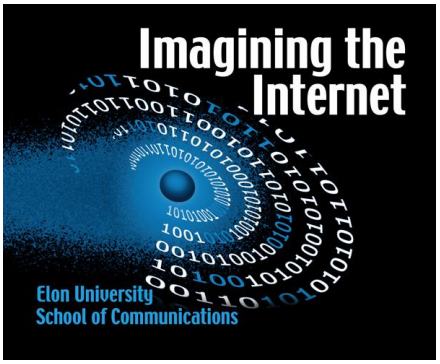
PewResearchCenter




# Big Data: Experts say new forms of information analysis will help people be more nimble and adaptive, but worry over humans' capacity to understand and use these new tools well

**Tech experts believe the vast quantities of data that humans and machines will be creating by the year 2020 could enhance productivity, improve organizational transparency, and expand the frontier of the "knowable future." But they worry about "humanity's dashboard" being in government and corporate hands and they are anxious about people's ability to analyze it wisely**

Janna Quitney Anderson, Elon University
Lee Rainie, Pew Research Center's Internet & American Life Project
July 20, 2012

Pew Research Center's Internet & American Life Project
An initiative of the Pew Research Center
1615 L St., NW – Suite 700
Washington, D.C. 20036
202-419-4500 | pewInternet.org

---

THE FUTURE OF THE INTERNET

**This publication is part of a Pew Research Center series that captures people's expectations for the future of the Internet, in the process presenting a snapshot of current attitudes. Find out more at: http://www.pewInternet.org/topics/Future-of-the-Internet.aspx and http://www.imaginingtheInternet.org.**

## Overview

We swim in a sea of data … and the sea level is rising rapidly.

Tens of millions of connected people, billions of sensors, trillions of transactions now work to create unimaginable amounts of information. An equivalent amount of data is generated by people simply going about their lives, creating what the McKinsey Global Institute calls "digital exhaust"—data given off as a byproduct of other activities such as their Internet browsing and searching or moving around with their smartphone in their pocket.[1]

Human-created information is only part of the story, a relatively shrinking part. Machines and implanted sensors in oceans, in the soil, in pallets of products, in gambling casino chips, in pet collars, and countless other places are generating data and sharing it directly with data "readers" and other machines that do not involve human intervention.

The projected growth of data from all kinds of sources is staggering—to the point where some worry that in the foreseeable future our digital systems of storage and dissemination will not be able to keep up with the simple act of finding places to keep the data and move it around to all those who are interested in it.[2]

Government leaders, scientists, corporate leaders, health officials, and education specialists are anxious to see if new kinds of analysis of large data sets can yield insights into how people behave, what they might buy, and how they might respond to new products, services, and public policy programs.

In March 2012, the White House Office of Science and Technology Policy (OSTP) announced a Big Data Research and Development Initiative, reporting that six U.S. government agencies would spend more than $200 million to help the government better organize and analyze large volumes of digital data.[3] The project is designed to focus on building technologies to collect, store and manage huge quantities of data. OSTP wants to use the technology to accelerate discovery in science and engineering fields and improve national security and education, the White House said.

How could Big Data be significant? A 2011 industry report by global management consulting firm McKinsey argued that five new kinds of value might come from abundant data: **1)** creating transparency in organizational activities that can be used to increase efficiency; **2)** enabling more thorough analysis of employee and systems performances in ways that allow experiments and feedback; **3)** segmenting populations in order to customize actions; **4)** replacing/supporting human decision making with automated algorithms; and **5)** innovating new business models, products, and services.

> "Our research finds that data can create significant value for the world economy, enhancing the productivity and competitiveness of companies and the public sector and creating substantial economic surplus for consumers. For instance, if

---

[1] McKinsey Global Institute. "Big data: The next frontier for innovation, competition, and productivity." May 2011. Available at: http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

[2] See, for instance, The Economist, "Data, data everywhere." Feb. 25, 2010. Available at: http://www.economist.com/node/15557443?story_id=15557443

[3] Office of Science and Technology Policy, Executive Office of the President, "Obama administration unveils Big Data initiative.". March 29, 2012. Available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

> US health care could use Big Data creatively and effectively to drive efficiency and quality, we estimate that the potential value from data in the sector could be more than $300 billion in value every year, two-thirds of which would be in the form of reducing national health care expenditures by about 8%. In the private sector, we estimate, for example, that a retailer using Big Data to the full has the potential to increase its operating margin by more than 60%."[4]

Indeed, the race to come up with special analytics and algorithms for working with data is driving more and more corporate activity. As the Economist reported:

> "Data are becoming the new raw material of business: an economic input almost on a par with capital and labour. 'Every day I wake up and ask, "how can I flow data better, manage data better, analyze data better?" says Rollin Ford, the CIO of Wal-Mart."[5]

Craig Mundie, chief research and strategy officer at Microsoft, was quoted later in the same story musing about the emergence of a "data-centered economy."

While enthusiasts see great potential for using Big Data, privacy advocates are worried as more and more data is collected about people—both as they knowingly disclose things in such things as their postings through social media and as they unknowingly share digital details about themselves as they march through life. Not only do the advocates worry about profiling, they also worry that those who crunch Big Data with algorithms might draw the wrong conclusions about who someone is, how she might behave in the future, and how to apply the correlations that will emerge in the data analysis.

There are also plenty of technical problems. Much of the data being generated now is "unstructured" and sloppily organized. Getting it into shape for analysis is no tiny task.

Imagine where we might be in 2020. The Pew Research Center's Internet & American Life Project and Elon University's Imagining the Internet Center asked digital stakeholders to weigh two scenarios for 2020, select the one most likely to evolve, and elaborate on the choice. One sketched out a relatively positive future where Big Data are drawn together in ways that will improve social, political, and economic intelligence. The other expressed the view that Big Data could cause more problems than it solves between now and 2020.

Respondents to our query rendered a decidedly split verdict.

53% agreed with the first statement:

> *Thanks to many changes, including the building of "the Internet of Things," human and machine analysis of large data sets will improve social, political, and economic intelligence by 2020. The rise of what is known as "Big Data" will facilitate things like "nowcasting" (real-time "forecasting" of events); the development of "inferential software" that assesses data patterns to project outcomes; and the creation of algorithms for advanced correlations that enable new understanding of the world. Overall, the rise of Big Data is a huge positive for society in nearly all respects.*

---

[4] McKinsey. Op cit.
[5] The Economist. Op cit.

39% agreed with the second statement, which posited:

> *Thanks to many changes, including the building of "the Internet of Things," human and machine analysis of Big Data will cause more problems than it solves by 2020. The existence of huge data sets for analysis will engender false confidence in our predictive powers and will lead many to make significant and hurtful mistakes. Moreover, analysis of Big Data will be misused by powerful people and institutions with selfish agendas who manipulate findings to make the case for what they want. And the advent of Big Data has a harmful impact because it serves the majority (at times inaccurately) while diminishing the minority and ignoring important outliers. Overall, the rise of Big Data is a big negative for society in nearly all respects.*

Respondents were not allowed to select both scenarios; the question was framed this way in order to encourage a spirited and deeply considered written elaboration about the potential of a future with unimaginable amounts of data available to people and organizations. While about half agreed with the statement that Big Data will yield a positive future, many who chose that view observed that this choice is their hope more than their prediction. A significant number of the survey participants said while they chose the positive or the negative result they expect the true outcome in 2020 will be a little bit of both scenarios.

Respondents were asked to read the alternative visions and give narrative explanations for their answers using the following guideline questions, "What impact will Big Data have in 2020? What are the positives, negatives, and shades of grey in the likely future you anticipate? How will use of Big Data change analysis of the world, change the way business decisions are made, change the way that people are understood?"

**Here are some of the major themes and arguments they made:**

## Those who see mostly positives for the future of Big Data share the upside

***By 2020, the use of Big Data will improve our understanding of ourselves and the world***.

- "Media and regulators are demonizing Big Data and its supposed threat to privacy," noted **Jeff Jarvis**, professor, pundit and blogger. "Such moral panics have occurred often thanks to changes in technology...But the moral of the story remains: there is value to be found in this data, value in our newfound publicness. Google's founders have urged government regulators not to require them to quickly delete searches because, in their patterns and anomalies, they have found the ability to track the outbreak of the flu before health officials could and they believe that by similarly tracking a pandemic, millions of lives could be saved. Demonizing data, big or small, is demonizing knowledge, and that is never wise."

  **Sean Mead**, director of analytics at Mead, Mead & Clark, Interbrand, added: "Large, publicly available data sets, easier tools, wider distribution of analytics skills, and early stage artificial intelligence software will lead to a burst of economic activity and increased productivity comparable to that of the Internet and PC revolutions of the mid to late 1990s. Social movements will arise to free up access to large data repositories, to restrict the development and use of AIs, and to 'liberate' AIs."

**David Weinberger** of Harvard University's Berkman Center observed, "We are just beginning to understand the range of problems Big Data can solve, even though it means acknowledging that we're less unpredictable, free, madcap creatures than we'd like to think. It also raises the prospect of some of our most important knowledge will consist of truths we can't understand because our pathetic human brains are just too small."

"Big Data is the new oil," said **Bryan Trogdon**, an entrepreneur and user-experience professional. "The companies, governments, and organizations that are able to mine this resource will have an enormous advantage over those that don't. With speed, agility, and innovation determining the winners and losers, Big Data allows us to move from a mindset of 'measure twice, cut once' to one of 'place small bets fast.'"

***"Nowcasting," real-time data analysis, and pattern recognition will surely get better.***

- **Hal Varian**, chief economist at Google, wrote: "I'm a big believer in nowcasting. Nearly every large company has a real-time data warehouse and has more timely data on the economy than our government agencies. In the next decade we will see a public/private partnership that allows the government to take advantage of some of these private-sector data stores. This is likely to lead to a better informed, more pro-active fiscal and monetary policy."

  "Global climate change will make it imperative that we proceed in this direction of nowcasting to make our societies more nimble and adaptive to both human-caused environmental events and extreme weather events or decadal scale changes," wrote **Gina Maranto**, co-director for ecosystem science and coordinator, graduate program in environmental science at the University of Miami. "Coupled with the data, though, we must have a much better understanding of decision making, which means extending knowledge about cognitive biases, about boundary work (scientists, citizens, and policymakers working together to weigh options on the basis not only of empirical evidence but also of values)."

  And **Tiffany Shlain**, director and producer of the film *Connected* and founder of The Webby Awards, maintained: "Big Data allows us to see patterns we have never seen before. This will clearly show us interdependence and connections that will lead to a new way of looking at everything. It will let us see the 'real-time' cause and effect of our actions. What we buy, eat, donate, and throw away will be visual in a real-time map to see the ripple effect of our actions. That could only lead to mores-conscious behavior."

***The good of Big Data will outweigh the bad. User innovation could lead the way, with "do-it-yourself analytics."***

- "The Internet magnifies the good, bad, and ugly of everyday life," said **danah boyd**, senior researcher for Microsoft Research. "Of course these things will be used for good. And of course they'll be used for bad and ugly. Science fiction gives us plenty of templates for imagining where that will go. But that dichotomy gets us nowhere. What will be interesting is how social dynamics, economic exchange, and information access are inflected in new ways that open up possibilities that we cannot yet imagine. This will

mean a loss of some aspects of society that we appreciate but also usher in new possibilities."

"Do-it-yourself analytics will help more people analyze and forecast than ever before," observed **Marjory S. Blumenthal**, associate provost at Georgetown University and adjunct staff officer at RAND. "This will have a variety of societal benefits and further innovation. It will also contribute to new kinds of crime."

## Some say the limitations of Big Data must be recognized

***Open access to tools and data "transparency" are necessary for people to provide information checks and balances. Are they enough?***

- "Big Data gives me hope about the possibilities of technology," said **Tom Hood**, CEO of the Maryland Association of CPAs. "Transparency, accountability, and the 'wisdom of the crowd' are all possible with the advent of Big Data combined with the tools to access and analyze the data in real time."

  **Richard Lowenberg**, director and broadband planner for the 1st-Mile Institute, urged, "Big Data should be developed within a context of openness and improved understandings of dynamic, complex whole ecosystems. There are difficult matters that must be addressed, which will take time and support, including: public- and private-sector entities agreeing to share data; providing frequently updated meta-data; openness and transparency; cost recovery; and technical standards."

***The Internet of Things will diffuse intelligence, but lots of technical hurdles must be overcome.***

- **Fred Hapgood**, a tech consultant who ran MIT's Nanosystems group in the 1990s, said, "I tend to think of the Internet of Things as multiplying points of interactivity—sensors and/or actuators—throughout the social landscape. As the cost of connectivity goes down the number of these points will go up, diffusing intelligence everywhere."

  An anonymous respondent wrote, "With the right legal and normative framework, the Internet of Things should make an astounding contribution to human life. The biggest obstacles to success are technological and behavioral; we need a rapid conversion to IPv6, and we need cooperation among all stakeholders to make the Internet of Things work. We also need global standards, not just US standards and practices, which draw practical and effective lines about how such a data trove may and may not be used consistent with human rights."

  An anonymous survey participant said, "Apparently this 'Internet of Things' idea is beginning to encourage yet another round of cow-eyed Utopian thinking. Big Data will yield some successes and a lot of failures, and most people will continue merely to muddle along, hoping not to be mugged too frequently by the well-intentioned (or not) entrepreneurs and bureaucrats who delight in trying to use this shiny new toy to fix the world."

***In the end, humans just won't be able to keep up***

- **Jeff Eisenach**, managing director, Navigant Economics LLC, a consulting business, formerly a senior policy expert with the US Federal Trade Commission, had this to say: "Big Data will not be so big. Most data will remain proprietary, or reside in incompatible formats and inaccessible databases where it cannot be used in 'real time.' The gap between what is theoretically possible and what is done (in terms of using real-time data to understand and forecast cultural, economic, and social phenomena) will continue to grow."

***Humans, rather than machines, will still be the most capable of extracting insight and making judgments using Big Data. Statistics can still lie.***

- "By 2020, most insights and significant advances will still be the result of trained, imaginative, inquisitive, and insightful minds," wrote **Donald G. Barnes**, visiting professor at Guangxi University in China.

  **David D. Burstein**, founder of Generation18, a youth-run voter-engagement organization, said, "As long as the growth of Big Data is coupled with growth of refined curation and curators it will be an asset. Without those curators the data will become more and more plentiful, more overwhelming and [it will] confuse our political and social conversations by an overabundance of numbers that can make any point we want to make them make."

## Those who see mostly negatives between now and 2020 share the down side

***Take off the rose-colored glasses: Big Data has the potential for significant negative impacts that may be impossible to avoid. "How to Lie with the Internet of Things" will be a best-seller.***

- "There is a need to think a bit more about the distribution of the harms that flow from the rise of big, medium, and little data gatherers, brokers, and users," observed communications expert **Oscar Gandy**. "If 'Big Data' could be used primarily for social benefit, rather than the pursuit of profit (and the social-control systems that support that effort), then I could 'sign on' to the data-driven future and its expression through the Internet of Things."

  "We can now make catastrophic miscalculations in nanoseconds and broadcast them universally. We have lost the balance inherent in 'lag time,'" added **Marcia Richards Suelzer**, senior analyst at Wolters Kluwer

  An anonymous survey participant wrote, "Big Data will generate misinformation and will be manipulated by people or institutions to display the findings they want. The general public will not understand the underlying conflicts and will naively trust the output. This is already happening and will only get worse as Big Data continues to evolve." Another anonymous respondent joked, "Upside: *How to Lie with the Internet of Things* becomes an underground bestseller."

***We won't have the human or technological capacity to analyze Big Data accurately and efficiently by 2020.***

- "A lot of 'Big Data' today is biased and missing context, as it's based on convenience samples or subsets," said **Dan Ness**, principal research analyst at MetaFacts. "We're seeing valiant, yet misguided attempts to apply the deep datasets to things that have limited relevance or applicability. They're being stretched to answer the wrong questions. I'm optimistic that by 2020, this will be increasingly clear and there will be true information pioneers who will think outside the Big Data box and base decisions on a broader and balanced view. Instead of relying on the 'lamppost light,' they will develop and use the equivalent of focused flashlights."

  **Mark Watson**, senior engineer for Netflix, said, "I expect this will be quite transformative for society, though perhaps not quite in just the next eight years."

  And **Christian Huitema**, distinguished engineer with Microsoft, said, "It will take much more than ten years to master the extraction of actual knowledge from Big Data sets."

***Respondents are concerned about the motives of governments and corporations, the entities that have the most data and the incentive to analyze it. Manipulation and surveillance are at the heart of their Big Data agendas.***

- "The world is too complicated to be usefully encompassed in such an undifferentiated Big Idea. Whose 'Big Data' are we talking about? Wall Street, Google, the NSA? I am small, so generally I do not like Big," wrote **John Pike**, director of GlobalSecurity.org

  An anonymous survey participant wrote, "Data aggregation is growing today for two main purposes: National security apparatus and ever-more-focused marketing (including political) databases. Neither of these are intended for the benefit of individual network users but rather look at users as either potential terrorists or as buyers of goods and services."

  Another anonymous respondent said, "Money will drive access to large data sets and the power needed to analyze and act on the results of the analysis. The end result will, in most cases, be more effective targeting of people with the goal of having them consume more goods, which I believe is a negative for society. I would not call that misuse, but I would call it a self-serving agenda."

  Another wrote, "It is unquestionably a great time to be a mathematician who is thrilled by unwieldy data sets. While many can be used in constructive, positive ways to improve life and services for many, Big Data will predominantly be used to feed people ads based on their behavior and friends, to analyze risk potential for health and other forms of insurance, and to essentially compartmentalize people and expose them more intensely to fewer and fewer things."

***The rich will profit from Big Data and the poor will not.***

- **Brian Harvey**, a lecturer at the University of California-Berkeley, wrote, "The collection of information is going to benefit the rich, at the expense of the poor. I suppose that for

a few people that counts as a positive outcome, but your two choices should have been 'will mostly benefit the rich' or 'will mostly benefit the poor,' rather than 'good for society' and 'bad for society.' There's no such thing as 'society.' There's only wealth and poverty, and class struggle. And yes, I know about farmers in Africa using their cell phones to track prices for produce in the big cities. That's great, but it's not enough."

**Frank Odasz**, president of Lone Eagle Consulting, said, "The politics of control and the politics of appearances will continue to make the rich richer and diminish the grassroots and disenfranchised until the politics of transparency make it necessary for the top down to partner meaningfully with the bottom up in visible, measurable ways. The grassroots boom in bottom-up innovation will increasingly find new ways to self-organize as evidenced in 2011 by the Occupy Wall Street and Arab Spring movements."

***Purposeful education about Big Data might include priming for the anticipation of manipulation. Maybe trust features can be built in.***

- **Heywood Sloane**, principal at CogniPower, said, "This isn't really a question about the Internet or Big Data—it's a question about who and how much people might abuse it (or anything else), intentionally or otherwise. That is a question that is always there—thus there is a need for a countervailing forces, competition, transparency, scrutiny, and/or other ways to guard against abuse. And then be prepared to misjudge sometimes."

  "Never underestimate the stupidity and basic sinfulness of humanity," reminded **Tom Rule**, educator, technology consultant, and musician based in Macon, Georgia.

  **Barry Parr**, owner and analyst for MediaSavvy, contributed this thought: "Better information is seldom the solution to any real-world social problems. It may be the solution to lots of business problems, but it's unlikely that the benefits will accrue to the public. We're more likely to lose privacy and freedom from the rise of Big Data."

  And an anonymous respondent commented, "Data is misused today for many reasons, the solution is not to restrict the collection of data, but rather to raise the level of awareness and education about how data can be misused and how to be confident that data is being fairly represented and actually answers the questions you think it does."

## Some share comprehensive views

A number of respondents articulated a view that could be summarized as: Humans seem to think they know more than they actually know. Still, despite all of our flaws, this new way of looking at the big picture could help. One version of this kind of summary thought was written by **Stowe Boyd**, principal at Stowe Boyd and The Messengers, a research, consulting, and media business based in New York City:

> Overall, the growth of the 'Internet of Things' and 'Big Data' will feed the development of new capabilities in sensing, understanding, and manipulating the world. However, the underlying analytic machinery (like Bruce Sterling's *Engines of Meaning*) will still require human cognition and curation to connect dots and see the big picture.

> And there will be dark episodes, too, since the brightest light casts the darkest shadow. There are opportunities for terrible applications, like the growth of the surveillance society, where the authorities watch everything and analyze our actions, behavior, and movements looking for patterns of illegality, something like a real-time *Minority Report*.
>
> On the other side, access to more large data can also be a blessing, so social advocacy groups may be able to amass information at a low- or zero-cost that would be unaffordable today. For example, consider the bottom-up creation of an alternative food system, outside the control of multinational agribusiness, and connecting local and regional food producers and consumers. Such a system, what I and others call Food Tech, might come together based on open data about people's consumption, farmers' production plans, and regional, cooperative logistics tools. So it will be a mixed bag, like most human technological advances.

**The view expressed by Jerry Michalski**, founder and president of Sociate and consultant for the Institute for the Future, weaves in the good, bad, and in between in a practical way:

> Humans consistently seem to think they know more than they actually know in retrospect. Our understanding of technological effects, for example, lags by many decades the inexorable effects of implementation. See Jerry Mander's great page about whether we would have let the car drive our evolution as much as it did had we known the consequences back then (in his book *In the Absence of the Sacred*).
>
> So the best-intentioned of humans will try to use Big Data to solve Big Problems, but are unlikely to do well at it. Big Ideas have driven innumerable bad decisions over time. Think of the Domino Theory, Eugenics, and racial superiority theories—even Survival of the Fittest. These all have led us into mess after mess.
>
> Meanwhile, the worst-intentioned will have at hand immensely powerful ways to do harm, from hidden manipulation of the population to all sorts of privacy invasions. A bunch of dystopian sci-fi movies don't seem like they're that far away from our reality. Also, data coming out of fMRI experiments will convince us we know how people make decisions, leading to more mistaken policies.
>
> There are a few bright spots on the horizon. When crowds of people work openly with one another around real data, they can make real progress. See Wikipedia, OpenStreetMap, CureTogether, PatientsLikeMe, and many other projects that weren't possible pre-Internet. We need small groups empowered by Big Data, then coordinating with other small groups everywhere to find what works pragmatically.
>
> Finally, Google's use of Big Data has found remarkably simple answers to thorny problems like spell checking and translation, not to mention nascent insights on pandemic tracking and more. I fear Google's monolithic power, but admire their more clear-cut approach.

Finally, **Patrick Tucker**, deputy editor of The Futurist magazine and director of communications for the World Future Society, sees the range of changes adding up to a new dimension he calls the "knowable future" extracted from the things that machines know better about us than we know ourselves:

> Computer science, data-mining, and a growing network of sensors and information-collection software programs are giving rise to a phenomenal occurrence, the knowable future. The rate by which we can predict aspects of the future is quickening as rapidly as is the spread of the Internet, because the two are inexorably linked. The Internet is turning prediction into an equation. In research centers across the country, mathematicians, statisticians, and computer scientists are using a global network of sensors and informational collection devices and programs to plot ever more credible and detailed forecasts and scenarios.
>
> Computer-aided prediction comes in a wide variety of forms and guises, from AI programs that chart potential flu outbreaks to expensive (yet imperfect) quant algorithms that anticipate outbreaks of stock market volatility. But the basic process is not dramatically different from what plays out when the human brain makes a prediction. These systems analyze sensed data in the context of stored information to extrapolate a pattern the same way the early earthquake warning system used its network of sensors to detect the P wave and thus project the S wave.
>
> What differs between these systems, between humans predictors and machine predictors, is the sensing tools. Humans are limited to two eyes, two ears, and a network of nerve endings. Computers can sense via a much wider menagerie of data collection tools.
>
> Many firms have gotten a lot better at predicting human patterns using those sense tools. Some, like Google, are already household names. In the coming years, Google is going to leverage the massive amount of user data that it collects on a minute by minute to basis to extrapolate trends in human activity and thus predict future activity. Google has been doing this with some success in terms of flu for several years now with its popular Flu Trends program. It works exactly how you would imagine that it would. We have found a close relationship between how many people search for flu-related topics and how many people actually have flu symptoms, says Google on its Flu Trends Web site. As Nicholas Christakis described in his book *Connected* querying activity and social network activity can reveal infectious disease trends long before data on those trends is released to the public by prudent government agencies.
>
> But does the same phenomenon, i.e. more querying equals more activity, hold true for subjects beyond influenza? Consider that in 2010 two Notre Dame researchers, Zhi Da and Penhji (Paul) Gao, showed that querying activity around particular companies can, somewhat reliably, predict a stock price increase for those companies.
>
> In many ways, Google is already in the process of becoming the world's first prediction engine, since prediction is key to its business model anyway. Not everyone realizes that Google makes 28% of its revenue through its Adsense program, which shows different ads to different users on the basis of different search terms. Better personalization in

terms of display ads is a function of prediction. Anticipating user behaviors, questions, and moods, strikes at the very heart of what Google's mission to 'organize the world's information.'

…. Services like Facebook and Google+ may help us to understand a lot more about our lives and our relationships than we did before these services came into existence. But Facebook's view into our lives and how our various social circles interact will always be clearer than will ours. The question becomes, who else gets to look through that microscope?

There are dangers associated with this phenomenon. Moveon.org president Eli Pariser, in his recently released book, *The Filter Bubble* describes it as a type of 'informational determinism,' the inevitable result of too much Web personalization. The Filter Bubble is a state where 'What you've clicked on in the past determines what you see next—a Web history you're doomed to repeat. You can get stuck in a static, ever-narrowing version of yourself--an endless you-loop.'

Google and Facebook are only the most obvious offenders. They're conspicuous because they're using that data to vend services to you. But you can always opt out of using Facebook, as millions already have. And while cutting Google out of your life isn't as easy as it was a decade ago, there are ways to use Google anonymously, and, indeed, to find information without using it at all. These are networks we opt in or out of….

Futurist machines are taking over the job of inventing the future. Their predictions have consequences in the real world because our interaction with the future as individuals, groups, and nations is an expression of both personal and national identity. Regardless of what may or may not happen, the future as an idea continually shapes buying, voting, and social behavior. The future is becoming increasingly knowable. We sit on the verge of a potentially tremendous revolution in science and technology. But even those aspects of the future that are the most potentially beneficial to humankind will have disastrous effects if we fail to plan for them.

## Survey Method:
## 'Tension pairs' were designed to provoke detailed elaborations

This material was gathered in the fifth "Future of the Internet" survey conducted by the Pew Research Center's Internet & American Life Project and Elon University's Imagining the Internet Center. The surveys are conducted through an online questionnaire sent to selected experts who are encouraged to share the link with informed friends, thus also involving the highly engaged Internet public. The surveys present potential-future scenarios to which respondents react with their expectations based on current knowledge and attitudes. You can view detailed results from the 2004, 2006, 2008, and 2010 surveys here: http://www.pewInternet.org/topics/Future-of-the-Internet.aspx and http://www.elon.edu/e-web/predictions/expertsurveys/default.xhtml. Expanded results are also published in the "Future of the Internet" book series published by Cambria Press.

The surveys are conducted to help accurately identify current attitudes about the potential future for networked communications and are not meant to imply any type of futures forecast.

Respondents to the Future of the Internet V survey, fielded from August 28 to Oct. 31, 2011, were asked to consider the future of the Internet-connected world between now and 2020. They were asked to assess eight different "tension pairs"—each pair offering two different 2020 scenarios with the same overall theme and opposite outcomes—and they were asked to select the one most likely choice of two statements. The tension pairs and their alternative outcomes were constructed to reflect previous statements about the likely evolution of the Internet. They were reviewed and edited by the Pew Internet Advisory Board. Results are being released in eight separate reports over the course of 2012.

**About the survey and the participants**

Please note that this survey is primarily aimed at eliciting focused observations on the likely impact and influence of the Internet – not on the respondents' choices from the pairs of predictive statements. Many times when respondents "voted" for one scenario over another, they responded in their elaboration that both outcomes are likely to a degree or that an outcome not offered would be their true choice. Survey participants were informed that "it is likely you will struggle with most or all of the choices and some may be impossible to decide; we hope that will inspire you to write responses that will explain your answer and illuminate important issues."

Experts were located in three ways. First, several thousand were identified in an extensive canvassing of scholarly, government, and business documents from the period 1990-1995 to see who had ventured predictions about the future impact of the Internet. Second several hundred of them have participated in the first four surveys conducted by Pew Internet and Elon University, and they were recontacted for this survey. Third, expert participants were selected due to their positions as stakeholders in the development of the Internet. The experts were invited to encourage people they know to also participate. Participants were allowed to remain anonymous; 57% shared their name in response to at least one question

Here are some of the respondents: danah boyd, Clay Shirky, Bob Frankston, Glenn Edens, Charlie Firestone, Amber Case, Paul Jones, Dave Crocker, Susan Crawford, Jonathan Grudin, Danny Sullivan, Patrick Tucker, Rob Atkinson, Raimundo Beca, Hal Varian, Richard Forno, Jeff Jarvis, David

Weinberger, Geoff Livingstone, Stowe Boyd, Link Hoewing, Christian Huitema, Steve Jones, Rebecca MacKinnon, Mike Liebhold, Sandra Braman, Ian Peter, Mack Reed, Seth Finkelstein, Jim Warren, Tiffany Shlain, Robert Cannon, and Bill Woodcock.

The respondents' remarks reflect their personal positions on the issues and are not the positions of their employers, however their leadership roles in key organizations help identify them as experts. Following is a representative list of some of the institutions at which respondents work or have affiliations or previous work experience: Google, the World Bank, Microsoft. Cisco Systems, Yahoo, Intel, IBM, Hewlett-Packard, Ericsson Research, Nokia, O'Reilly Media, Verizon Communications, Institute for the Future, Federal Communications Commission, World Wide Web Consortium, National Geographic Society, Association of Internet Researchers, Internet2, Internet Society, Institute for the Future, Santa Fe Institute, Harvard University, MIT, Yale University, Georgetown University, Oxford Internet Institute, Princeton University, Carnegie-Mellon University, University of Pennsylvania, University of California-Berkeley, Columbia University, University of Southern California, Cornell University, University of North Carolina, Purdue University, Duke University, Syracuse University, New York University, Ohio University, Georgia Institute of Technology, Florida State University, University of Kentucky, University of Texas, University of Maryland, University of Kansas, University of Illinois, and Boston College.

While many respondents are at the pinnacle of Internet leadership, some of the survey respondents are "working in the trenches" of building the web. Most of the people in this latter segment of responders came to the survey by invitation because they are on the email list of the Pew Internet & American Life Project, they responded to notices about the survey on social media sites or they were invited by the expert invitees. They are not necessarily opinion leaders for their industries or well-known futurists, but it is striking how much their views are distributed in ways that parallel those who are celebrated in the technology field.

While a wide range of opinion from experts, organizations, and interested institutions was sought, this survey should not be taken as a representative canvassing of Internet experts. By design, this survey was an "opt in," self-selecting effort. That process does not yield a random, representative sample. The quantitative results are based on a non-random online sample of 1,021 Internet experts and other Internet users, recruited by email invitation, Twitter, Google+, or Facebook. Since the data are based on a non-random sample, a margin of error cannot be computed, and results are not projectable to any population other than the respondents in this sample.

When asked about their primary workplace, 40% of the survey participants identified themselves as a research scientist or as employed by a college or university; 12% said they were employed by a company whose focus is on information technology; 11% said they work at a non-profit organization; 8% said they work at a consulting business, 10% said they work at a company that uses information technology extensively; 5% noted they work for a government agency; 2% said they work for a publication or media company.

When asked about their "primary area of Internet interest," 15% identified themselves as research scientists; 11% said they were futurists or consultants; 11% said they were entrepreneurs or business leaders; 11% as authors, editors or journalists; 10% as technology developers or administrators; 6% as advocates or activist users; 5% as legislators, politicians or lawyers; 3% as pioneers or originators; and 28% specified their primary area of interest as "other."

## Main Findings: Influence of Big Data in 2020

| % | TOTAL RESPONSES | Tension pair on future of Big Data |
|---|---|---|
| | 53 | Thanks to many changes, including the building of "the Internet of Things," human and machine analysis of large data sets will improve social, political, and economic intelligence by 2020. The rise of what is known as "Big Data" will facilitate things like "nowcasting" (real-time "forecasting" of events); the development of "inferential software" that assesses data patterns to project outcomes; and the creation of algorithms for advanced correlations that enable new understanding of the world. Overall, the rise of Big Data is a huge positive for society in nearly all respects. |
| | 39 | Thanks to many changes, including the building of "the Internet of Things," human and machine analysis of Big Data will cause more problems than it solves by 2020. The existence of huge data sets for analysis will engender false confidence in our predictive powers and will lead many to make significant and hurtful mistakes. Moreover, analysis of Big Data will be misused by powerful people and institutions with selfish agendas who manipulate findings to make the case for what they want. And the advent of Big Data has a harmful impact because it serves the majority (at times inaccurately) while diminishing the minority and ignoring important outliers. Overall, the rise of Big Data is a big negative for society in nearly all respects. |
| | 8 | Did not respond |

**PLEASE ELABORATE:** What impact will Big Data have in 2020? What are the positives, negatives, and shades of grey in the likely future you anticipate? How will use of Big Data change analysis of the world, change the way business decisions are made, change the way that people are understood? (If you want your answer cited to you, please begin your elaboration by typing your name and professional identity. Otherwise your comment will be anonymous.)

***Note:*** The survey results are based on a non-random online sample of 1,021 Internet experts and other Internet users, recruited via email invitation, conference invitation, or link shared on Twitter, Google Plus or Facebook from the Pew Research Center's Internet & American Life Project and Elon University. Since the data are based on a non-random sample, a margin of error cannot be computed, and the results are not projectable to any population other than the people participating in this sample. The "predictive" scenarios used in this tension pair were composed based on current popular speculation. They were created to elicit thoughtful responses to commonly found speculative futures thinking on this topic in 2011; this is not a formal forecast.

## Respondents' thoughts

One major sign of the sanctification of Big Data as a topic of interest with vast potential emerged in March this year when the National Science Foundation and National Institutes of Health joined forces "to develop new methods to derive knowledge from data; construct new infrastructure to manage, curate and serve data to communities; and forge new approaches for

associated education and training," NSF Director Subra Suresh announced in a letter to researchers in engineers, computers, and information science.[6] He said the "program aims to advance the core scientific and technological means of managing, analyzing, visualizing, and extracting information from large, diverse, distributed, and heterogeneous data sets in order to accelerate progress in science and engineering research."

The effort could hardly begin soon enough. Many are excited about the prospects for analyzing Big Data. Rick Smolan, creator of the "Day in the Life" photography series, is in the middle of a project he calls "The Human Face of Big Data," documenting the collection and uses of data. He says that Big Data has the potential to be "humanity's dashboard," an intelligent tool that can help combat poverty, crime, and pollution.[7]

Still, there is uncertainty about how effective it will be. One illustration is a recent survey of chief marketing officers at major corporations: 75% of survey respondents said they believed that leveraging data will help their companies dramatically improve their business, yet more than half said they currently lack the tools to mine true customer insights from the data generated by digital and offline efforts.[8] In the survey, 58% of respondents said they lacked the skills and technology to perform analytics on marketing data, and more than 70% said they aren't able to leverage the value of customer data.

There is already evidence in everyday life of use of Big Data:

- Every time Google suggests a spelling change in a search query, it's because previous queries on the same subject used different spellings that were found more useful. The firm's analysis of trillions of search queries yields those spell-change suggestions.[9] Google economist Hal Varian has talked about the firm's ability to spot trends from search queries allow it to forecast economic and public health trends.

- Every time someone gets a call from a credit/debit card company about "unusual activity" on their cards, the call is arriving because firms are churning through billions of transactions looking for anomalies in consumer behavior that are potentially associated with fraud or identity theft.[10]

- In April, Forbes ran through examples of Big Data operations at well known firms:[11] "Netflix, for example, takes all of its customers' viewing habits and movie ratings and runs them through a sophisticated algorithm to generate the 5-star recommendation system tailored for each subscriber. Amazon.com does a form of this, too. Online dating site OKCupid generates a steady stream of often hilarious insights into modern romance by sifting through its user profiles looking for correlations. An iPhone app

---

[6] See "Core Techniques and Technologies for Advancing Big Data Science & Engineering." National Science Foundation. Available at: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767

[7] Steve Lohr, "The Age of Big Data." New York Times. Feb. 11, 2012. Available at http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html

[8] "CMOs: Big data is a game changer, but most say they're not in the game yet." Tech Journal. March 30, 2012. Available at: http://www.techjournal.org/2012/03/cmos-big-data-is-a-game-changer-but-most-say-theyre-not-in-the-game-yet-infographic/?utm_source=contactology&utm_medium=email&utm_campaign=eWire_%2803-30-12%29

[9] See for more details: http://answers.google.com/answers/threadview/id/526503.html

[10] "7 reasons your credit card gets blocked." Fox Business News. Available at: http://www.foxbusiness.com/personal-finance/2010/08/06/reasons-credit-card-gets-blocked/

[11] Bruce Upbin, "How Intuit Uses Big Data for The Little Guy." Forbes. April 26, 2012. Available at: http://www.forbes.com/sites/bruceupbin/2012/04/26/how-intuit-uses-big-data-for-the-little-guy/

called Ness uses your own social network and preferences to generate a personalized restaurant search engine."

- The "Target Snafu" got major attention at an O'Reilly Strata conference on Big Data last spring. As Patrick Tucker blogged from the conference for the World Future Society: The *New York Times* reported in February that retailer Target "used customer data and predictive analytics to figure out that one of their customers was pregnant, and *even more remarkably,* what trimester she was in. They emailed her some promotional material and the girl's father discovered his daughter was pregnant based on the coupons she started receiving from a big box retailer, which gave rise to an awkward conversation, no doubt."[12]

This growing focus on Big Data prompted us to pose two scenarios eliciting expert views on how things might unfold by the year 2020.

After being asked to choose one of the two 2020 scenarios presented in this survey question, respondents were also asked, "What impact will Big Data have in 2020? What are the positives, negatives, and shades of grey in the likely future you anticipate? How will use of Big Data change analysis of the world, change the way business decisions are made, change the way that people are understood?"

A number of survey participants questioned the language used to describe the positive-outcome scenario. "Massive increases in the volume and availability of data will certainly improve the power of analytical and predictive tools, but there will not be some kind of major shift to more 'knowable' outcomes," wrote an anonymous respondent. "Having more data does not change the fact that there are too many interoperating variables for meaningful prediction to be possible for many things—e.g. weather. To the extent that people are saying that certain things are more predictable than they used to be they are either lying or using magical thinking."

Another wrote, "As part of the Big Data sector, I have only modest expectations of its positive impacts. There is very little evidence that there is growing practice of 'evidence based decision-making.'" However another anonymous respondent disagreed, saying, "As more people enter the digital age there will be more minds working to improve the way people communicate, curate information, and even predict events. Studying how people interact based on time, day, use of language, and comparing it to particular events it would not be impossible to say there could be patterns for identification of future events. If such a process of identifying algorithmic processes doesn't occur by 2020, it will be well on its way."

Others directed their predictions to the scenarios. What follows is a selection from the hundreds of written responses survey participants shared when answering this question. About half of the expert survey respondents elected to remain anonymous, not taking credit for their remarks. Because people's expertise is an important element of their participation in the conversation, the formal report primarily includes the comments of those who took credit for what they said. The full set of expert responses, anonymous and not, can be found online at http://www.imaginingtheInternet.org. The selected statements that follow here are grouped under headings that indicate some of the major themes emerging from the overall responses.

---

[12] Patrick Tucker, "The Three Things You Need to Know About Big Data, Right Now." World Future Society. Available at: http://www.wfs.org/content/three-things-you-need-know-about-big-data-right-now

## By 2020 we should be seeing progress in the use of Big Data to improve our understanding of ourselves and the world

Many respondents felt sure that Big Data analysis will have progressed to the point in 2020 where practical, everyday applications of it will show up in people's and organizations' lives and provide help.

**Bryan Trogdon**, entrepreneur and Semantic Web evangelist, said, "Big Data is the new oil. The companies, governments and organizations that are able to mine this resource will have an enormous advantage over those that don't. With speed, agility, and innovation determining the winners and losers, Big Data allows us to move from a mindset of 'measure twice, cut once' to one of 'place small bets fast.'"

**Paul Jones**, clinical associate professor at the University of North Carolina-Chapel Hill, says a lot of evolution will take place in the next few years. "I expect misuse and regulation responding to that misuse in the near term," he wrote. "By 2020 behaviors and actions surrounding Big Data will be normalized and a lot less scary and chaotic. The rewards that can be reaped from understanding the world through Big Data are giant and capable of changing society for the better." **Ross Rader**, board member of the Canadian Internet Registration Authority, agreed. "Only by 2020 will we have enough of a fundamental understanding to start truly doing great things with Big Data. We will make a lot of mistakes in the next ten years, endure predictions about 'The Death of Big Data' and slowly but surely, we will develop the tools and understanding necessary to turn the rise of Big Data into a positive force for change."

**Amber Case**, CEO of Geoloqi, expects positive progress. "When data can't speak to each other, time and effort is wasted," she pointed out. "In many cases the use of analytics is a way of understanding long-term trends or identifying emergent behavior that may evolve into long-term problems. As with any natural process, there will be mistakes and errors, but there will also be great benefits, one of which is the reduction of the time and space it takes to get work done or understand a process."

The ongoing evolution of code is seen as a plus. **Laura Lee Dooley**, online engagement architect and strategist for the World Resources Institute, wrote, "Building on XML, we will enhance and enforce a structured language labeling method for data gathering so data can be incorporated into datasets more easily and seamlessly. This would allow more time for analysis, requiring less time for data formatting and cleaning. This will also enable researchers to quickly respond to information needs by providing mashups of data which can inform quick decision-making."

**Don Hausrath**, retired from the US Information Agency, sees positives. "Big Data will prevail," he wrote. "Be it the design of war strategies by a UNIVAC in Bethesda during the Vietnam War, to the design of the BART System in Berkeley, gaming systems using sophisticated data sets are better at identifying solutions. In fact, the use of non-traditional statistical analysis in the BART System contributed to the winning of a Nobel Prize in Economics to one of the consultants. It is absolutely false that Big Data will diminish our lives. The use of modern statistical analysis is such that nuanced results are not only possible, but routine."

“More information will be beneficial in all sorts of ways we can't even fathom right now, namely because we don't have the data,” said **John Capone**, freelance writer and journalist, former editor of MediaPost Communications publications.

An anonymous respondent said, “Time to embrace something that is bigger than our brains, but also to put our brains to use to manage the input and control the analysis. A win-win. We think harder and get smarter.”

Some respondents shared their enthusiasm about the benefits of real-time data. **Hal Varian**, chief economist at Google noted, “I’m a big believer in nowcasting. Nearly every large company has a real-time data warehouse and has more timely data on the economy than our government agencies. In the next decade we will see a public/private partnership that allows the government to take advantage of some of these private sector data stores. This is likely to lead to a better informed, more pro-active fiscal and monetary policy.”

**Gina Maranto**, co-director for ecosystem science and policy at the University of Miami, said, “I believe, with Hans Rosling, that the more data we analyze, the better off we will be. Global climate change will make it imperative that we proceed in this direction of nowcasting to make our societies more nimble and adaptive to both human caused environmental events (e.g., Deepwater Horizon) and extreme weather events or decadal scale changes such as droughts. Coupled with the data, though, we must have a much better understanding of decision making, which means extending knowledge about cognitive biases, about boundary work (scientists, citizens, and policymakers working together to weigh options on the basis not only of empirical evidence but also of values).”

**Tiffany Shlain**, director and producer of the film ‘Connected’ and founder of The Webby Awards, wrote, “Big Data allows us to see patterns we have never seen before. This will clearly show us interdependence and connections that will lead to a new way of looking at everything. It will let us see the ‘real-time’ cause and effect of our actions. What we buy, eat, donate, and throw away will be visual in a real-time map to see the ripple effect of our actions. That could only lead to mores-conscious behavior.”

## Some responses that concentrated on the Internet of Things (a source of Big Data) came from people arguing that we will see impressive gains

While a number of respondents expressed little confidence in much additional, useful development of the Internet of Things by 2020, many see it developing. The Internet of Things is the mixture of connected “smart objects”—devices with IP-enabled sensors and readers, RFID tags, and other identifying digital information that can feed material to machines for analysis.

“The huge prospects for the 'Internet of Things' tip me to checking the first choice,” wrote **Fred Hapgood**, technology author and consultant and moderator of the Nanosystems Interest Group at MIT in the 1990s. “I tend to think of the Internet of Things as multiplying points of interactivity—sensors and/or actuators—throughout the social landscape. As the cost of connectivity goes down the number of these points will go up, diffusing intelligence everywhere.”

An anonymous survey participant wrote, “With the right legal and normative framework, the Internet of Things should make an astounding contribution to human life. The biggest obstacles to success are technological and behavioral, we need a rapid conversion to IPv6, and we need

cooperation among all stakeholders to make the Internet of Things work. We also need global standards, not just US standards and practices, which draw practical and effective lines about how such a data trove may and may not be used consistent with human rights."

**Bob Frankston**, computing pioneer, co-developer of VisiCalc, and ACM Fellow, noted, "The Internet of Things is less about massive data than meta objects. We'll have to learn how to hide from Big Data in plain sight. I do worry about the tyranny of the major less because of Big Data than because of today's self-terrorized society seeking solace in the past."

**Bruce Nordman**, research scientist at Lawrence Berkeley National Laboratory and Internet Engineering Task Force group leader, wrote, "This topic relates directly to some of my own work on the Internet of Things. Data that is much more available in quantity, cost, and quality will be a marked feature of the coming decade, but much of that will be 'Little Data,' which is useful mostly or entirely only locally (for practical or privacy concerns). I will want data possibly related to my health kept as private as possible. My house should enable control for light, heat, sound, image, etc. that enhances my experiences and convenience, and saves resources. For example, lighting will increasingly respond to occupancy or 'presence' (not just that someone is present, but who they are, how many they are, and what activity engaged in), and so provide better lighting services, automatically, and at less net energy than before. However, who outside the building should care about the details? No one. Big Data will be a net plus, but a sizeable amount of problems will be created by it as well, particularly around security and privacy."

An anonymous respondent wrote, "We are on a path that will make very large datasets available to study the world around us. The emergence of ubiquitous, high-speed wireless environments will enable the deployment of low-cost sensors. These sensors will provide unprecedented quantities of data. Businesses are presently leading the way in 'predictive analytics.' Government has recently become attentive to such tools. In the near term (2020), these Big Data sets will begin coming on-line and professional analysts will begin using the information to make informed policy choices. Over the longer-term, the potential for abuse is strong. It is unclear that politically driven people will possess the will or skill to properly interpret data analyses. Any real abuses are likely to accumulated in the more distant future, beyond 2020."

Internet Society leader Rajnesh Singh, regional director for Asia, warned, "Embedding Internet technology in various 'things' will help us improve our lives. However, it is equally important to ensure that we use this responsibly and not too much power and control is held by any one entity. There must be appropriate checks and balances, accountability, and transparency. A lot more work needs to be done by all stakeholders to ensure we get there, and use such technology for the advancement of mankind—not its control."

An anonymous respondent said, "A risk is that the Big Data available could be used—in a Wild West of privacy rights—as a new gold mine for aggressive Internet companies. It will depend very much on the capacity governments (and the future Internet governance bodies) will have to avoid the risk that the data provided by the Internet of Things (IoT) will become the same that is today the data derived from search engines. In terms of benefits from IoT for the environment, I don't believe that their impact will be so relevant as you could believe. It will take a lot of time to standardize and to integrate existing networks and the databases of IoT. None of the existing companies will accept being maginalized via the IoT game; there will be fierce resistance to integration."

**Charlie Breindahl**, a part-time lecturer at the University of Copenhagen, predicted, “Most things—even the cheapest and most banal, such as paper clips—will carry an individual identity at some point in the future. We are already in the middle of the revolution and we now have available for analysis an unprecedented amount of data. We should get used to the idea that the important question is how much data we can afford to throw away, not how much data we can afford to collect. We now know much more than we used to know, but our knowledge merely points to new needs for research.”

**Barry Chudakov**, a consultant and visiting research fellow in the McLuhan Program in Culture and Technology at the University of Toronto developed the following scenario: “By 2020 our every movement (or click or emotion) is someone’s business model. We will first build narratives and then a worldview around that. Considering the ability to take vast quantities of data and find meaning in it through pattern-finding and analytics, we will eventually employ these analytics not only in finance, healthcare, marketing and IT, but in what we hear, see and encounter as the world goes by us and through us. There will be a dawning reality that our identities are already tied to our data. In essence, in some measure our identities are our data. Big data and the Internet of Things become an arbiter, a shibboleth, an agent of triage. As the world becomes increasingly interconnected, information holds things together: it is a binding agent for systems. As such it is not only a new decider of what’s important or not, it is a new proxy it can stand in place of anyone. By 2020 data becomes a new belief system. In human history we’ve had this sort of binder before and we used the Latin base religare, meaning to bind together, to embody this concept. Information, in the form of Big Data and the Internet of Things, becomes religion.”

A doubtful anonymous respondent observed, “Apparently this 'Internet of Things' idea is beginning to encourage yet another round of cow-eyed Utopian thinking. Big Data will yield some successes and a lot of failures, and most people will continue merely to muddle along, hoping not to be mugged too frequently by the well-intentioned (or not) entrepreneurs and bureaucrats who delight in trying to use this shiny new toy to fix the world.”

## Many expect or at least hope that the good will outweigh the bad; but some worry the balance of impacts will tip the other way

Many respondents in this sampling had a strong sense of both the benefits and problems that will emerge as Big Data becomes a great reality in corporate, government, and social life. They spoke about both dimensions of impact. Some tended to accentuate the positive even as they cautioned about coping with the negative; others worried about the things breaking more bad than good.

Here is how **danah boyd**, senior researcher with professional affiliations and work based at Microsoft Research, sees the balance of forces: “The Internet magnifies the good, bad, and ugly of everyday life. Of course these things will be used for good. And of course they'll be used for bad and ugly. Science fiction gives us plenty of templates for imagining where that will go. But that dichotomy gets us nowhere. What will be interesting is how social dynamics, economic exchange, and information access are inflected in new ways that open up possibilities that we cannot yet imagine. This will mean a loss of some aspects of society that we appreciate but also usher in new possibilities.”

**Marjory S. Blumenthal**, associate provost at Georgetown University and adjunct staff officer at RAND Corporation, predicted, "Do-it-yourself analytics will help more people analyze and forecast than ever before. This will have a variety of societal benefits and further innovation. It will also contribute to new kinds of crime."

Professional programmer **Seth Finkelstein** responded, "This is a question where I want to answer *both*. The 'choices' above are both true in their descriptions. I finally went with 'negative' because I've been advocating for years that data-mining businesses are not good models for government. But this is just the latest version of 'computers and society.'"

**Perry Hewitt**, director of digital communications and communications services at Harvard University, wrote,"'Nowcasting' is sure to stumble many times before it stands, and companies will control software tools in ways that make us all profoundly and correctly suspicious. However, fearing Big Data feels like fearing fire: it exists, its capacity to do damage is enormous, and yet it illuminates such that there is no going back. For every health care data aggregator that makes us cringe, there is, one hopes, an Esther Duflo [a MacArthur Foundation Fellow for her work on improving the lives of the world's poorest people]. Using data can inform social solutions."

**Larry Lannom**, director of information management technology and vice president at the Corporation for National Research Initiatives, wrote, "Added data will enhance our understanding of the physical world and the real-time tracking of objects in motion, e.g., shipments and inventories, and will increase the efficiency of various economic activities. Privacy will continue to be a large challenge."

**Mark Walsh**, cofounder of geniusrocket.com, said, "Sadly, this is a question that will *definitely* have different answers by category. IBM Smarter Planet will make energy use and traffic congestion get better. Big Data works. Politicians will be fed Big Data results by lobbyists to support a given conclusion, and bad things will happen. On and on down the line you will see that dichotomy: Business vs. lobbyists. One will work for positive, one for negative."

**Ted M. Coopman**, a faculty member at San Jose State University and member of the executive committee of the Association of Internet Researchers, explained, "While the ability to process huge amounts of data will bring many benefits, the lack of a theoretical coherency and understanding of how large and complex systems work will cause major problems to arise. The focus of Big Data on financial markets has not increased our understanding of how our complex and global economies work. Being able to identify variables does not lead to an understanding of them. Massive complex systems are very hard to predict. Moreover, just because we understand more does not mean we can take actions that do not create more friction or introduce variables that result in unintended consequences. At the end of the data you must act on the data and that is where we run into problems. There will always be more known unknowns and unknown unknowns than known knowns. I think that more data will only increase the former more than the latter."

**Sam Punnett**, president of FAD Research Inc., observed, "As with any new technology its arrival is a mixed blessing fraught with the peril of our decision-making organizations to utilize new potentials. The two most obvious cases in point are intelligence-gathering systems used for national security and the information systems currently employed to manage international financial markets. Both have manifested unintended consequences—in the one case a failure to

properly act on information available and in the other hugely intricate and extreme fluctuations in markets that no one can explain. I am optimistic for the potential of the Internet of Things deployed on a manageable scale. The great caution with more-ambitious systems is an over reliance on seemingly rational systems to provide total unassailable solutions. The potential for these systems to be abused or to fail to take into account unforeseen circumstances is real, underscoring the need for the design of such systems to be founded on well-considered principles addressing information privacy and civil liberties as well as the realization that the systems are constructs from data using rules. Rules created by people, with all their foibles and imperfections, are subject to the occasional 'black swan' conditions of unimagined outcomes."

**Caroline Haythornthwaite**, director and professor at the School of Library, Archival, and Information Studies of the University of British Columbia, wrote, "With any change there are equal and opposite reactions. Greater data aggregation will create privacy issues; greater visualizations will hide algorithms for generating these appealing data presentations."

She warned: "As Herbert Simon said a number of years ago, algorithms will disappear into machines and then not be reexamined."

**Stephen Masiclat**, associate professor of communications, Syracuse University, predicted, "Big Data use will be the norm for all business and an increasing sector of the population will eventually be in the business of explaining Big Data insights to people not trained to understand the statistical mechanics and limits of the systems. This will not be a universal good: in America especially people dislike the idea of classification. As our data become more granular and our analysis more refined, we'll likely see more class stratification driven by marketers and other business operations. But the benefits will very likely outweigh these negatives as we will be able to do more things more cost-effectively with the insights gained from more data."

## Open access to tools and data 'transparency' are necessary for people to provide information checks and balances. Are they enough to tilt the impacts in a positive direction?

Some respondents said the future will be positive if access to data is offered on an equal basis to all, and even "private" organizations make most of their data sets or all of them open and free. This is often referred to as data "transparency." An anonymous respondent wrote, "I'm personally very involved in this trend and I am thrilled how consistently people are pushing for open data." Another wrote, "If Big Data is not also Wide Data (that is, dispersed among as many players and citizens as possible) then it will be a negative overall."

**Alex Halavais**, vice president of the Association of Internet Researchers and author of *Search Engine Society,* wrote, "The real power of 'Big Data' will come depending largely on the degree to which it is held in private hands or openly available. Openly available data, and widespread tools for manipulating it, will create new ways of understanding and governing ourselves as individuals and as societies."

**Cyprien Lomas**, director at The Learning Centre for Land and Food Systems at the University of British Columbia, urged, "Along with the rise of Big Data should come equal and open access to the data so that assumptions can be checked and double checked and to foster a culture of looking for results in data. Access to the *same* data should allow for thousands of parallel

experiments to be run by amateurs. This ecosystem should allow the discovery of new patterns and meanings in the Big Data."

**Tom Hood**, CEO of the Maryland Association of CPAs, responded, "Big Data gives me hope about the possibilities of technology. Transparency, accountability, and the 'wisdom of the crowd' are all possible with the advent of Big Data combined with the tools to access and analyze the data in real time. Many examples are already in progress. In the accounting profession there is the advent of XBRL (eXtensible Business Reporting Language), an open source, and standardized business reporting language, which is a subset of XML. This is already being deployed with mandatory financial reporting with the SEC, FDIC, and is being proposed for government spending accountability via the DATA Act of 2011 (Digital Accountability and Transparency Act). The use of XBRL is also being used for reducing the compliance burden for businesses and government with many governments around the world (Netherlands, Australia, New Zealand, and the UK). The risks of a negative scenario revolve around data integrity and security issues. If these allow for manipulation and distortion from those in power, then the public trust will erode and a very negative scenario will rise. However, if the data is set free and there are tools for the many to access and analyze the data, then I would expect the positive scenario to be the most likely."

**Donald Neal**, senior research programmer at the University of Waikato, based in Hamilton, New Zealand, and others see promise in enabling all to easily understand the world better through data. Neal wrote, "One consequence of 'The Cloud' is that tools for Big Data analysis could be available to anyone."

**Nathan Swartzendruber**, technology education at SWON Libraries Consortium, warned the data must be open. "In order for Big Data to have a positive impact on society overall, it has to be transparent," he said. "Ordinary citizens would have to be able to query the data set and discover real answers, regardless of the light that shows on individuals or corporations or governments. There is too much at stake for these parties to allow open, transparent access to this data. As long as some data sets or parts of data sets are hidden, there is room for misuse and manipulation. I think this manipulation is sure to take place. Unless Big Data is democratized on a massive scale, it will overall have a negative impact on society. Right now, I don't see much hope for such a democratization."

**Richard Lowenberg**, director of the 1st-Mile Institute and network activist since early 1970s, noted, "Big Data should be developed within a context of openness and improved understandings of dynamic, complex whole ecosystems. There are difficult matters that must be addressed, which will take time and support, including: public and private sector entities agreeing to share data; providing frequently updated meta-data; openness and transparency; cost recovery; and technical standards."

**Cathy Cavanaugh**, associate professor of educational technology at the University of Florida-Gainesville, predicted that in this sort of world, "because people will be able to quickly create their own data manipulation apps, public datasets will be used widely for answering questions. In many cases, data analysis will augment user satisfaction and ratings, and in some cases archived user satisfaction and reports will be analyzed as the data, bringing balance to decision-making through the use of large amounts of objective and subjective information."

But **Sean Mead**, director of solutions architecture, valuation, and analytics for Mead, Mead & Clark, Interbrand, expects there will have to be a public outcry to open data to the public and there may be an AI liberation movement. “Large, publicly available data sets, easier tools, wider distribution of analytics skills, and early stage artificial intelligence software will lead to a burst of economic activity and increased productivity comparable to that of the Internet and PC revolutions of the mid to late 1990s,” he predicted. “Social movements will arise to free up access to large data repositories, to restrict the development and use of AIs, and to 'liberate' AIs.”

Some respondents said they don’t think most people will be able to identify or assess complex data sets, with or without tools and open access.

An anonymous respondent wrote, “Collection is likely to be imperceptible to most, unless law and regulation make it overt and provide the individual choice. Analysis likely will suffer from a divorce in knowledge and context between the orderers and the providers of the analysis. No example currently is better than that between the avaricious ignorance of bank executives and the technologists' naiveté about the realities of collateralized debt obligations. Reliance will lead to increasingly unstable processes where only those able to use Big Data will be able to protect themselves, with the individual increasingly at risk. Rapid program stock trading is a current, pernicious example.”

Another anonymous survey participant said, “The false-positive rate will continue to grow, but the general population doesn't understand ROC curves or false-positive/true-positive comparisons today or in the future. The few people who will understand the dangers of 'Big Data' will have high cognitive abilities and training. The general population will continue to rely on crappy results because they know no better.”

## Take off the rose-colored glasses, some argued. Big Data has the potential for significant “distribution of harms” that may be impossible to avoid

For a number of respondents, the upside of Big Data is not yet clear enough compared to the foreseeable difficulties it will create.

Longtime technology analyst **Oscar Gandy**, emeritus professor of communication at the University of Pennsylvania, was one of the forceful advocates of this view: “I recently published a book *Coming to Terms with Chance* that largely mirrors the arguments at the core of the second option. In that book and in my view more generally, there is a need to think a bit more about the distribution of the harms that flow from the rise of big, medium, and little data gatherers, brokers, and users. If ‘Big Data’ could be used primarily for social benefit, rather than the pursuit of profit (and the social control systems that support that effort), then I could ‘sign on’ to the data driven future and its expression through the Internet of Things.”

**Michael Goodson**, assistant project scientist at the University of California-Davis, wrote, “My answer is a reluctant acknowledgment of what I perceive as human nature. Basing my opinion on how effectively marketing works on many people—convincing them to do things other than what is in their personal interest—it seems likely that powerful people and institutions will use all of the data at their disposal to affect events according to their interests.”

“While the rise of Big Data yields some positives, I fear that it will mostly result in increased surveillance and more-targeted marketing efforts,” wrote **Melinda Blau**, freelance journalist and

the author of 13 books, including *Consequential Strangers: The Power of People Who Don't Seem to Matter But Really Do.*

A warning tone was taken by **Sivasubramanian Muthusamy**, president of the Internet Society India chapter in Chennai and founder and CEO of InternetStudio. "The Internet and the Internet of Things together with the accompanying explosion in the capacity to process data will indeed facilitate positive progress, but at the same time, the data explosion will definitely cause more problems than it solves in future," he wrote. "Separating necessary data from unnecessary data will pose peculiar challenges. Also, data analysis alone does not guarantee optimal decisions and optimal outcomes because there are several factors beyond data—a point that is prone to be missed in the quest for more and more data. Such volumes of data call for more elaborate data management infrastructure and complex tools for analysis, which will inevitably leave all the data in the custody of very large enterprises, good and bad, and in the hands of governments. There is tremendous power associated with such a wealth of information. It is unlikely that this power will always be used with infallible ethical standards. In particular the rise of Big Data is likely to lead to a situation where everyone is tracked every moment everywhere, out of a pointless concern for security and in a misguided quest for control."

**James A. Danowski**, a professor of communication at Northwestern University and program planner for Open-Source Intelligence and Web Mining 2011, wrote, "Mining, analytics, shortening time to predicting trends are the intensive focus of most of the information segments of the knowledge sector. Misuse will increase as cyberwarfare, as one manifestation, is projected to become much more prevalent and more state-sponsored than it currently is. Government intelligence sources are currently funding research to detect deception in social media and develop ways to counteract it, technologies which can be readily repurposed for manipulation of new 'public opinion' sources."

An anonymous survey participant said, "The hype surrounding Big Data is almost as frenzied as the hype surrounding the efficient markets theory in the 1990s, and look where that led us. While improvements in data collection and processing will lead to vast improvements in our understanding in many areas, it is not a panacea. We have had decades of experience analyzing corporate financials, stock exchanges, and market indices, and yet we still cannot predict what will happen next. Many dynamic systems, of which the stock market is one, do not lend themselves to predictable patterns, no matter how much data you gather or how much computing power you apply. I worry that overconfidence in Big Data and all-seeing algorithms will lead to terrible errors. And I worry that these systems can be gamed, made to deliver false or misleading results."

One anonymous respondent joked, "Upside: *How to Lie with the Internet of Things* becomes an underground bestseller."

Another anonymous respondent wrote, "The manpower required to appropriately tag and accurately merge all current data sets is prohibitively excessive. And that doesn't consider the new data sets being created every day. Consequently, Big Data will generate misinformation and will be manipulated by people or institutions to display the findings they want. The general public will not understand the underlying conflicts and will naively trust the output. This is already happening and will only get worse as Big Data continues to evolve."

Several people expressed concerns for the human individual in a world of Big Data. An anonymous respondent said, “We will become more addicted to what the databases tell us. It might impair risk-taking for the good. We'll depend more on models than instincts.”

**Leara Rhodes**, an associate professor of journalism and international communications at the University of Georgia, said, “Any data can be misused, information is power, and if someone has a lot of information, whether or not it is accurate, complete, or truthful is harder and harder to prove. Group thinking takes over. Diversity in thinking cycles is so important that to override it with the majority point of view will be harmful to our society and it will push people to conform and not maintain their cultural identities.”

An anonymous respondent wrote, “Somewhere along the timeline of life, logical basic understandings that root each of us to the others need to take precedence. I don't think technology or Big Data can do that or should try to tell us the future or how or what to believe. I'm not ready for that conceptual change. Technology moves fast, people not so much.”

**Stan Stark**, a consultant at Heuroes Consulting, responded, “Too much trust will be given to predictive analytics of Big Data, thereby clouding and 'greying' decisions made by big business to the detriment of their performance in customer service arenas. They will 'assume' their analytics are correct in all decision making and lose focus on 'pre' Big Data techniques that were more personalized.”

An anonymous survey respondent said, “We still haven't figured out the implications of chaos theory, and if 'Big Data' and futurecasting aren’t perfect examples of chaos-based information, then I don't know what is. Generically, we're not prepared for this great a lack of privacy; we're even less prepared for data of this magnitude available only to the powerful, rich, or connected.”

An anonymous participant wrote, “Two points: First: Big Data is not Big Knowledge. We are opening a fire hose of data pointed at ourselves, but aside from developing higher density storage, we're not doing much in terms of handling it. Major challenges here will be developing 'perceptual filters' for these data flows (analogous to the ones in our minds that allow us to, for example, not spend the whole day paying attention to the fact we're wearing socks): throwing away data points that are not likely to become knowledge, that are not likely to ever be accessed, that will only serve to take up space on a hard drive and complicate further analysis of interesting events. Second: legal protections for the citizenry (in those jurisdictions which are not decidedly autocratic) are lacking, and will be essential to prevent corporate or governmental abuse of the insights available about people through widely aggregated data, as well as through new surveillance techniques.”

Another anonymous respondent noted, “In 2020, few people understand 'Big Data' as no more than conventional 20th century statistics applied to variables measuring highly superficial and ephemeral presences in physical and cyberspace. This information will continue to be imbued with magical power to predict, but will nonetheless fail to determine the unplanned behavior of individuals who are subject to highly emergent social cues. The rise of Big Data is not a negative, but many will lose interest as the cost to acquire and maintain the data exceeds the derived benefit.”

**Jon Lebkowsky**, principal at Polycot Associates LLC and president of the Electronic Frontier Foundation-Austin, observed, “We've seen so many situations where the gloss of statistical analysis misrepresents the reality of the data analyzed. It's too easy to bend the analysis to

serve a specific goal or intention. I'm also concerned about individual data ownership and privacy issues in the world of Big Data. This is an area where the outcomes could probably be improved by regulation, but regulation is currently out of style."

One anonymous respondent shared a criticism that many survey respondents raised—that organizations that possess data are not going to be swapping files with each other, even when it is of benefit to the greater good. "I don't believe there will be large data sets that are shared across corporations, government, and universities at a widespread level like discussed in the above scenarios."

**Marcia Richards Suelzer**, senior writer and analyst at Wolters Kluwer, warned, "The biggest risk is the speed and access that the Internet provides. We can now make catastrophic miscalculations in nanoseconds and broadcast them universally. We have lost the balance inherent in 'lag time.'"

And **Barry Parr**, owner and analyst for MediaSavvy, said, "Better information is seldom the solution to any real-world social problems. It may be the solution to lots of business problems, but it's unlikely that the benefits will accrue to the public. We're more likely to lose privacy and freedom from the rise of Big Data."

## We won't have the human or technological capacity to analyze Big Data accurately and efficiently. Analysts might be looking for insight in all the wrong places

Some challenged the 2020 timeline presented in the scenario descriptions. **Mark Watson**, senior engineer for Netflix, said, "I expect this will be quite transformative for society, though perhaps not quite in just the next eight years."

Others who argued a similar line described what they think is a fundamental mismatch between the volumes of data being generated and human capacity—even with the assistance of machines—to work with large sets of data, to share sets of data, and to derive significant, accurate results.

**Mike Liebhold**, senior researcher and distinguished fellow at The Institute for the Future, predicted, "The constraints of appreciating the benefits of Big Data will be the speed of adoption of open APIs, linked data, and interoperable metadata. Continued concerns over privacy and security will constrain the utility of Big Data for inference visualization and personal analytics."

**Christian Huitema**, distinguished engineer at Microsoft, said, "Unsupervised machine learning is hard. There are many examples of supervised machine learning, but these are driven by subject-matter experts that guide the machine towards specific discoveries. It will take much more than ten years to master the extraction of actual knowledge from Big Data sets."

**Bill St. Arnaud**, consultant at SURFnet, the national education and research network building The Netherlands' next-generation Internet, noted, "The benefits and impacts will be much smaller and take a longer time to develop. Manipulating and correlating Big Data sets is hard work."

And an anonymous respondent said, “The fact that most data is unstructured is a huge issue, and I doubt that we will solve the problems associated with getting meaning from that morass.” Another anonymous survey participant wrote, “Certainly in 2020 Big Data will be more risky than trustworthy. We just won't have enough experience—the equivalent of the 100-year flood in forecasting terms—and so our systems will 'look good' on some basic problems but prove to make whoppers of mistakes.”

**Dan Ness**, principal research analyst at MetaFacts, producers of the Technology User Profile, told a tale in his response: “There's an old story about a passerby who comes across a drunk man standing under a lamppost looking for his keys. The passerby joins in the search and doesn't see anything. He asks and learns that the keys didn't fall anywhere near the lamppost, but that the drunk was looking near the lamppost because that's where the light was. A lot of 'Big Data' today is biased and missing context, as it's based on convenience samples or subsets. We're seeing valiant, yet misguided attempts to apply the deep datasets to things that have limited relevance or applicability. They're being stretched to answer the wrong questions. I'm optimistic that by 2020, this will be increasingly clear and there will be true information pioneers who will think outside the Big Data box and base decisions on a broader and balanced view. Instead of relying on the 'lamppost light,' they will develop and use the equivalent of focused flashlights.”

Seattle-based consultant **Tom Whitmore** said, “There will be a rising need not for statistical analysts, but for people who will do 'forensic data analysis'—what was actually measured to generate this datum that I'm looking at, and how close is it to what I really wanted to see measured? As more and more large data sets get generated, there will be more and more of a problem with this. Everyone knows what a datum is, and what a comparison is. And it's very clear if one begins to look that the definitions used by different people have very different implications for the meaning that can be derived. Exploratory data analysis can show you what's interesting in a large batch of numbers, whether the interesting things that are discovered reflect something useful about the labels attached to those numbers is a completely different question, and a number without appropriate descriptive attachments isn't a datum—it's just a number.”

Futurist **John Smart** says Big Data will be a huge positive, but not until the semantic Web becomes fully functional, around 2030. “Lots of folks and companies will over claim what Big Data can do for us in the next decade, and are already doing so, but that's just a mild negative. Such hype causes overinvestment in underperforming platforms and other problems, but they are mild. Once we have cybertwins (semi-intelligent agents) interfacing with us and a valuecosm in 2030, all the smallest social values groups will have their own online lobbies, and be able to find subcultures that support and advance their values. In the meantime, expect the typical chaos, hype, and inefficiencies that tech innovation always brings.”

A number of respondents said the second, negative scenario will be likely in 2020, but by 2030 or after we may have adapted and evolved to reach the point at which the positive scenario will be most prevalent. An anonymous survey participant said, “Option one would be desirable, but option two is more likely, at least for 2020. In 2020, many questions related to justice, majority vs. minority decisions, etc., will not be solved and the algorithms will still be too machine-like and not humanized enough. Option one might be a long-term scenario.”

**Jonathan Grudin**, principal researcher at Microsoft, predicted, “Data mining will be used more, but by 2020 it will still be in fairly limited ways for limited purposes, and won't have that much of an effect, though of course those marketing it will amplify the benefits. But it will probably be 2030 before it really gets powerful. Will the effects be a net plus or minus? For twenty years the direct marketing and other people have been doing this, has that been a net plus or minus? I like the advances in weather and traffic prediction. I like it when my supermarket actually offers me free items or heavily discounted items that I have actually bought there in the past, rather than random coupons. I don't expect data mining to make massive advances over this kind of stuff by 2020. It will be here faster than you think. It is like three releases of the Mac or Windows OS ago—how revolutionary have the changes been since then? I guess we have an iPad and a Kinect now, but none of this has radically transformed the lives of ‘most’ people for good or ill.”

**J. Meryl Krieger**, a sociologist at Indiana University Purdue University-Indianapolis, said, “We don't have the resources to process the data and analyze it adequately for its meaning. Also, vast quantities of quantitative information are lovely, but without the contextualization and detail that come from interviews, observation, and other qualitative techniques that vast quantity of information is essentially meaningless. In other words, that's nice but so what? Until we commit adequate resources (which currently are not available—I point out the emphasis of our current society towards monetization and specialization) towards interpretation and explanation ‘The Internet of Things’ remains a great idea and that's about it. In terms of values, it totally depends on where you sit. There are going to be people who are terrified of what comes out of finding out what people actually do; they are much more interested in having the world reflect what they know and understand and find ‘difference’ to be incredibly threatening. Such folks will always try to manipulate data sets. On the other hand are the idealists who likewise want diversity to always be a good thing and will try to manipulate data sets to reflect their vision. Integrity in scholarship is the key here - way too many people have an agenda they are pursuing. This is the threat to ‘The Internet of Things’ not the information itself.”

**Tapio Varis**, professor emeritus at the University of Tampere and principal research associate with UN Educational, Scientific, and Cultural Organization (UNESCO), noted, “The general lack of trust and confidence and the gigantic misuse of existing Big Data for monitoring and intelligence will slow down and backset progress.”

**Rich Tatum**, the research analyst for Zondervan, a religious publishing house, agreed that major trust issues lie ahead. “Such analysis will enable deception on an ever greater scale,” he wrote. “What will matter most will not be who you trust for news, or what outlet you trust, but who owns the data you use for news. And this kind of data and analysis will not be cheap.”

**Nikki Reynolds**, director of instructional technology services at Hamilton College, says overconfidence is a big risk. “We are already using data modeling to make big mistakes,” she pointed out. “In most ways, I doubt that the use of Big Data will be any more or less faulty than our current uses of the data and models we have accessible today. The real estate and sub-prime mortgage disasters are a clear case of those problems now. The best analyses of the roots of those mistakes I have read and heard point to overconfidence in poorly understood, very complex risk models, and the refusal to recognize that the worst *will* happen. When the probability of a situation occurring is 1,000 to one, that means the situation will occur, just not very often, over the long run, although possibly on two or more consecutive occasions over the short run. People seem not to pay attention to that when making decisions. Perhaps we are not really good at thinking about the long run, and ultimately having to put the 'disaster recovery

plan' into action. In any case, I don't think Big Data will make the problem of poor judgment when assessing the consequences of risk any smaller. Anything forecast based on any data is just a model, and not an event controller. We, as a species, will continue to make decisions on a short runway and we will get caught out. So, will Big Data make the consequences of our mistakes worse? Yes, and no. While the ramifications of a mistake may become more far-reaching, it will also be much harder to 'hide' mistakes and their consequences, because of the level of connectivity that we have. We've seen the accelerating potential of that connectivity in recent political events, and even in recent environmental disasters. When someone starts a protest, others know of it right away, from the immediate observers. Decisions about whether to join the protest don't have to wait for the publication of a newspaper, or the 'film at eleven'. When an oil rig blows, the whole world knows within hours. Governments and scientists swing into action immediately—not always smoothly and certainly not always cooperatively, but the response is immediate. I'm betting that our ability to respond to crises is going to increase just as rapidly, perhaps more rapidly, than our ability to create crises. I hope I'm right."

**Jeniece Lusk**, an assistant research director with a PhD in applied sociology who works at an Atlanta information technology company, wrote, "As an applied sociologist, I religiously believe in the ability of the humans who must interpret and create these data sets to muck it all up, intentionally or not. The media, the research, the Internet is all driven by humans. Human error can mess up even the best of data collection, analysis, and dissemination. Beyond that, we are unable to ever suggest or predict in a generalizable way until confidence intervals are 100% and the Census doesn't need to impute data we're not going to be able to read the future or become psychic-statisticians. Unfortunately, you won't be able to convince some audiences or segments of that, because if someone with an authoritative position tells you something that a computer calculated, you might as well call it absolute truth (unless it doesn't match up with their belief system, of course)."

And one participant noted that we don't use the data we already have. "Modern society already ignores a century of social science research when determining programs and policies," said **Cheryl Russell**, editorial director for New Strategist Publications and author of the Demo Memo Blog. "It is doubtful that the leaders of 2020 will be any more able or willing than our leaders today to use social science findings to improve our lives."

## Some concentrated their focus on the role of human judgment in the process of Big Data analysis and response

An anonymous respondent said, "The old lesson that correlation is not causation seems never to be learned. The control over data means that inaccurate data is hard to identify and correct. I see that the problems will only increase with the size of the datasets. Most emphasis seems to be given to doing clever things with data rather than ensuring its validity or giving the right people control over it."

**Michel J. Menou**, visiting professor at the department of information studies at University College London, noted, "The intelligence of systems cannot substitute for the intelligence of the individuals and organizations that use them. Since efforts are focused on the development of technology at the expense of education, consciousness raising, and democratic control, negative effects are the more likely to occur."

**Tom Rule**, an educator and technology consultant based in Macon, Georgia, wrote: “Never underestimate the stupidity and basic sinfulness of humanity.”

**William L Schrader**, an independent consultant who founded PSINet in 1989, provided a few more details. “The fact is: people are people,” he wrote. “The rich get richer and the powerful stay that way. All tools will be used by the rich for gain and by the powerful to remain so. However, the activists in the world will also have access to Big Data, and big tools, in fact, it will be innovators and activists who create those very tools. In the end, the bag is always mixed; much as the Internet brought us distance learning and distance medicine (as predicted in the 1980s), it also brought humans global access to child pornography, the opportunity to phish financial and identity information for illegal activity, and to assist governments in monitoring and controlling their populations. Simultaneously, we saw how the Internet played an integral role in the overthrow of several governments during 2009-2011 and that activity will continue. Yes, the answer is 'both,' positive and negative.”

**Miguel Alcaine**, head of the International Telecommunication Union’s area office, Tegucigalpa, Honduras, responded, “If some high-placed people believe this type of technology can predict the unpredictable, there will be cases where this technology will be overextended and misused. Human judgment cannot be replaced by technology, the former being the responsible for decisions.”

**David D. Burstein**, founder of Generation18, a youth-run voter-engagement organization, said the human element trumps all of the technology. “As long as the growth of Big Data is coupled with growth of refined curation and curators it will be an asset,” he wrote. “Without those curators the data will become more and more plentiful, more overwhelming and confuse our political and social conversations by an overabundance of numbers that can make any point we want to make them make.”

**Donald G. Barnes**, visiting professor at Guangxi University in China and former director of the Science Advisory Board at US Environmental Protection Agency, noted, “Big Data has possibilities and will result in some, but limited, number of discoveries. However, a vision of relying on results of the analysis of Big Data as the major source of breakthrough information and insights is unwarranted. Past and current examples of the analysis of Big Data suggests that we should be cautious about the fruitfulness of this type of analysis; e.g., the Department of Homeland Security being hamstrung by the torrent of information intercepted on the Internet and the limited payoff from the use of massive information sources in combinatorial chemistry and bioinformatics. The underlying problem is one of signal-to-noise; i.e., with more information, the challenge of detecting the signal can be even larger. By 2020, most insights and significant advances will still be the result of trained, imaginative, inquisitive, and insightful minds.”

**David Kirschner**, a PhD candidate and research assistant at Nanyang Technological University in Singapore, wrote, “People put way too much faith in statistics and quantitative analysis of giant data sets. It leads us to believe we can predict and forecast much better than we actually can. Forecasting leads to people assuming outcomes that don't necessarily happen, and that has real-life implications for people who win or lose based on these predictions. We also assume that we can trust interpretations of this data. Interpretations are made by people, people in positions of power who have their own agendas, and those are the interpretations people generally trust. Not smart, but we don't know any better because we believe what we're told by

'experts' and we don't have any means to find out what's really going on, reasons for this or that. It's all efficiently masked in bureaucracy. Very dangerous!"

**Jeffrey Alexander**, senior science and technology policy analyst at the Center for Science, Technology & Economic Development at SRI International said the human factor in analysis is crucial. "While 2020 is too soon for the emergence of true artificial intelligence and predictive power, the ability to manipulate social, physical, and informational inputs on a large scale will reveal new insights into behaviors and human development," he wrote. "The greater danger lies beyond 2020, when machine learning may become so effective that it crowds out human judgment."

An anonymous respondent predicted a withdrawal by some, writing, "I expect a backlash against Big Data to occur sooner rather than later, and expect to see a movement toward people reducing their presence on the grid. There is still a large percentage of the population who have a low level of Internet presence (anyone over the age of 45, for example) and this will offer the necessary contrasts for this to occur."

## People are concerned about the power agendas of governments and corporations, the interests with the most Big Data resources

A variety of responses focused on the collectors of Big Data and their motives. Among those for whom that was the framework for their answers, many were wary and full of warning about how the data could be used.

**Ed Lyell,** a professor at Adams State College, wrote: "I see two major negatives overwhelming the positive. **1)** Our trust in econometric models made the great world economic crash more likely. By getting better and better at predicting the specifics of the near future we had models that ignored big system changes and the power of market corruption by those on top. Predictive models are all subject to a movement to the mean, ignoring the rising system change caused by falling off a cliff not seen by the models and not looked for by the humans trusting models. **2)** Like *1984* and *Brave New World*, books my generation knows well, I have seen government, and even more big business use massive personalized data to control people, to not just respond to their needs but to create needs. Government has made us fearful and willing to accept increasing limits on personal freedom because of our insecurity. The wealthy elite (top 2% or so) can purchase TV ads and other media to get Congress elected for their purposes. Now big corporations can be even more active on the top of the table. This increasing power shifted to the top is moving the United States away from democracy into what I now see as a plutocracy. The RFID's in our clothes and products make tracking the outliers easier and perhaps puts people at more risk in the future."

An anonymous respondent wrote, "The correct choice depends largely on our collective choice. In the end, I selected the more pessimistic scenario because that is the choice we are in right now—the one where corporations with no sense of values, morality, or conscience make choices for humans (choices affecting humans, but motivated by mere profit for shareholders). Consider the number of TV programs dedicated to 'investing' money compared to the number of TV programs dedicated to ending poverty. As things stand right now, there is little doubt that the second option is the correct one. But fortunately, that could change if humans decide to take charge, and return corporations to their subservient role."

Another anonymous survey participant predicted, “Both outcomes will occur, concurrently, in many complex intertwined ways. Even liberal governments will feel compelled to accumulate and use data against their citizens, in many of these countries corporations run amok will do the job.”

**Julia Takahashi**, editor and publisher at Diisynology.com, wrote, “By 2020, most Internet users will be used to receiving algorithmic recommendations and will either give them little notice or will have found ways to circumvent them. In the United States a large majority of people dislike feeling that they are being manipulated or presented with fewer choices and the online retailing community is going to have to deal with this. At a community, regional, state, or national planning level there will be more use of Big Data and it will have to compete with political attitudes which seem to be trending towards suspicion of 'Big Data.' Corporations will most likely be the largest users of Big Data and may find that the data out is only as good as the data in and the suppositions that went into planning the output. I think we will see some major mistakes.”

Most of the respondents who commented with concerns over government and/or corporate control of data chose to remain anonymous. Here are more of their observations:

> — “I started dealing with data aggregation in the 1970s and have a copy of the 1970s US Health Education and Welfare report on computers, privacy, and databases on the bookshelf where I am typing this. Data aggregation is growing today for two main purposes: National security apparatus and ever more focused marketing (including political) databases. Neither of these are intended for the benefit of individual network users but rather look at users as either potential terrorists or as buyers of goods and services. Already it costs a lot for people to fetch the results of some of these things, even simple things like credit scores are available to the data subject only for a fee. Information is power, and power will cost money.”
>
> — “Whenever corporations or governments get involved with anything, they rarely behave in what could be considered an 'altruistic' fashion. Corporations will monopolize Big Data to make money; an unethical government administration could use it to wreak havoc on private lives, which I believe is already happening in the United States, under the auspices of preventing child pornography and exploitation. While certainly a worthwhile endeavor, there are implications to eroding citizens' rights to privacy to carry out sting operations. This sets an uneasy precedent where suspicion of activity can trump proof, and entrapment follows closely behind. I see this being a major problem for journalists and political bloggers in the future.”
>
> — “The false confidence already plagues risk-management 'professionals.' No one looking at the Big Databases predicted the criminal activities of the financial sector, anything could be changed to look a certain way, and any channel to data could be clogged, muddled, or dirtied to the point where an independent analysis is undermined. Files have and will be deleted on demand. There is no moral code in the algorithms, no ethics, no enforcement. These tools are only indexes pointing to areas of further research. Without a more robust system of checks and balances and independent watchdogs, these systems will not guarantee fidelity to the truth.”

— “Money will drive access to large data sets and the power needed to analyze and act on the results of the analysis. The end result will, in most cases, be more effective targeting of people with the goal of having them consume more goods, which I believe is a negative for society. I would not call that misuse, but I would call it a self-serving agenda.”

— “Data is not information, and information is not knowledge, and knowledge is not wisdom. Conducting things as they had been conducted up-to-date, the finest information will serve 'elastic' statistics, neo-Nazi supremacy visions, wars based on 'reliable intelligence relative to mass destruction weapons' or faked president elections. The ethic-control thing will turn more and more important as the power of the Internet gives power to some men.”

— “Unless some major political upheaval changes the balance of power in the world, Big Data will be primarily in the hands of the increasingly small group of the rich and powerful. The tendency of those with immense power is to use tools such as Big Data to increase their power. Therefore, if the current direction of international power structures continues and power is increasingly concentrated in the hands of a few, the capabilities of Big Data will be used to further augment that power and will not be used for the good of the community.”

— “The majority of Big Data is and will continue to be in the hands of corporate interests which by definition are selfish bastards.”

— “Big Data will probably be a cause of reduced freedom and privacy, and it will give advantages to companies that can spend money on analysis.”

— “It is unquestionably a great time to be a mathematician who is thrilled by unwieldy data sets. While many can be used in constructive, positive ways to improve life and services for many, Big Data will predominantly be used to feed people ads based on their behavior and friends, to analyze risk potential for health and other forms of insurance, and to essentially compartmentalize people and expose them more intensely to fewer and fewer things.”

— “Humanity will always have greed and corruption and deception, and this can only be mitigated by insightful analysis of fact, open sharing of information and logical decision making.”

## Some predict that algorithms will most negatively impact the lives of those who are already disadvantaged

Other answers were related to those in the previous section in that they explored the motives and behaviors of the data collectors. But their material concentrated on the effect of Big Data on those who are not themselves powerful.

**Fred Stutzman**, a postdoctoral fellow at Carnegie Mellon University and creator of the software Freedom, Anti-Social, and ClaimID, said, “We must stay mindful that Big Data is not a complete lens, especially when interpreting the human condition.”

**Steve Sawyer**, professor and associate dean of research at Syracuse University; an expert of more than 20 years of research on the Internet, computing, and work, wrote, “Our vision of the data is based on our vision of the world, and this vision is not very broad-minded when it comes to Big Data. We tend to emphasize the parietal insights of a particular form of economic thinking, and we tend to frame social analyses through a form of soft colonialism. Such bias, combined with the arrogance of technical competence, will create huge disparities between 'what the data say' and the lives of billions of people.”

**Brian Harvey**, a lecturer at the University of California-Berkeley, noted, “The collection of information is going to benefit the rich, at the expense of the poor. I suppose that for a few people that counts as a positive outcome, but your two choices should have been ‘will mostly benefit the rich’ or ‘will mostly benefit the poor,’ rather than ‘good for society’ and ‘bad for society.’ There's no such thing as ‘society.’ There's only wealth and poverty, and class struggle. And yes, I know about farmers in Africa using their cell phones to track prices for produce in the big cities. That's great, but it's not enough.”

**Ebenezer Baldwin Bowles**, owner and managing editor of corndancer.com, wrote, “With Big Data comes Great Power, and neither shall be used wisely for the common good. The objective is not to reveal opportunity for the elimination of scarcity among the many, but to identify fertile ground for exploitation and control.”

**Paul McFate**, an online communications specialist based in Provo, Utah, said, “New media channels will continue to splinter consumers and enhance the social divide. Intelligent people will use the information well, but the average person will continue to look for bright shiny objects that will entertain. Abusive people will continue to abuse. Providing access to data does not change moral behavior.”

**Daren C. Brabham**, an assistant professor of communications at the University of North Carolina-Chapel Hill, said, “Our reliance on algorithms is already proven to be problematic, evidenced by the fickle nature of the stock markets and other things. As we keep funneling the best and brightest mathematicians into algorithm-focused professions (like finance), we'll continue to abstract real labor and real human concerns further away from real consequences and circumstances. This is a massive ethical problem, too.”

**Paul Gardner-Stephen**, a telecommunications fellow at Flinders University, observed, “While many benefits will arise from an Internet of Things, while the things remain in the possession of very few centralised interests, it will present great potential for abuse. History tells us that where such potential exists and sufficient money and civil-control can be made by the abuses, that such abuses will continue in increasing measure. Face recognition and tracking alone present the simple means to create an almost inescapable police state. Hats and coats such as those worn by Spy-vs-Spy will become more appealing, although ultimately ineffective as statistical and probabilistic algorithms allow the tracking of even persons cloaked (literally or by other means).”

**Arthur Asa Berger**, professor emeritus of communications at San Francisco State University, said, “While the Internet does allow dissidents to have a voice, for the main part they are not heard relative to the power of the dominant elites, members of the ruling class, etc.”

**Frank Odasz**, president of Lone Eagle Consulting, a company specializing in Internet training for rural, remote, and indigenous learners, wrote, “The politics of control and the politics of

appearances will continue to make the rich richer and diminish the grassroots and disenfranchised until the politics of transparency make it necessary for the top down to partner meaningfully with the bottom up in visible, measurable ways. The grassroots boom in bottom-up innovation will increasingly find new ways to self-organize as evidenced in 2011 by the Occupy Wall Street and Arab Spring movements."

**David A.H. Brown**, executive director of Brown Governance Inc., a consulting business based in Toronto, Canada, noted, "Democratization is the issue; this has tremendous implications for social structure and social order (increasing pressure by 'have-nots' on 'elites') as well as privacy, family, and culture. A big unanswered question is who will control Big Data? Whomever controls the information will have greater power and influence, and they may use this for positive or negative results."

## Purposeful education about Big Data might include warnings that anticipate manipulation of data analysis outcomes; trust features might be built into the data scrutiny

Some respondents wondered if some negatives of Big Data might be mitigated by more serious study and purposeful planning.

**John Horrigan**, vice president of TechNet, a research organization, said, "'Big Data' is very much undiscovered country for citizens and policymakers, and its beneficial potential depends on getting governance and citizen education right. The tech sector is typically pretty good at ushering in new applications in a secure way. But it is easy to underplay the importance of educating the public on what this all means, not least to foster widespread use of the affordances of 'Big Data' (e.g., health care delivery, home energy management). So a word of caution is in order if such efforts are not undertaken."

**Maureen Hilyard**, development programme coordinator for the New Zealand High Commission and vice chair of the board of the Pacific Chapter of the Internet Society, responded, "The big issue to do with the Internet is *trust* in where the information comes from and how it is used. As long as people know who is offering the information and can trust its source, then there is a better understanding of what the world is about. However, confidence in the technology declines when it is misused and people are harmed as a result of miscommunication or false communication. Education in the appropriate use of the Internet, taking more flexible advantage of the diversity of access that the World Wide Web offers, and the security of online information during financial transactions, are what I see are the big issues for the future, to ensure that the Internet is used appropriately and safely, and provides the positive impacts for which it has the greatest potential."

**Hugh F. Cline**, adjunct professor of sociology and education at Columbia University, wrote, "It will be necessary to regulate these activities to ensure that they are used for the benefits of all peoples. Furthermore, it will be necessary to educate ourselves to ensure that we can recognize abuses and self-servicing frauds."

An anonymous respondent commented, "Data is misused today for many reasons, the solution is not to restrict the collection of data, but rather to raise the level of awareness and education

about how data can be misused and how to be confident that data is being fairly represented and actually answers the questions you think it does."

**John Kelly** of the Monitor Group says that savvy individuals can fight back if those in power misrepresent to the public what the data is showing. "The positive outcomes for 'Big Data' will depend on the general availability of powerful analytic and visualization tools and widespread fluency in their use. Even if consumer and pro-consumer data dashboards are less advanced than those employed by commercial business and government intelligence agencies, the ability of individuals and small organizations to develop analytically rigorous counter narratives complete with dynamic 3D graphs that go up on YouTube and get picked up by CNN could provide a check on the presumptuous uses of data mining."

**Marcel Bullinga**, futurist and author of W*elcome to the Future Cloud - 2025 in 100 Predictions,* observed, "Big Data can be manipulated as well as Small Data. It is not about big or small! It is about embedding all data with trust and privacy features. We must develop the 'Cloud Seal' and wrap all data in such a notary-like seal."

**Heywood Sloane**, principal at CogniPower, a consulting business, said, "This isn't really a question about the Internet or Big Data—it's a question about who and how much people might abuse it (or anything else), intentionally or otherwise. That is a question that is always there—thus there is a need for a countervailing forces, competition, transparency, scrutiny, and/or other ways to guard against abuse. And then be prepared to misjudge sometimes."

## Humans seem to think they know more than they actually know, but despite all of our flaws, looking at the big picture usually helps

In addition to the longer observations by Stowe Boyd, Jerry Michalski, and Patrick Tucker that were shared in the opening "Overview" of this report (on pages 9-12) several respondents wrote extended and thoughtful answers that included historical perspective and general observations about human nature and society.

**Kevin Novak**, co-chair for the eGov Working Group of the World Wide Web Consortium, speaker and author on electronic government and consultant to the World Bank on the eTransform Initiative, observed: "Society often finds itself in a perplexed state as it attempts to understand large issues, solutions, and items given the diversity of environments, actions, and opinions throughout diverse cultures of the world. It is often challenging, given this diversity, to determine what is the best course of action/plan to move forward. A growing mass of data can help better inform decision-making, identify trends, and connect data bits to see a larger picture than what was previously known. Big Data will, however, offer challenges unless the tools, methods, and technologies are developed that can aid in relating unstructured data together to tell a story. The tools, methods, and technologies will be the challenge in 2020, not the availability of the data itself. Society will continue to struggle with privacy. As more and more of our lives are archived and mined on the Web, opportunists will continue to explore ways to exploit the available data for their own, not-so-honest means. How we respond and manage should continue to be a major focus in the Internet community through 2020. We must understand the challenges and opportunities, know the gaps that exist, and offer the best chances for addressing these."

**Michael Castengera**, senior lecturer at the Grady College of Journalism, wrote: “At one point, futurists talked about the development of a 'global brain' through the Internet. That scenario may appear to be hyperbole but not if you differentiate between the definitions of 'brain' and 'mind.' The brain now creates its own version of algorithms, which allow for advanced correlations and new understanding. In the continuing evolution of the Internet, the algorithms more and more connected disjointed data in a way that mimics the brain's connecting synapses. As developments continue, the analogy could be drawn of an Internet that has an autonomous and semi-autonomous nervous system. Control of that development will be held by held by major institutions—not just corporations, but the direction of that development will be determined by individuals. It is the debate about what is the nature of the Internet and what it is to be—a debate that will pit individuals against institutions. My concern is that individuality will be lost to an institutional hegemony not because of their 'selfish agendas' so much as individuals' complacency and acceptance of continuing personal intrusions.”

And an anonymous respondent responded: “The more that data sets are open and accessible, entrepreneurial Web-savvy types will harness that raw material for different ends, and many times these may be philanthropic. We will see more visual representations of large data sets that will enable people to see the impacts of their activities as they play out in other parts of the world. Big Data will be used to forecast and predict, more simulations will be played out, and these simulations will help people to understand the complexity of our correlation to each other, as beings on this planet and beyond. People will try to ‘fix’ or ‘game’ scenarios based on simulations. We’ve already seen this in the past decade with the Wall Street crisis, but systems of this size and complexity are dynamic and self-regenerative. The realization of dynamic and emergent systems as a natural order will cause people to realize the foolishness of trying to game systems to the Nth degree. We will see the rise of more algorithmic thinking among average people, and the application of increasingly sophisticated algorithms to make sense of large-scale financial, environmental, epidemiological, and other forms of data. Innovations will be lauded as long as they register a blip in the range of large-scale emergent phenomena.”

## About the Pew Research Center's Internet & American Life Project

The Pew Research Center's Internet & American Life Project is one of seven projects that make up the Pew Research Center, a nonpartisan, nonprofit "fact tank" that provides information on the issues, attitudes and trends shaping America and the world. The Project produces reports exploring the impact of the Internet on families, communities, work and home, daily life, education, health care, and civic and political life. The Project aims to be an authoritative source on the evolution of the Internet through surveys that examine how Americans use the Internet and how their activities affect their lives.

The Pew Internet Project takes no positions on policy issues related to the Internet or other communications technologies. It does not endorse technologies, industry sectors, companies, nonprofit organizations, or individuals.

URL: http://www.pewInternet.org

## About the Imagining the Internet Center at Elon University

The Imagining the Internet Center's mission is to explore and provide insights into emerging network innovations, global development, dynamics, diffusion and governance. Its research holds a mirror to humanity's use of communications technologies, informs policy development, exposes potential futures and provides a historic record. It works to illuminate issues in order to serve the greater good, making its work public, free and open. The center is a network of Elon University faculty, students, staff, alumni, advisers, and friends working to identify, explore, and engage with the challenges and opportunities of evolving communications forms and issues. They investigate the tangible and potential pros and cons of new-media channels through active research. Among the spectrum of issues addressed are power, politics, privacy, property, augmented and virtual reality, control, and the rapid changes spurred by accelerating technology.

The Imagining the Internet Center sponsors work that brings people together to share their visions for the future of communications and the future of the world.

URL: http://www.imaginingtheInternet.org

## Methodology

The survey results are based on a non-random, opt-in, online sample of 1,021 Internet experts and other Internet users, recruited via email invitation, Twitter or Facebook from the Pew Research Center's Internet & American Life Project and the Imagining the Internet Center at Elon University. Since the data are based on a non-random sample, a margin of error cannot be computed, and the results are not projectable to any population other than the experts in this sample.