A DELICATE BALANCE: Organizational Barriers to Evidence-Based Management

BY JAMES GUSZCZA AND JOHN LUCKER
> ILLUSTRATIONS BY ANTHONY FREDA
A DELICATE BALANCE: 
Organizational Barriers to Evidence-Based Management

BY JAMES GUSZCZA AND JOHN LUCKER > ILLUSTRATIONS BY ANTHONY FREDA

“The most difficult subjects can be explained to the most slow-witted man if he has not formed any idea of them already; but the simplest thing cannot be made clear to the most intelligent man if he is firmly persuaded that he knows already, without a shadow of doubt, what is laid before him.” — Leo Tolstoy

FUTURE PRESENT

Over a hundred years ago, H. G. Wells stated that statistical thinking would one day be as necessary for efficient citizenship as the ability to read and write.¹ Wells’ prescient comment is equally true of management and organizational behavior in the age of big data and business analytics. In domains as varied as professional sports, medicine, consumer business, financial services and government operations, a consensus has rapidly developed about the power of statistical thinking to help experts make better decisions and businesses improve their operations. And a stream of best-selling books, movies and podcasts on the topic has piqued societal awareness of analytics as a catalyst for fresh thinking and change.²
However, in our many years as consultants, we have found that the realized benefits of business analytics are unevenly distributed across domains, and even among different organizations within the same domains. One might chalk this up to the fact that intelligently working with data and doing statistical analysis is hard work, involving specialist skills. Fair enough. But this turns out to be only part of the problem.

Generally speaking, data analysis is only part of an “analytics” project; and ironically it often isn’t the hardest part. It is not uncommon for sophisticated technical work to end up on the cutting room floor—resulting in unrealized value—for reasons having more to do with human and organizational behavior than the finer points of data quality or statistical methodology.

In previous articles, we have discussed the ways in which business analytics is more than a story about arcane statistical algorithms, big data management and information technology. Certainly the newfound prominence of business analytics owes much to Moore’s Law and its corollaries. But business analytics is not ultimately about technology and technique any more than architecture is about blueprints and drafting tools. Well-conceived analytics projects are directed at the central problems and processes in the domain at hand. For example, in medicine this might mean making more reliable diagnoses and triage decisions. In insurance this might mean making better underwriting, pricing or claim settlement decisions. In a human resources context this might mean making better hiring and talent management decisions.

A common challenge in such applications is achieving a realistic compromise between a current-state business or decision process and an envisioned ideal that could in theory be achieved with perfect data and the best available analytics. We have encountered many organizations that, often out of a combination of inertia, competing priorities and culture of skepticism about the effectiveness of business analytics, spend years deliberating before taking the first step toward embracing analytical methods. Others eagerly embrace the notion of analytics but treat it as an “all-or-nothing” proposition requiring data or algorithmic perfection before actions can be taken. Some organizations swing from one extreme to the other.

Of course the preferred point is somewhere between these extremes: in many business settings, analytics is best viewed as an iterative process of continued improvements and data-driven refinements of core business operations. In such settings, either extreme skepticism leading to inaction or extreme aspiration leading to analysis paralysis is suboptimal. Such extreme attitudes and approaches are often borne both of sketchy notions of how analytics works and poor communication between the technical people who analyze data and the decision-makers for whom the fruits of their efforts are intended.
The legendary statistician John Tukey memorably characterized large-scale data analysis as “the collision between statistics and computing.” Similarly, if not properly planned and managed, business analytics projects can feel like a “collision between data analysis and business decision making.” If they are to enjoy the benefits of analytical methods, organizations should strive to avoid collisions and promote evolutions, syntheses and collaborations among people with differing skills and perspectives.

To that end we humbly offer a taxonomy—based on our observations in the field—of what can go wrong. A connecting theme of our observations is that analytics projects are often stymied because of failures to appreciate that both data-driven analytics and expert decision making have strengths as well as limitations and that the strengths and limitations of each must be counterbalanced with those of the other. The image of “data mining” should give way to the image of “data dialogues.”

A MIDDLE PATH

An intriguing aspect of business analytics is its near-universal applicability, yet this also accounts for why it can be such a slippery topic to discuss. Analytics projects take on vastly different aspects in different contexts. For example, the authors have built credit-scoring models using tens of millions of data points as well as analyzed human resources databases containing mere hundreds of data points. Size matters, but it’s not decisive. Similar comments can be made about data quality and completeness, the relative appropriateness of “supervised” versus “unsupervised” learning techniques, the relative appropriateness of experimental versus observational data, appropriate validation methodologies and so on. Such considerations are context-dependent and can vary in veracity and business significance in real-world settings. To bring order to the kaleidoscopic—and ever expanding—variety of applications and methodologies, a classification scheme might be helpful.

One way to classify analytics projects is by the degree to which decision making can be outsourced to computer algorithms. Some of the more prominent examples of business analytics hinge on computer algorithms that serve as decision-making “robots” whose day-to-day functioning involves a minimum of human intervention. Think of Netflix using collaborative filtering algorithms to suggest new titles based on a customer’s viewing history. In each case an algorithmic, data-driven approach both refines and scales up a traditional mode of doing business: Savvy booksellers and video store clerks can be very adept at recommending books and
movies to their loyal customers. But even Quentin Tarantino in his video store heyday could not make movie recommendations on the scale of Netflix’s recommendation engines.

At the other extreme, consider an executive at a global reinsurance company recommending how much capital to set aside in reserves as a cushion for adverse events. Any such manager worth her salt will make the decision in the light of highly sophisticated analyses of past loss trends, correlations among the risks in a portfolio, and stochastic simulations of future economic conditions and other macro factors. But the decision remains solidly with the executive and is unlikely to be left up to the indications of a purely automated algorithm.

The term “business analytics” is broad enough to apply equally well to each of these extreme instances. In each case, data analysis is used to guide a business decision, and the result is decisions that are—on average and in the long run (think the Law of Large Numbers)—better than those that would result from unaided judgment. In the book and movie examples, the machine learning algorithms and induction rules simply replace human decision-makers
(the store clerks); in the reinsurance example, the analytical results serve as inputs to a decision that remains fully under the purview of a human decision-maker.

Our focus in this article is the broad swath of analytics applications falling at various points on the spectrum bounded by these two extremes. It is in this “middle realm” that the success of business analytics can be most surprising and sometimes downright counterintuitive. Medical decision making offers a good paradigm example. Here, highly trained professionals—medical doctors—regularly make decisions under uncertainty. Which of two patients arriving in full crisis at the emergency room complaining of chest pains should be admitted first? Given a positive outcome of an imperfect test, should a patient be treated for a rare disease? Should a risky operation be recommended to a patient? It is hard to imagine such decisions being turned over to a purely algorithmic process similar to the ones used to recommend movies or make targeted marketing decisions. The stakes are too high and the evidence too subtle and complex to turn over to a purely automated decision process.

Yet one could also argue that, precisely because the stakes are so high and the evidence so subtle and complex, the opposite strategy of entrusting medical decision making to the unchecked professional judgment of doctors is similarly suboptimal. In *Blink*, Malcolm Gladwell provides a memorable example that illustrates the point. Gladwell’s anecdote begins in the late 1990s at the resource-strapped Cook County Hospital emergency room. (As it happens this was the very emergency room that inspired the television show *E.R.*) Brendan Reilly, the chairman of the hospital’s Department of Medicine faced 250,000 patients visiting the Emergency Department (ED) each year. An average of 30 arriving patients per day complained of chest pains and worried they were having heart attacks. This presented ED physicians with the formidable problem of rapidly deciding which patients to send to intensive care, which to send to intermediate care, and which to send home. In a controlled experiment, Reilly found that a computer-driven decision-rule protocol was markedly more accurate than the unaided judgment of physicians. In a *JAMA* article summarizing his work, Reilly reported that 84 percent of the physicians he surveyed believed that the decision-rule approach improved patient care. Reilly himself concluded that the analytics-driven rules approach improved efficacy without compromising patient safety.

More generally, there is now considerable evidence that Computerized Decision Support Systems (CDSS) can improve both practitioner performance and patient outcomes. And Atul Gawande has written eloquently about the power of simple checklists—which in other domains would be called business rules—to improve the delivery of medical care. In a *New York Times* op-ed, none other than Billy Beane
joined the chorus of medical, political and business leaders who prescribe a data-driven, evidence-based approach to medical care analogous to the evidence-based methods he famously used to bring the Oakland A’s up in the ranks.¹²,¹³

Obviously none of this work suggests that physicians could or should be replaced with purely automated decision protocols. What it does suggest is that purely “clinical” decision making—one extreme end of our spectrum—is likely a sub-optimal model for much medical decision making. In many situations, physicians make better decisions armed with data-driven predictive models, decision-rule sets and checklists than they do relying on unaided professional judgment.

What can be applied to medical decision making can also be applied to decision problems in domains as diverse as human resources and talent management (Moneyball has become a classic example), risk management, insurance and loan underwriting, fraud detection, caseworker deployment, retail pricing and understanding the organizational drivers of employee resignations. In each of these domains—and many others—evidence-based methods have been shown to outperform the unaided judgment of trained professionals.¹⁴ We believe that, as with medical decision making, the rise of data-driven decision making does not presage the end of professional judgment in any of these fields. There will always be a need for HR managers and talent scouts to make hiring decisions; risk and insurance professionals to make risk management, underwriting and investment decisions; and caseworkers in government, business, and education to make various decisions serving citizens, customers, employees and students.

While the march of business analytics will not replace professional judgment, it can continually transform, enhance and refocus professional judgment. This is a major reason it would be a mistake to view business analytics as a technical domain beginning with data analysis and ending with computer algorithm implementation. Professional judgment enters this process at two crucial points. First, professional judgment and domain knowledge should be used to frame, prioritize and inform specific steps in the process of analyzing data to build predictive models or craft rule sets and checklists. Second, no predictive model or decision rule is complete or infallible: Human judgment is needed to decide when to use, temper or simply ignore model indications. There is no simple recipe for doing this. The process is typically a pragmatic blend of art, science and case-specific business strategy. In short, both analytical methods and the traditional decision processes they are intended to improve have strengths and weaknesses that should be pragmatically counterbalanced.¹⁵
ANGELS AND DEMONS

Scott Fitzgerald famously wrote that “the test of a first-rate intelligence is the ability to hold two opposing ideas in mind at the same time and retain the ability to function.” A similar comment applies to an organization’s ability to execute on analytics. A prerequisite for achieving organizational buy-in of analytics is understanding, forming a strategy around, and communicating the required interplay between analytical methods and the best available domain knowledge and judgment.

The biggest challenges of executing on analytics are often found where algorithmic indications should be integrated with human professional judgment. Because of the range of personnel involved, this is an inherently organizational issue. Unsurprisingly, challenges often arise from such sources as office politics, inertia, principal/agent issues and organizational dynamics. Such generic project implementation issues often take on added force because business analytics may often be poorly or inconsistently understood by the various stakeholders within the organization.

As a result, one often encounters extreme or overly simplistic attitudes about predictive analytics. At one end of the continuum is the sort of extreme skepticism and hostility to analytical methods dramatized in such books as *Moneyball* and *Super Crunchers*. At the other, models are tacitly regarded as repositories of truth rather than provisional, imperfect decision aids that should be continually monitored and subjected to critical thinking. Models are either demons or angels. We believe that such extreme conceptions are at the root of many of the organizational biases that we have observed over the years.

POINT: EQUATIONS TRUMP EXPERTS

“We tell ourselves stories in order to live.” — Joan Didion

Business analytics is typically viewed as a “techy” or “geeky” subject because of its statistics and machine learning subject matter as well as the need for such IT-heavy contributions as data warehousing, systems implementation and dashboard reporting. We tend to regard business analytics in the context of what economists call “human capital.” After all, decision making and decision makers—a.k.a. people—are central to all enterprises, and decades of academic research and business experience suggest that data-driven methods can help even highly trained domain experts make better decisions.

This is not just because our databases are now so deep and rich or that we now possess powerful analytical tools and techniques. It is also because we human
beings are so surprisingly bad at weighing evidence, juggling probabilities and making consistent, coherent decisions in the face of uncertainty. Business analytics is therefore as much about human psychology as it is about data and algorithms.

As an example, take a moment to form a mental image about Linda. Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Now that you’ve formed your mental image, rank these three scenarios in order of likelihood:

- Linda is active in the feminist movement.
- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

The pioneering psychologists Daniel Kahneman and Amos Tversky posed precisely this question to groups of students at several major universities. Kahneman discusses this experiment in *Thinking, Fast and Slow*. Not surprisingly, most of the students felt that being active in the feminist movement was the most likely scenario given what we know of Linda. But at each university, between 85 percent and 90 percent of them also felt that being a bank teller was the least likely of the three scenarios. In other words, they judged that Linda’s being a feminist bank teller is more likely than Linda’s being a bank teller. But a moment’s reflection reveals that this cannot possibly be the case: Feminist bank tellers are a subset of all bank tellers! The probability of being a feminist bank teller must therefore be lower than the probability of being a bank teller.

In this example, our intuitions can lead us badly astray in a way that is as surprising as it is straightforward. Kahneman attributes phenomena such as the Linda story to a certain type of mental process that he called “Type 1.” Type 1 mental processes are fairly automatic, effortless and place a premium on “associative coherence.” In contrast, “Type 2” mental processes are controlled, effortful and place a premium on logical coherence. Although we fancy ourselves primarily Type 2 creatures, many of our mental operations are Type 1 in nature. And—here’s the rub—Type 1 mental processes are very poor at statistical reasoning. This is a major—and, in business, too often neglected—reason why analytical methods are taking root in broad swaths of business, government and medicine. Models can serve as correctives for the bounded rationality and biased cognition of human decision-makers.

Ironically, the dominance of Type 1 thinking can also lead to the organizational resistance to the very analytics initiatives that can help organizations become more
“Type 2” in nature. A major culprit is the so-called “overconfidence bias.” So far are we from being naturally statistical thinkers and rational decision-makers that Kahneman characterizes the mind as a “machine for jumping to conclusions.” He comments that “neither the quantity nor the quality of evidence counts for much in subjective confidence. The confidence that individuals have in their beliefs depends mostly on the quality of the story that they can tell about what they see, even if they see little.” (Italics added.) This is why human experts’ confidence in their own judgments systematically exceeds those judgments’ accuracy. Kahneman calls this phenomenon “the illusion of validity.” It is no wonder that the corrective power of predictive models is so counterintuitive to people making decisions in the field.

This helps explain a phenomenon we have long noticed in our consulting work: Often it is senior leaders and decision-makers who are skeptical about the economic value of predictive models. In light of Kahneman’s observations, this makes sense. After all, such individuals have had the longest time to form an “associatively coherent” body of narratives pertaining to their domains: which draftees will make the best baseball player; which student to admit; which intern to hire; which insurance risks will profit the company; which medical protocols can be cut short. Perhaps their eminence has resulted, in part, from their skill at weaving convincing narratives that impress their colleagues. Their seniority lends them an air of authority, and indeed part of their success might be attributable to their charisma and ability to convince their colleagues with their narrative accounts.21 Unfortunately, given the authority that such individuals enjoy within their organizations, their resistance can seriously hinder the progress of analytics projects.
We have witnessed situations in which a few well-positioned skeptics have wielded disproportionate influence over the fate of predictive modeling projects. Consistent with Kahneman’s discussion, such people tend to disbelieve models and be most confident in the accuracy of their own judgments. In conversation and in meetings, they often emphasize a relatively small number of instances where a model makes counterintuitive predictions, and deemphasize the unproblematic majority of instances. We have seen convincing “anti-modeling” narratives wrapped around memorable cases where a model appears to make a novice error that no competent human expert would ever make. The appropriate response is to analyze such cases with the perspective that (a) models combine the information that they are presented with and (b) no model is perfect, and analyzing anomalies and outliers is a standard way to improve a model. In analytically minded organizations, this is the natural response. But in cultures where anti-model skepticism dominates, such narratives can take on a life of their own.22

Another key finding of behavioral economics is the surprising prevalence of the so-called “availability heuristic:” One’s estimate of an event’s likelihood is affected by how easily it comes to mind. For example, people fear perishing in an airplane accident more than perishing in an auto accident even though the former is actuarially less likely; in academic studies, people have been willing to pay more for terrorism insurance than insurance that covers multiple perils including terrorism; and people tend to estimate that words ending in “ing” are more frequent than words whose penultimate letter is “n.” We have seen examples of apparent model failure lead to conclusions that the model in question is not to be trusted. In these situations, offering statistical evidence of high model accuracy and segmentation power on out-of-sample validation data is only weakly effective against such “cognitively available” stories. The irony is amusing and frustrating in equal measures: The very types of cognitive biases that the model is intended to ameliorate are themselves responsible for institutional “organ rejection” of the model.

Such problems are cultural rather than technical in nature and therefore do not lend themselves to easy answers. Achieving proper communication, unbiased assessments and organizational buy-in are often no less challenging than achieving technical excellence.

COUNTERPOINT: ALL MODELS ARE WRONG

“Any sufficiently advanced technology is indistinguishable from magic.”
— Arthur C. Clarke
We have just discussed an organizational bias that might be called “model accuracy neglect”—the tendency to overestimate the accuracy of one’s own judgments and regard predictive models with undue skepticism. It is also worthwhile to explore a set of organizational biases that tends to the opposite direction: undue deference to analytical techniques and practitioners, and lack of critical thinking in model design and execution.

George Box, one of the world’s preeminent statisticians, is widely known outside the statistical community for his aphorism, “all models are wrong, but some are useful.” It is a sign of the times that one now hears academic statisticians regularly quoted at business conferences and in the popular press. Box’s motto expresses a subtle idea in a mere eight words. But perhaps this idea is too subtle. For Box’s message is often distorted (as in “it’s not too bad to bend the rules”) in ways that lead to this second type of organizational bias.

Two themes are important. First, it is important not to lose sight of the practical context of modeling projects: The goal is not “absolute truth” of the sort sought in fields like mathematics and physics. Rather, it is improved decisions. Second, it is important to have a realistic conception of what models can and cannot do. At the opposite end of the spectrum from “model accuracy neglect” lies another type of organizational bias that might be called “magical thinking about analytics.”

Business analytics practitioners are often motivated by the sheer pleasure of using mathematics and scientific reasoning to arrive at useful facts and insights. The authors remember hearing about a prominent executive of a major insurance company—an actuary by training—who was spotted reading one of Einstein’s original essays on relativity theory while traveling on the corporate jet. The scientific motivation is both admirable and valuable and should be encouraged by organizations wishing to become more analytically oriented. At the same time it is important to remember that the goal of any business analytics project is not “Truth with a capital T” but converting raw data into insights, inferences or predictive models that can lead to better decisions. The goal is not “Truth” but “true enough to be useful.”

This is the essence of Box’s motto, one that becomes clearer in one of its less quotable versions: “Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.” The thought seems transparent to the point of requiring no comment. Yet in practice we see it violated frequently and in a variety of ways. Examples include:

- **The data perfection syndrome**: Organizations often defer analytics projects until such time as an elaborate analytics data warehouse has been constructed. One often hears comments like “first we need to get our data house in
order.” Fair enough, but in many situations this can amount to leaving on the table millions of dollars of savings that could be realized from imperfect and provisional—yet practically effective—models built with imperfect data. A common sentiment is that one’s data needs to be in excellent shape in order to begin analysis. This is typically a mistake. Just as “all models are wrong,” one could also say that “all databases are incomplete.” We have found that, more often than not, something useful can be gleaned even from highly imperfect data. Indeed, analyzing provisional or imperfect data can help focus the organizations’ thinking about what new data elements to collect or how to improve the collection of existing data elements. Furthermore, incomplete data can often be augmented by publicly available or third-party data sources.

• The super-model syndrome: An analogous organizational bias is failing to distinguish between a good-enough, “satisficing” model and a theoretically

Analytics experts are humans, too. And just like the decision-makers whom models are intended to help, analytics experts can be overly confident both in their abilities and in the accuracy of their judgments. This can be exacerbated by the fact that analytics experts possess uncommon skills that many consider advanced or esoteric.
ideal model. As with holding out for perfect data, significant benefits are often sacrificed by engaging in a snark hunt for model perfection or failing to account for the opportunity cost of striving for greater degrees of accuracy. We believe that this organizational tendency results at least in part from a “magical” view of models as repositories of truth rather than inherently imperfect but (in varying degrees) useful decision tools.

- **Outsourced critical thinking:** A related organizational bias is a naïve belief that “the answers are all in the data” or “the quants have figured this out for us.” These are perhaps not bad guiding principles in data-rich, low-risk situations such as recommending books and movies. But in cases where the data are messy, incomplete, ambiguous and/or of limited quantity, considerable institutional knowledge, domain expertise and common sense is needed to effectively make sense of it. Popular phrases such as “data mining” might be partly to blame here. Mining for nuggets of gold is a helpful metaphorical image for a certain kind of algorithmic-powered knowledge discovery. But real-world data analysis more often resembles a dialogue between indications from data and the active hypothesis formation and critical thinking of the data analyst. Furthermore, there is no way to guarantee that the people within an organization best equipped to analyze the data (analysts) are also in the best position to interpret the results. We have been privy to a number of predictive modeling projects that ended badly because the business people outsourced necessary critical thinking entirely to analytics personnel who, while skilled, did not have the appropriate perspective to properly design the analysis and interpret the results. In more than one case, we have witnessed the results of analysts who actually built models to predict the wrong quantity—a decision that should have been discussed and signed off on near the beginning of the project!

- **Over-confident analysts:** Analytics experts are humans, too. And just like the decision-makers whom models are intended to help, analytics experts can be overly confident both in their abilities and in the accuracy of their judgments. This can be exacerbated by the fact that analytics experts possess uncommon skills that many consider advanced or esoteric. However, specialist quantitative skills are not the same thing as critical thinking ability. To take one example, we encounter arguments from authority with unfortunate regularity. This has manifested itself (for example) in analysts stonewalling or rejecting useful methods that do not conform to textbook assumptions; electing to predict an easy (such as a binary) quantity that conforms to textbook assumptions rather than attempting to predict a more complex (for example highly skewed) quantity that would yield more powerful results; or mistaking statistical
significance for business significance. More generally, we have nearly all felt the ramifications of an outright modeling blunder. As was widely reported in the wake of the 2008 market downturn, at least one rating agency used a model of home price changes that could not accept negative numbers. In such cases the damage can be mitigated or avoided by injecting critical thinking, checks and balances, and communication among people with a variety of perspectives into the process. Analytics should be viewed neither as an “ivory tower” nor a “back room” exercise.

- **Glamorous models:** Here another George Box quote is apropos: “Statisticians, like artists, have the bad habit of falling in love with their models.” A common manifestation of this tendency is continuing to refine an analysis or model past the point of diminishing returns. A less obvious manifestation is failing to appreciate—or failing to communicate—a model’s limitations, assumptions or inherent risks. Once again, a dramatic example came to light after the market downturn. The statistician and Wall Street quant David X. Li, at one time called “the world’s most influential actuar,” became famous for a model that greatly simplified the complex relationships among the various securities underlying collateralized debt obligations. Li’s model seemingly offered its users the ability to price complex securities that had been considered too difficult to price. Unfortunately, the model was too simple to support its widespread use. Box’s aphorism notwithstanding, it was not Li himself who fell in love with his model; it was the larger derivatives pricing world. Well before the 2008 crash, Li both articulated the limitations of his model and nicely captured a type of organizational bias in the adoption of models: “The most dangerous part … is when people believe everything coming out of [the model].”

Li’s comment speaks to the dangers of “magical thinking” about analytics and models: the notion that models are repositories of truth rather than inherently provisional and imperfect—but useful—tools for guiding actions.

In an interview, economist and *Financial Times* columnist John Kay provided a clear statement of this position, one that is perhaps less open to misinterpretation than Box’s concise motto. Kay was asked why investment models, built by people with quantitative PhDs from elite universities, appeared to fail. Kay replied:

“Put simply, people made the mistake of believing the model. The people who built them—the mathematics PhDs—didn’t know very much about the world. The people who knew about the world didn’t understand the mathematics. Both groups had inappropriate confidence in the value of these models. They aren’t useless—but models can only illuminate the world, never be a substitute for judgment.”
ACHIEVING BALANCE

Organizations wishing to be first-rate analytical competitors should therefore cultivate the ability to function without losing sight of two opposing ideas about business analytics. On the one hand, in domain after domain, many models have been shown to be effective in helping human specialists make decisions more consistently, accurately and economically. Models are useful. On the other hand, models in these domains often tend to be not repositories of “truth” but rather inherently provisional decision tools that benefit from continual improvement. The goal is therefore not so much to choose between specialists and equations but rather to set up a virtuous cycle whereby one continually works to improve the functioning of the other.

While there is no simple recipe for achieving this, promoting dialogue between groups with different perspectives and skills is a good way to begin. Modelers can do more effective work when they are in continuous dialog with the decision-makers for whom their work is intended. Not incidentally, this also helps reduce the chances of nasty downstream surprises, expensive implementation snags and unmet expectations that manifest themselves only at the close of a project. Conversely, such dialogue can help achieve organizational buy-in of analytics in an organic, incremental way rather than via a collision between data analysis and traditional judgment-driven modes of decision making. In many organizations, promoting such communication is as important an executive function as articulating a strategic vision for analytics in the first place.

Above all such dialogue can help the organization avoid the extremes of skepticism-induced inaction and delay resulting from pursuing unnecessary degrees of perfection. Both extremes are expensive places to reside. DR

James Guszcza is the national predictive analytics lead for Deloitte Consulting LLP’s Advanced Analytics & Modeling practice. He is also an assistant professor of Actuarial Science in the School of Business at the University of Wisconsin–Madison.

John Lucker is a principal and the Advanced Analytics Human Capital market leader in Deloitte Consulting LLP. He is also a U.S. leader in Deloitte Touche Tohmatsu Limited’s Deloitte Analytics group.

Endnotes
1. Wells actually wrote, “The time may not be very remote when it will be understood that for complete initiation as an efficient citizen of one of the new great complex world wide states that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and write.” H. G. Wells, Mankind in the Making (1904). Wells is commonly paraphrased as having written “statistical thinking.”

2. A recent example: Michael Lewis’ book Moneyball, a book which we view as popularizing the concept of “actuarial versus clinical judgment,” has recently been turned into a major Brad Pitt movie. (A landmark academic article in this field is “Clinical versus Actuarial Judgment” by R.M. Dawes, D. Faust, and P.E. Meehl, Science, 31 March 1989 <http://www.sciencemag.org/content/243/4899/1668>). A second example is a journalist at Slate magazine taking an online Stanford University Machine Learning class and blogging about the experience. See “Blogging the Stanford Machine Learning Class” by Chris Wilson, Slate, October 18, 2011. <http://www.slate.com/articles/technology/future_tense/features/2011/learning_machine/stanford_machine_learning_class_week_1_what_what_richard_scary_0.html>

Endnotes

4. Supervised learning involves predicting or explaining a well-defined target variable. Regression analysis is a common example of supervised learning. Unsupervised learning involves finding “interesting” patterns, associations, or groupings in a multidimensional database. Consumer segmentation is an example of unsupervised learning.

5. See for example, “Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model,” by Yehuda Koren: <http://public.research.att.com/~volinsky/netflix/kdd08koren.pdf> Koren is a member of the team that won the Netflix Grand Prize.

6. One of our favorite counterintuitive examples was used to open Ian Ayres’ book Super Crunchers: the Princeton Econo-
ist Orley Aschenfelter has successfully used simple regression models to predict the future value of fine Bordeaux vintages from basic information about growing season temperatures and rainfall. The initial reaction of eminent wine critics was one of dismay and disbelief. This is understandable because one would intuitively think that judging wine quality would be an example where objective, statistical analysis is helpless against the nuanced perceptions of a sophisticated palate. A rich, oaky blend of data, scholarship, and tasting reports is available at Aschenfeter’s website: <http://www. liquidasset.com/>.


15. In his book Super Crunchers, Ian Ayres elegantly captures this thought: “The rise of statistical thinking does not mean the end of intuition or expertise. Rather, [it] underscores how intuition will be reinvented to coexist with statistical thinking. Increasingly, decision makers will switch back and forth between their intuitions and data-based decision making. Their intuitions will guide them to ask new questions of the data that nonintuitive number crunchers would miss. And data-based bases will increasingly allow decision makers to test their intuitions—not just once, but on an ongoing basis . . . while there is now great conflict between dyed-in-the-wool intuitivists and the new breed of number crunchers, the future is likely to show that these tools are complements rather than substitutes. Each form of decision making can pragmatically counterbalance the greatest weaknesses of the other.” (page 195)

16. Daniel Kahneman also discusses this issue in the chapter “The Hostility to Algorithms” in Thinking, Fast and Slow.


18. The University of Minnesota psychologist Paul Meehl was a pioneering figure in the academic study of what has come to be called “actuarial versus clinical prediction.” Towards the end of his career, Meehl commented, “There is no controversy in social science which shows such a large body of quantitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing over 100 investigations, predicting everything from the outcome of football games to the diagnosis of liver disease, and when you can hardly come up with half a dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion.” — “Causes and Effects of My Disturbing Little Book,” Journal of Personality Assessment 50, pp. 370–375. In Thinking, Fast and Slow, (Chapter 21—“Intuitions vs. Formulas”) Daniel Kahneman reports that “Meehl . . . was one my heroes from the time I read his Clinical vs. Statistical Prediction: a Theoretical Analysis and a Review of the Evidence.”
Endnotes


20. In a recent edge.org master class, Kahneman reminisced that his seminal research in cognitive heuristics and biases was in part motivated by his experience teaching a statistics class. He found the material he was teaching very unintuitive and began to wonder whether this was due to a fact of human psychology that humans are not “good intuitive statisticians.” See “The Marvels and Flaws of Intuitive Thinking” by Daniel Kahneman at http://edge.org/conversation/the-marvels-and-flaws-of-intuitive-thinking.

21. Consistent with this hypothesis is the work of the University of Pennsylvania psychologist Philip E. Tetlock. In his book *Expert Political Judgment: How Good Is It?* Tetlock discussed a study of many thousands of predictions made by experts in a variety of fields. Tetlock found that the experts performed little better than random chance, and worse than statistical algorithms. Furthermore, the more prominent experts fared worse than their less celebrated counterparts. It is likely that some of the high-profile experts’ success is due to their overconfidence as well as the narrative appeal of their forecasts, rather than to the accuracy of their predictions. Tetlock wrote that, “there is no reason for supposing that contributors to top journals—distinguished political scientists, area study specialists, economists, and so on—are any better than journalists or attentive readers of the *New York Times* in ‘reading’ emerging situations.” For an informative review of Tetlock’s book, see “Everybody’s an Expert” by Louis Menand in the December 5, 2005 *New Yorker*.

22. This is an example of a phenomenon that Timur Kuran and Cass Sunstein call the availability cascade: a collective belief formation process in which a perception or attitude becomes steadily more plausible as it becomes more prominent in a group’s discourse. Kuran, Timur and Sunstein, Cass R., “Availability Cascades and Risk Regulation,” *Stanford Law Review*, 51 (April 1999): 683–768


24. Ibid.

25. For example, one of us had a recent conversation with an executive at a financial services company who had spent years overseeing the development of a large analytical data warehouse without having a clear idea of what the data would be used for.

26. Although sometimes we wonder. One of us made the mistake of trying Daniel Kahneman’s “Linda” experiment on a group of senior actuarial science majors at The University of Wisconsin. They got it right with hardly a moment’s thought.

27. Deirdre McCloskey wrote about this issue in her book *The Cult of Statistical Significance*.


