthur C. Clarke; both Clarke's original novel and screenplay name-check Minsky.

21. Responding to Stephen Hawking's warning about AI portending "the end of the human race," MIT's Andrew McAfee commented that today's narrow AI technologies "are examples of what can be accomplished with extraordinary amounts of computing power, oceans of data and software that learns from being shown lots of examples. Systems such as these will improve our lives and change our economies, but they are about as likely to rise up against us as forklifts are." See Andrew McAfee, "The march of artificial intelligence is a long way off," *Financial Times*, December 7, 2014, www.ft.com/intl/cms/s/0/cc9ad870-7cb5-11e4-9a86-00144feabdc0.html#axzz479lVmOvB.

22. Chris Anderson, "The end of theory: The data deluge makes the scientific method obsolete," *Wired*, June 2008, www.wired.com/2008/06/pb-theory/ and Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Boston: Eamon Dolan/Houghton Mifflin Harcourt, 2013).

23. David Lazer et al., "The parable of Google flu: Traps in big data analysis," *Science* 343 (March 2014), http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf.

24. Ibid.

25. Ibid.

26. David Lazer and Ryan Kennedy, "What we can learn from the epic failure of Google Flu Trends," *Wired*, October 1, 2015, www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/. Lazer might see the GFT failure as the high-water mark of the early "correlations in big data are enough" form of hype around big data; shortly after *Wired* published Lazer's essay, Tim Harford published an essay headlined, "Big data: Are we making a big mistake?" *Financial Times*, March 28, 2014, www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html. Harford quoted the Cambridge University statistician David Spiegalhalter, commenting that, "There are a lot of small data problems that occur in big data. . . . They don't disappear because you've got lots of the stuff. They get worse." Around the same time, New York University's Gary Marcus and Ernest Davis published an essay headlined "Eight (no, nine!) problems with big data," which concluded that "although big data is very good at detecting correlations . . . it never tells us which correlations are meaningful." *New York Times*, April 7, 2014, www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html. Our own thoughts on big data—for the record, published in January 2013!—appeared in the *Deloitte Review* article "Too big to ignore," January 31, 2013, http://dupress.com/articles/too-big-to-ignore/.

27. Tetlock and Gardner, *Superforecasting*; see the "An Optimistic Skeptic" chapter.

28. Garry Kasparov, "The chess master and the computer," *New York Review of Books*, February 11, 2010, www.nybooks.com/articles/2010/02/11/the-chess-master-and-the-computer/.

29. The recently published *Handbook of Collective Intelligence*, edited by Thomas W. Malone and Michael S. Bernstein (Cambridge, MA: MIT, 2015), is an effort to coalesce this emerging multidisciplinary field. True to their subject, the editors maintained an online draft version to elicit comments from a wide variety of people; this online version can be found at http://cci.mit.edu/CIchapterlinks.html.

30. Aristotle, *Politics*, translated with notes by Ernest Barker (London: Oxford University Press, 1972), p. 123. Also see James Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations* (New York: Doubleday, 2004).

31. The opposite can also be true. As discussed by Cass R. Sunstein and Reid Hastie in *Wiser: Getting Beyond Groupthink to Make Groups Smarter* (Cambridge, MA: Harvard Business Review Press, 2014), poorly structured and managed groups can amplify individual-level cognitive biases. The book also includes a useful survey of research into groupthink and collective intelligence. For a shorter summary of many of the key ideas, see James Guszcza, "From groupthink to collective intelligence: a conversation with Cass Sunstein," *Deloitte Review* 17, July 27, 2015, http://dupress.com/articles/groupthink-collective-intelligence-cass-sunstein-interview/.

32. Galton was also a second cousin of Charles Darwin and, unfortunately, is also remembered as the father of eugenics.

33. This experiment has been recreated numerous times; a recent popular example is the study conducted by National Public Radio for the *Planet Money* episode "How much does this cow weigh?" August 7, 2015, www.npr.org/sections/money/2015/08/07/430372183/episode-644-how-much-does-this-cow-weigh.

34. Hayek, a famous proponent of unfettered markets (as opposed to socialism and central economic planning), argued, "The crucial function of the market process [is] that it enables us to make effective use of information about thousands of facts of which nobody can have full knowledge." Friedrich A. Hayek, "Coping with ignorance," *Imprimis* 7(7), July 1978, pp 1–6, https://imprimis.hillsdale.edu/coping-with-ignorance/.

35. Charles R. Plott and Kay-Yut Chen. "Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem," 2002.

36. Bo Cowgill, Justin Wolfers, and Eric Zitzewitz, "Using prediction markets to track information flows: Evidence from Google," January 6, 2008, www.stat.berkeley.edu/~aldous/157/Papers/GooglePredictionMarketPaper.pdf.

37. For a fuller discussion, see *Wiser* by Sunstein and Hastie.

38. Poorly structured teams, suffering from (for example) domineering leadership or a lack of diversity of opinions expressed, can achieve the opposite of collective intelligence, popularly known as *groupthink*. A major theme of *Wiser* is that such teams can amplify and cascade many of the individual-level cognitive biases identified by Kahneman and Tversky and their followers; this is the modern way of understanding the intuitive concept of groupthink, introduced by Irving Janus in his 1972 book *Victims of Groupthink*. Recent collective intelligence research suggests that conversational turn-taking, the presence of women, and the presence of individuals possessing high degrees of social perception are predictive of "smart teams." Interestingly, these factors are considerably more predictive of group intelligence than the average or maximum IQ of team members.

39. Jason Matheny and Steve Rieber, "Aggregative Contingent Estimation (ACE)," IARPA, www.iarpa.gov/index.php/research-programs/ace, accessed April 29, 2016.

40. Tetlock and Gardner, *Superforecasting*, p. 180.

41. The technique, developed by Daniel Kahneman and Amos Tversky, is known as reference class forecasting and is intended to counter such cognitive biases as overconfidence and base rate neglect. See Kahneman, *Thinking, Fast and Slow*, p. 245. In "From Nobel Prize to project management: Getting risks right," Bent Flyvbjerg discusses a practical example of reference class forecasting involving cost forecasting for large transportation infrastructure projects. *Project Management Journal* 37, no. 3, (August 2006): pp. 5–15, http://arxiv.org/pdf/1302.3642.pdf.

42. Tam Hunt, "How I became a superforecaster," *Slate*, November 18, 2015, www.slate.com/articles/technology/future_tense/2015/11/good_judgment_project_how_i_became_a_superforecaster_for_the_intelligence.html.

43. Chris DeRose, "How Google uses data to build a better worker," *Atlantic*, October 2013, www.theatlantic.com/business/archive/2013/10/how-google-uses-data-to-build-a-better-worker/280347/. Laszlo Bock, *Work Rules! Insights From Inside Google That Will Transform How You Live and Lead* (New York: Twelve, 2015). In the "Don't Trust Your Gut" chapter, Bock discusses Google's use of "wisdom of the crowds" methods to improve its hiring decisions.