

Deloitte Review

ISSUE 14 | 2014

Complimentary article reprint



The *Personalized* and the **Personal**

Socially Responsible
Innovation through Big Data

BY JAMES GUSZCZA, DAVID SCHWEIDEL AND
SHANTANU DUTTA

> ILLUSTRATION BY LIVIA CIVES

Deloitte.

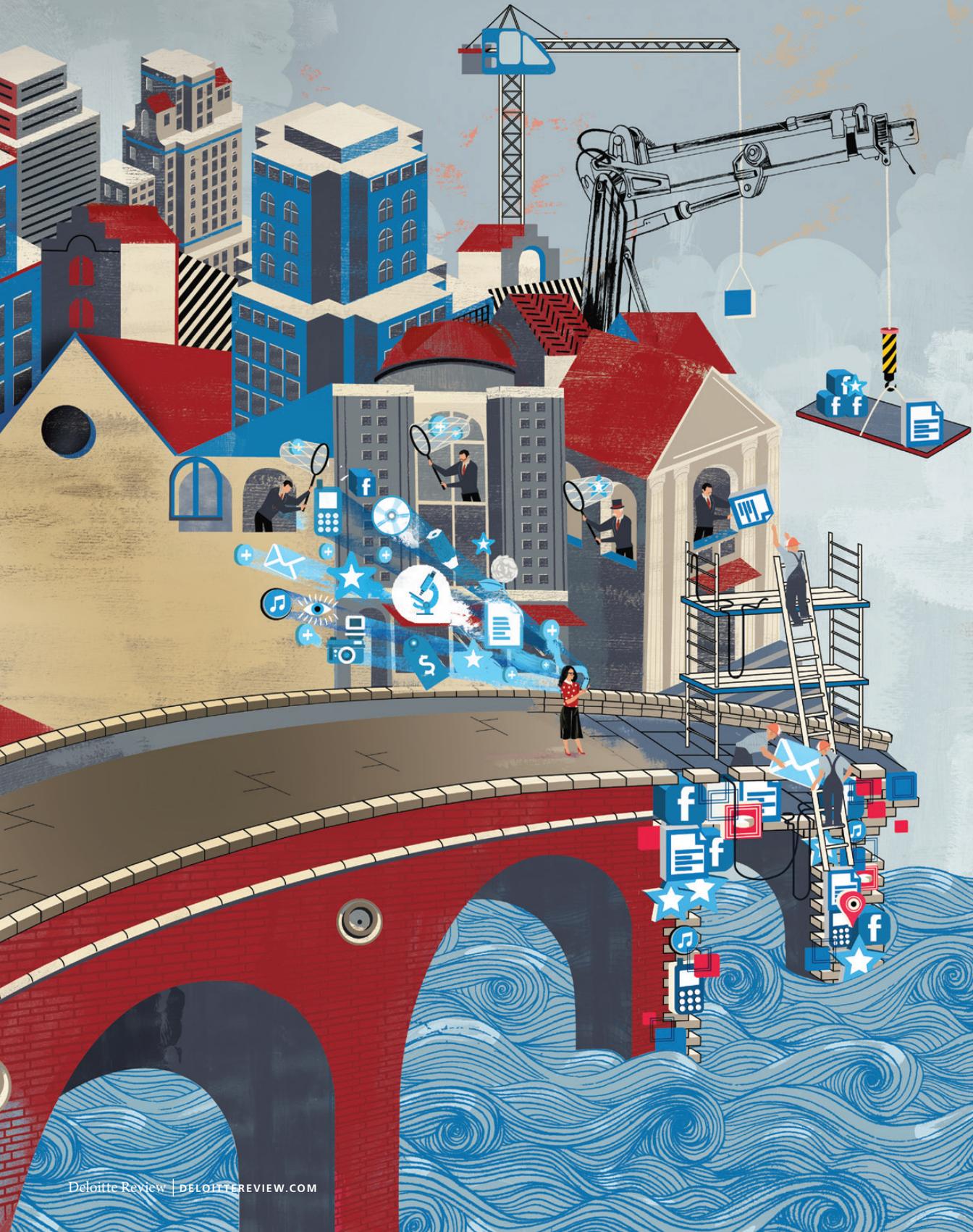
About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms. Please see www.deloitte.com/us/about for a detailed description of the legal structure of Deloitte LLP and its subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

Copyright 2012 Deloitte Development LLC. All rights reserved.
Member of Deloitte Touche Tohmatsu Limited



The *Personalized* and the **Personal**

Socially responsible innovation through big data

BY JAMES GUSZCZA, DAVID SCHWEIDEL, AND SHANTANU DUTTA
> ILLUSTRATION BY LIVIA CIVES

“The best minds of my generation are thinking about how to make people click ads. That sucks.”

—Jeff Hammerbacher, Cloudera founder

YOU DON'T NEED A WEATHERMAN TO KNOW WHICH WAY THE WIND blows. “The web has become an integral part of our lives. A trace of our use of it can reveal very intimate personal things... Whom would you trust to decide when to access it, or even to keep it secure?” So writes Tim Berners-Lee, the inventor of the World Wide Web, in a recent statement sent to the *Financial Times*.¹ While Berners-Lee is a singular figure, his comment is of a piece with of a broader societal unease with the growing ubiquity of data capture and data analytic technologies in many areas of business and society.

This all sounds like bad news about life in the age of electronically captured behavioral data. But there is a positive message to be discussed as well. Beyond focusing on accuracy and privacy, organizations can consider innovative ways of using their stores of data to give back to the people who generated it.



ics, but rather the questionable use of data or analytics models to support (whether intentionally or unintentionally) poor decisions or ethically questionable practices. With characteristic pithiness, mathbabe blogger Cathy O’Neil discusses a genre of what she calls “creepy models.”⁶ In short, there is a growing recognition that, as with all sciences, technologies, and business methods, data science can be used for both socially desirable and undesirable ends.

Societal views surrounding big data technologies, business models, and government policies are therefore set to evolve as kaleidoscopically as have the technologies themselves in recent years. Informed observers anticipate a trough in the enthusiasm surrounding business analytics created by excessive big data boosterism in recent years.⁷ Such boom-and-bust patterns are of course common to many

Other examples are not hard to find. Relaxations of digital privacy policies are routinely followed by criticism both in the blogosphere and in prominent mainstream forums.² Regulators call for restrictions on the collection and use of electronic data with increasing regularity. Dave Eggers’s latest novel *The Circle* concerns life at a fictional mash-up of Silicon Valley behemoths, which is given to promulgating such Orwellian pronouncements as “SECRETS ARE LIES” and “PRIVACY IS THEFT.” Eggers’s book is the first fictional work to be excerpted as a cover feature for the *New York Times Magazine*³ and became a best-seller upon its release.

Nor is such discomfort restricted to journalists, politicians, and novelists. Preeminent network theorist Albert-László Barabási recently penned an opinion piece titled “Scientists must spearhead ethical use of big data.”⁴ Not unrelated is an attitude of “data skepticism” that is increasingly harbored among prominent data scientists.⁵ This is not skepticism about the inherent value of data science or business analyt-

business ideas that capture the collective imagination of the business press. But the emerging attitude of data skepticism pertains to societal aspects of data analytics that are not shared by other technologies that have marched through the various stations of the hype cycle. As we will argue shortly, this owes less to the vaunted bigness (in the sense of raw data volume, variety, velocity) of the data than to a psychological, behavioral, and social content that many find invasive or “creepy.”

This climate will motivate many data-centric organizations to more closely evaluate the ethical and societal implications of their data strategies. But it would be a missed opportunity to view the situation as a tug-of-war between business opportunity and social acceptability. Focusing on the ethical and societal aspects of big data can lead to innovative strategies for achieving sustainable success in the marketplace.

In other words the story of responsible approaches to big data need not simply be one of legal and regulatory compliance; it can be about innovation-fueled profitability and growth as well. But this requires confronting head-on why many people find certain uses of data creepy. Issues of data accuracy and privacy protection are of obvious importance, but more can be said.

The issue is less one of *big* data than *behavioral* data. Ever more detailed information about what we buy, what we think, how we spend our time, and with whom we spend it, is digitally captured and available to infer personal traits that until recently have been hidden from view. Such portraits are often based on data that are not always knowingly divulged and data mashups that are impossible to anticipate. And they can be at odds not only with our public personas but with our personal self-conceptions as well. It is understandable that many find this prospect unsettling.

This all sounds like bad news about life in the age of electronically captured behavioral data. But there is a positive message to be discussed as well. Beyond focusing on accuracy and privacy, organizations can consider innovative ways of using their stores of data to give back to the people who generated it. One promising avenue is, like the above-mentioned promises and perils, rooted in considerations of human behavior.

A considerable amount of contemporary behavioral science research focuses on the difficulty people experience making important choices and following through on their goals. This research can serve as a wellspring of ideas for innovative data products and services that use people’s data in ways that enable them to make better decisions. Data analytics in business need not conform to the “us versus them” story that is currently taking shape. To the contrary, when motivated by and brought to life with ethical thinking as well as insights from the behavioral sciences, data analytics offers a distinctly 21st century approach to doing well by doing good.

A SMALL NOTE ABOUT BIG DATA

A semantic confusion surrounds (some might say smothers) the term big data. This confusion engenders its own type of data skepticism—namely the mistaken belief that it's all marketing hype and little of substance is at stake. There is indeed legitimate criticism to be made of the journalistic and marketing excesses heaped on the term. We believe that once these excesses have been put in their place, much of the confusion and skepticism will dissolve.

Much of the confusion surrounding big data owes to the fact that the term is used in two distinct senses.⁸ Officially, big data denotes data sources whose very size and complexity create problems for standard data management and analysis tools. Streaming data from digital sensors, audio and video recording devices, mobile computing devices, Internet searches, and social networking technologies are all examples. This is the petabyte-class data that calls for such next-generation data management technologies as MapReduce, Hadoop, and NoSQL. It is the sort of data that justifiably excites business leaders and such leading researchers as Albert-László Barabási and Sandy Pentland.

However most references to big data pertain simply to whatever data is useful in enabling business analytics endeavors. And most such endeavors do not (at least at first) call for data that is “big” in the “3V” (volume, variety, velocity) sense. The book *Moneyball* provides the classic (and now, with the Brad Pitt movie, photogenic) example: Billy Beane’s innovative data and analysis-driven strategy for scouting baseball players upended a major industry by exploiting an inefficient market for talent. The strategy did not require the use of the cloud, petabyte-class data, Hadoop clusters, or anything of the sort. Yet in the press, such stories are often described in big data terms.⁹

Our point is certainly not to argue that true 3V big data is unimportant, or that the colloquial use of big data should go away. Only that the term is typically best understood along similar lines as terms like “rocket scientist”: Unless otherwise specified, it’s probably safest to assume that the term is being used loosely or metaphorically. Most of the time big data is used not in the sense of “the very size of the data is a problem” but in the sense of “data that is too rich or complex to analyze well in a spreadsheet and without concepts from university-level statistics.” We believe this simple clarification would go some way toward dampening the disruptive boom-and-bust effect that that hype cycle threatens for business analytics.

A more fundamental point is that organizations’ focus should be not on “big” data per se, but on identifying and using the right data.¹⁰ And even before the collection of data begins, a strategic objective should be clearly articulated and leadership must be in place to ensure that the resulting findings will be acted upon and used to make improved decisions. Oftentimes the answers are easy. Asking the right questions, and having the organizational will to act upon them, is the hard part. These commonsense, yet crucial, observations are obscured by the tendency to focus on the technological, rather than the strategic and scientific, aspects of business analytics.

LET'S GIVE CREDIT SOME THEORY

An early example of behavioral “digital traces” revealing hidden personal traits involves precisely the type of data that kicked off the business analytics revolution in the early 1960s: personal credit data. Decades before Billy Beane transformed the baseball scouting profession through the use of analytics, credit scoring transformed the somewhat less glamorous job of the bank loan officer. Rather than employ experts to evaluate loan applications on a case-by-case basis, it turns out to be dramatically more accurate and economical for lenders to employ credit scoring algorithms that combine many dozens of weakly predictive data elements into predictive models that are stronger than the sum of their parts. Today most personal loans are underwritten, and their interest rates largely determined, using credit scores.

Over thirty years later, the personal insurance industry made a striking discovery: Personal credit data is also predictive of which policyholders are more or less likely to experience automobile accidents and claim-worthy events on their homeowners policies. And the effect isn't even subtle. Not only is personal credit score predictive of personal auto insurance claim experience, actuaries commonly view it as one of the top predictors.¹¹

Many (including one of us, whose first predictive modeling job was to build a credit scoring model for an insurance company) at first find this relationship surprising, and perhaps a bit unsettling. Why should bill-paying behavior be relevant to car accident propensity? After all, there is no causal relationship connecting the former with the latter.

Or is there? In fact there is evidence supporting what many analysts intuitively infer from the unambiguous patterns they encounter in the data. Namely, a more subtle, indirect causal relationship is likely at play. Upon researching the issue, University of Texas-Austin professors Patrick Brockett and Linda Golden stated that:

... Individualized biological and psycho-behavioral correlates provide a connection between credit scores and automobile insurance losses. Credit scores, like good student discounts and marital status, tap a dimension of responsibility and stability for the individual that can permeate multiple areas of behavior.¹²

In practical terms, the Brockett-Golden conclusion is that the credit score encapsulates personal finance behavior that serves as an observable proxy for something that is unobservable but strongly predictive: a set of behavioral traits that lead one to have a higher or lower chance of experiencing an auto accident or committing a violation than otherwise similar people. One type of behavior—borrowing and repaying—is likely an outward manifestation of a set of underlying behavioral

traits that affect another type of behavior: defensive or aggressive driving.

More metaphorically, credit provides an unexpected window into the soul, one that has unambiguous business implications. As soon as one company pioneered the use of credit scores to select and price auto insurance risks, its competitors had little choice but to quickly follow suit to avoid the very real danger of adverse selection.

KNOWING ME, KNOWING YOU

The credit scoring story is but one among many. A recent example comes from the world of US electoral politics. The 2012 Obama reelection campaign built and analyzed a comprehensive database combining demographic information as well as past voting and lifestyle behaviors. Doing so enabled the campaign to characterize individual voters in terms of their likely “persuadeability” in response to specific types of messages, delivered by specific types of campaign workers, or over specific media.¹³ Interestingly, the campaign paired this analysis of behavioral data with the insights of a team of behavioral scientists. The team, for example, formulated behavioral science-informed strategies for countering false rumors and prompting voters to follow through on their commitments to vote.¹⁴

A more surprising example comes from the fledgling domain of neuroeconomics. The ratio of the lengths of financial traders’ ring to index fingers (“2D:4D”) turns out to be predictive of the number of years they stay in the business, as well as their long-term profitability.¹⁵ This might strike one as one of the innumerable scientific publications that makes the headlines but fails to replicate.¹⁶ But it turns out that 2D:4D, like credit score, is a proxy for a highly predictive but hard to observe trait: Low 2D:4D is a rough proxy for high exposure to testosterone while in the uterus. The 2D:4D ratio has also been correlated with the risks of heart disease and diabetes, alcoholism, school exam scores, and measures of business leadership and innovation. Echoing a point made above, the 2D:4D measurement hardly constitutes big data. But it is highly predictive of a range of behaviors.

The 2D:4D ratio is a somewhat unusual example of a physical proxy for various latent psychological traits that cannot be directly observed. As the credit scoring discussion suggests, the Internet age presents us with the ability to construct *digital* proxies for private or otherwise unobservable personal traits. Striking confirmation of this is provided by a recent study of social networking “likes” (voluntary indications offered by online social network site users of their positive associations with various pieces of online content). Research conducted at the University of Cambridge Psychometrics Centre focused on the “likes” of a sample of 58,000 users. They found that they were able to predict ethnic origin (Caucasian or African

American) with 95 percent accuracy; male sexual orientation with 88 percent accuracy; political leanings (Democrat or Republican) with 85 percent accuracy; religion (Christian or Muslim) with 82 percent accuracy, and so on.¹⁷ The researchers also found weaker but still significant correlations between social networks and such latent psychological traits as intelligence, extraversion, and emotional stability. The research employed standard analytic techniques (Principal Components Analysis, regression modeling, cross-validation) common to much business analytics work.

While this particular study focuses on only one type of digital record (likes), it underlines a broader point about the predictive power of behavioral data. Considered in isolation, most individual digital traces (such as watching a movie, buying a bottle of wine, or expressing approval of a brand) convey fairly little. But when used to create composite measures—much as credit scores are synthesized from multitudes of financial transactions—they collectively provide strong indications of personal attributes and psychological traits that might be considered private.

Predicting behavior is of central importance in any number of domains ranging over marketing, insurance, health care, government, and education. The implication of these examples is that predictive behavioral patterns can be found even in unlikely and unexpected places. And as more of our daily activities become digitally mediated, the opportunities for data-driven behavioral predictions increase considerably.

...many of the digital traces we routinely leave behind as we go about our quotidian affairs will likely prove useful in ways that are hard to imagine, and perhaps hard to accept, today.



O BRAVE NEW WORLD THAT HAS SUCH ALGORITHMS IN IT

The credit example is an early, instructive instance of a process that has been taking place in business and society for decades and is taking on ever-greater velocity in the age of big data and business analytics. The story can be told something like this: Once upon a time everyday purchasing, borrowing, and payment activities went largely unrecorded, reflected only in highly aggregated accounting statistics. As information technology became widely available, such transactions began to be captured electronically—people began to leave behind digital traces of their personal finance transactions. At first this was for ease of bookkeeping, but businesspeople eventually realized the extent to which past behavior is predictive of future behavior. For decades, these predictions were made in essentially the same domain—consumer finance—in which the data was generated. Much later, it was discovered that the data was surprisingly effective even well outside its domain of origin (personal insurance). And today, there is a gradual awareness of the biological, psychological, and behavioral traits underpinning the stable predictive relationships that we have stumbled upon.

In hindsight, this story was a bellwether for a process that now repeats itself ever more rapidly. At first it was only our personal finance behaviors that were digitally captured. Today, we leave behind “digital traces” each time we call or email a friend, read an online news article, watch a program or movie on cable TV or a streaming service, make a social or professional connection, shop for a product, or express an opinion or preference via social media. Nor does the list end there. Many drivers now have GPS devices installed in their cars that stream data about their driving behavior to their insurers. Analogously, the “quantified self” movement has led to the retail availability of inexpensive self-tracking devices that, connected via Bluetooth to our smartphones, capture fine-grained details about our sleep, diet, and exercise behavior. Ever more aspects of our daily affairs and routine behaviors are digitally mediated.¹⁸

The business analytics implications are both obvious and enormous. Credit scores predict insurance risk behavior because in a sense they “know us,” perhaps in some sense better than we know ourselves.¹⁹ Because of this, there is in principle no limit to their range of applicability. For example, landlords commonly use credit scores as a “window into the soul” of prospective tenants as a way of selecting tenants who will pay the rent and look after their apartments. Similarly, one of us had the opportunity to study the relationship between credit score and the likelihood that a noncustodial parent will fall behind on his or her child support payments. As one might expect, the relationship was strong.

Similarly, many of the digital traces we routinely leave behind as we go about our quotidian affairs will likely prove useful in ways that are hard to imagine, and perhaps hard to accept, today. To illustrate, let's arbitrarily recombine a few of the examples from the above section. Will educators find that poor—or virtuous—study habits are contagious in digitally captured social networks? Or predicted by streaming online game-playing data? Conversely, might online gaming companies be able to use digitally captured information from Massively Open Online Courses (MOOCs) to target specific games to specific individuals? Will our social networks increasingly determine the range of opinions presented to us via digital news media? Will health care providers and insurers find that retail scanner data is predictive of health expenditures or various disease risks? Might auto insurers find that 2D:4D is an effective predictor of aggressive driving behavior? Or that texting while driving is, like smoking and obesity, contagious in social networks? Will governments find it desirable to use all manner of social network, purchasing, web browsing, and attitudinal digital traces to predict ahead of time which citizens are more likely than others to commit tax fraud or various types of white-collar or violent crimes?

The last scenario was dramatized in the movie *Minority Report*, based on the Philip K. Dick story of the same name. We now know that the psychic powers of Dick's fictional "pre-cogs" are unnecessary. For the purpose of making predictions about individual-level behaviors, the accumulation and statistical analysis of a sufficient variety of digital traces often suffices.

The reader can form his or her own judgments about the relative plausibility of these particular speculations. While we happen to find each of them scientifically plausible, their social desirability is another matter entirely. In a recent interview, the pioneering computational social scientist Sandy Pentland puts the matter starkly:

[An] important issue with Big Data is that since this data is mostly about people, there are enormous issues about privacy, data ownership, and data control. You can imagine using Big Data to make a world that is incredibly invasive, incredibly 'Big Brother'... George Orwell was not nearly creative enough when he wrote 1984.²⁰

Pentland's comment telegraphs the sorts of issues that increasingly occupy business leaders pondering their current and future uses of behavioral, or otherwise personal, information. As the comment suggests, the issue goes beyond questions of "what can we do" to "what should we do?" That is to say, the issue is one of corporate ethics.

'TIS NEW TO THEE

Different people will report different levels of queasiness at various private- or public-sector uses of behavioral, social network, and personal data. This is evidenced by generational differences in the way seemingly private information is shared on social networking sites; political and philosophical differences that lead different people to support different levels of privacy regarding data that can be used for public safety; and societal differences that manifest themselves in varying data privacy laws across countries. Furthermore, there is no reason to assume that what an individual or a population finds acceptable is stable over time. People's attitudes can be expected to evolve along with the evolution of technology and innovations in the use of data.²¹

There can be no easy answers to questions such as this, nor will we attempt any. We will be content to make only one modest suggestion, which is that organizations adopt a certain focus as they strategize future strategic uses of data. Namely: consider whether the customers, citizens, patients, or students whose data is being used and who are being targeted perceive the value they receive in return to be at least as great as the costs in terms of both personal privacy loss and general societal change.

The idea might initially seem Pollyannaish but in fact carries practical benefits. Most obviously, individuals and regulators will tend not to resist innovations that they perceive as offering favorable cost-benefit trade-offs. For example, many people are willing to use data-capturing grocery store club cards in return for discounted groceries and personalized offers. Similarly, many are reasonably sanguine about using free email services in full knowledge that their e-mails are being text-mined to support targeted ad placement. These are both prominent big data-driven business analytics strategies that, until now, have caused fairly little regulatory or public relations backlash. One can infer that most supermarket shoppers and users of free e-mail services ultimately perceive that the benefits outweigh the costs. So long as this balance remains stable the grocers and e-mail providers can pursue their business models without the burden of undue media or regulatory scrutiny.

Avoiding burdensome regulation or scrutiny, while valuable, is a "negative" benefit in the sense of avoiding something bad. But the customer-centric focus can also be the source of positive benefits in the form of innovative products and business models. A few examples will illustrate the concept.

Insurance: Auto insurers are now considering the use of continuously streaming driving behavior, captured via GPS devices, as inputs into a core analytical function: better actuarial pricing and segmentation of their populations of drivers. But the level of detail now capable of being captured is such that insurers can additionally contemplate providing additional sources of value to policyholders in the form

of “data products.” For example, young drivers and their families could be offered subscriptions of personalized progress reports about their driving behavior. For some student drivers, taking corners too quickly might need the most attention; for others it might be lane changing.²² Maybe some drivers will, in the right circumstances and with the right instructors, improve more quickly than others. Perhaps the data captured could be analyzed at the macro scale to help suggest what works, what doesn’t, and which sorts of student drivers would benefit from which sorts of instruction. Not only could individualized reports be contextualized by comparing the driver’s behavior to absolute standards of safety, but the insurer could use its large stores of data to compare the young driver’s behavior to that of his or her peers. Behavioral economics teaches us that peer effects are influential in affecting behavior. Such reports could therefore serve as nudges to prompt better driving. Such innovative data products would be an economical strategy for simultaneously distinguishing an insurer from its competitors, building goodwill, and providing a service that benefits society as a whole.

Retail: At least one retailer, Tesco, is investigating the use of its club card data to do more than make promotions or upsell complementary goods (“Full fat ice cream? How about some chocolate sauce to go with it?”) The retailer is analyzing individual shoppers’ club card transactions in order to target those customers whose behavior indicates risk of such conditions as obesity or diabetes. In an interview Tesco’s CEO commented,

The information provided by (the) club card is invaluable... Our customers have told us they’d like help in choosing healthy options, so on an individual level, we want to see whether customers would welcome tailored suggestions for how they could shop more healthily. Customers would need to opt in of course, but we think it could be a really innovative way of highlighting those healthier options.²³

Tesco is complementing such efforts with behavioral economics-motivated nudge tactics to prompt customers to visit the fruit and vegetable aisle.²⁴

Consumer finance: In *Nudge*, Richard Thaler and Cass Sunstein describe many personal finance decisions as “fraught choices.” By this they mean that such decisions are made infrequently, require expert knowledge to be made well, and have effects that are felt only in the distant future. Behavioral economics teaches us that people tend to put off important decisions and also become overwhelmed (suffer “decision fatigue”) when faced with an overabundance of choices and information. Just as supermarkets can investigate the use of data to help consumers make better dietary choices, retail banks and investment companies can investigate strategies that are at once data-driven and behavioral science-informed to enable their customers to improve their financial literacy and make better savings and

investment decisions. For example, decision fatigue could be addressed with predictive “choice models” and customer segmentation analyses to create personalized menus of choices appropriate to the individual or household. Such models could be brought to life using such “choice architecture” principles as simple language that avoids excessive detail, inspired information visualization, analytically determined personalized default choices, and various types of commitment mechanisms.

This handful of ideas illustrates a larger, broadly applicable idea. In each domain, data-rich organizations face a choice. They can focus solely on using novel data sources to gain market share and ever more effectively perform such (crucial) analytical tasks as actuarial segmentation or target marketing. Or they can look beyond this and investigate ways to give back to their customers with innovative products and services.

While there is no limit to the types of innovations that can be considered, one promising framework is nicely exemplified by the use of grocery club card data to help people make healthier dietary choices. Namely, combine the use of data analytic methods and behavioral economics principles to provide people with information and tools that can help them—on an “opt-in” basis—achieve their personal goals.

NEW BOTTLES FOR NEW WINE

An overarching theme of behavioral economics is that even when people possess the relevant knowledge and harbor the best of intentions, they simply fail to act effectively. For example, most of us intend to be responsible drivers. Yet behind the wheel, some of us become aggressive, others grow absentminded, and still others give in to the potentially fatal temptation to check just that one e-mail. Many of us would like to exercise more and lose weight. Yet we give in to temptation and make impulse buys at the stores and skip visits to the gym in spite of our best intentions. And many of us for years put off tough—or even easy—“fraught” choices affecting our financial futures. We spend disproportionate time and effort making fairly minor short-term decisions (“do these eyeglasses make me look fat?”), and avoid spending time on crucial long-term decisions (should I rebalance my investment portfolio?) Nor are these isolated quirks. In *Thinking, Fast and Slow*, Daniel Kahneman discusses the general tendency for the human mind, when faced with an

important task or question, to naturally (and often unconsciously) replace it with an easier one. Inertia, peer effects, and cognitive biases compound the problem.

The general implication is that people can benefit from behavioral science-informed tools of precisely the sort sketched in the previous section. This suggests a path to innovation that is quite distinct from the use of new data sources in traditional ways. Beyond pursuing traditional targeting and segmenting-based growth strategies, organizations can also offer customers data products and services designed to help them make better choices and stick to their commitments. Doing so can help differentiate an organization from more commodity-oriented competitors and build goodwill in the process. It is interesting to note that the preceding examples involve precisely the sorts of organizations (insurers, realtors, banks) that the public indicates that it trusts least to use data responsibly and fairly.

Much of the public's skepticism about corporate and governmental uses of big data arises from the very source of that data's power: Personal behaviors that have long remained hidden from view now leave behind digital traces that can be used in ways that are often difficult for either the expert or the layperson to foresee. Without a sense of trust that the aggregators and users of such data truly have the public's interests in mind, there is little reason to expect this sense of skepticism to diminish over time.²⁵

The possibility of creating data products and services fueled by fine-grained behavioral information and shaped by the principles of behavioral science and choice architecture offers a framework for innovations that enhance rather than diminish public trust. The idea is one of many and certainly does not exhaust the topic of socially responsible big data innovation. It is reasonable to hypothesize that the organizations willing to take such ideas on board will be best positioned in the long run to distinguish themselves through superior, consumer-oriented product design and enjoy sustained profitability and growth. **DR**

James Guszcza is a senior fellow of the Singapore Deloitte Analytics Institute and the US National Predictive Analytics Lead for Deloitte Consulting's Actuarial, Risk, and Advanced Analytics practice.

David Schweidel is an associate professor of marketing at Emory University's Goizueta Business School.

Shantanu Dutta is the Dave and Jeanne Tappan Chair in Marketing at the Marshall School of Business, University of Southern California and visiting professor at Singapore Management University.

The authors would like to thank Kevin Bingham, a principal with Deloitte Consulting LLP, and David Steier, a director with Deloitte Consulting LLP, for many helpful comments.

Endnotes

1. Quoted in the *Financial Times* June 7, 2013. <<http://blogs.ft.com/tech-blog/2013/06/tim-berners-lee-prism/>>
2. See, for example, “Facebook’s New Rules” by Joe Nocera in the *New York Times*, October 18, 2013.
3. “We Like You So Much and Want to Know You Better” by Dave Eggers, *The New York Times Magazine*, September 29, 2013
4. See “Scientists must spearhead ethical use of big data” by Albert-László Barabási, *Politico*, September 30 2013
5. A thoughtful, prescient discussion is “Management’s Great Addiction” by Susan Webber, *Across the Board*, May/June 2006. Cathy O’Neil’s essay “On Being a Data Skeptic” is a recent attempt to update Webber’s essay. See also “Data Skepticism” by Mike Loukides, O’Reilly Radar, April 11, 2013.
6. See “Creepy model watch”, mathbabe.org, February 21, 2012
7. See for example “Big data is dead, long live big data: Thoughts heading to Strata” by Mike Loukides, O’Reilly Radar, February 23, 2013. We refer to the “trough of disillusionment” following the “peak of inflated expectations” in the Gartner Hype Cycle.
8. See “Too Big to Ignore” by James Guszczka, David Steier, Vivekanand Gopalkrishnan, Harvey Lewis, and John Lucker, *Deloitte Review* Issue 12, 2013.
9. This is not to say that big data in the strict sense of the term doesn’t apply to sports analytics. Only that the original Moneyball story—and many analogous stories both in an outside of the sports domain—did not involve anything approaching petabyte class data or the technologies needed to support it.
10. James Guszczka, David Steier, John Lucker, Vivekanand Gopalkrishnan, and Harvey Lewis, 2013. “Too Big Too Ignore”, *Deloitte Review* Issue 12. Also accessible at <http://cdn.dupress.com/wp-content/uploads/2013/01/DR12_Too_Big_to_Ignore1.pdf?99d1b1>
11. For a recent discussion see the US Federal Trade Commission report, “Credit-based Insurance Scores: Impacts on Consumers of Automobile Insurance”
12. Patrick L. Brockett and Linda L. Golden. 2007. “Biological and Psychobehavioral Correlates of Credit Scores and Automobile Insurance Losses: Toward an Explication of Why Credit Scoring Works.” *Journal of Risk and Insurance* 74, 23–63.
13. See “How President Obama’s campaign used big data to rally individual voters” by Sasha Issenberg, *MIT Technology Review*, December 17, 2012 and “Inside the Secret World of the Data Crunchers Who Helped Obama Win” by Michael Scherer, *Time*, November 7, 2012.
14. See “Academic ‘Dream Team’ Helped Obama’s Effort”, *The New York Times*, November 12, 2012.
15. See “Second-to-fourth digit ratio predicts success among high-frequency financial traders” by John M. Coates, Mark Gurnell, and Aldo Rustchini, *Proceedings of the National Academy of Sciences*, January 2009.
16. This is a major topic, related to big data, worthy of separate discussion. See for example “Trouble in the lab”, *The Economist*, October 19, 2013. For an intuitive discussion of one of the key issues see the “green jelly bean” discussion in our previous article “Too Big to Ignore”, *Deloitte Review* Issue 12.



17. “See Private traits and attributes are predictable from digital records of human behavior” by Michal Kosinski, David Stillwell, and Thore Graepel, *Proceedings of the National Academy of Sciences*. An unexpected pleasure of computational social science is the occasional (intentional or unintentional) humor making its way into sober scientific discussions. For example in discussing the predictive power of Likes the authors report that, “Good predictors of male homosexuality included ‘No H8 Campaign,’ ‘Mac Cosmetics,’ and ‘Wicked The Musical,’ whereas good predictors of male heterosexuality included ‘Wu-Tang Clan,’ ‘Shaq,’ and ‘Being Confused After Waking Up From Naps.”
18. In their book *Big Data: A Revolution that Will Transform How we Live, Work, and Think*, Viktor Mayer-Schönberger and Kenneth Cukier introduce the handy term “datification” to describe the digital capture of various aspects of our everyday activities.
19. A well-known cognitive bias is overconfidence bias, which means that people’s subjective confidence in their abilities and judgments often outstrip their objective reality or accuracy. For example, the great majority of American drivers rate themselves as better than the median. Such effects are also called “Lake Wobegon” effects, in homage to Garrison Keillor’s radio show *The Prairie Home Companion* which takes place in the town of Lake Wobegon where “all the children are above average”. Predictive models in contrast suffer no such cognitive biases.
20. See the edge.org conversation with Alex (Sandy) Pentland “Reinventing Society in the Wake of Big Data”.
21. This suggests that better understanding the various levels of social acceptability to various sub-populations is itself a matter of business analytics and evidence-based decision-making. Surveys, randomized field studies, and social media analytics—of course all beginning and ending with common sense and critical thinking—can presumably improve on unaided intuition in cases where the social acceptability of a business practice is ambiguous, varying amongst individuals, or dynamically changing over time.
22. The idea is somewhat analogous to one that is often discussed in the context of Massively Online Open Courses [MOOCs]. MOOCs enable the reporting of fine-grained details of the learning behavior of individual students.
23. See “Tesco: Clubcard data could be used to help shoppers make healthier eating choices”, *The Grocer*, May 23, 2013.
24. See “Buying pizza? Expect a healthy reminder from Tesco: Supermarket giant will use loyalty card data to see who is eating what”, *Daily Mail*, May 27, 2013. The “nudge tactic” mentioned in this article is in fact strikingly close to the hypothetical example of “choice architecture” *Nudge* by Richard Thaler and Cass Sunstein. They hypothesized a cafeteria whose manager put the fruits and vegetables close to the front of the line, and the sugary snacks and deserts closer to the end. The idea is that this arrangement of items would prompt diners to fill their plates more with healthy items, less with unhealthy items. This simple example illustrates Thaler and Sunstein’s concept of “libertarian paternalism”. The cafeteria arrangement is “paternalistic” in the sense that she is looking out for the wellbeing of her clients. Yet it is simultaneously “libertarian” in that diners are free to choose any and all items regardless of how they are arranged.
25. Axiom, one of the largest US data aggregators, recently made headlines with its decision to allow individuals to peer into information that the company gathers about them. Comments by Axiom’s chief strategy and marketing officer suggest that at least part of the motivation is to enhance the public’s trust: “I want to be open about what we are doing and I think there are some misperceptions about what happens and what does not happen... The more that we can talk in a straightforward manner, the more likely people are to understand it and draw their own conclusions based on the facts as opposed to supposition.” See “Finally You’ll Get To See The Secret Consumer Dossier They Have On You”, *Forbes*, June 25, 2013.

