7-2014

Understanding the paradigm shift to computational social science in the presence of big data

Ray M. CHANG
Singapore Management University, mrchang@smu.edu.sg

Robert J. Kauffman
Singapore Management University, rkauffman@smu.edu.sg

Young Ok KWON
Sookmyung Women's University

DOI: https://doi.org/10.1016/j.dss.2013.08.008

Citation

CHANG, Ray M.; Kauffman, Robert J.; and KWON, Young Ok. Understanding the paradigm shift to computational social science in the presence of big data. (2014). *Decision Support Systems*. 63, 67-80. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/2108

Published in Decision Support Systems, 2014 July, Volume 63, pp. 67-80.
http://doi.org/10.1016/j.dss.2013.08.008

Understanding the paradigm shift to computational social science in the presence of big data

Ray M. Chang [a,1], Robert J. Kauffman [b,2], YoungOk Kwon [c,*]

[a] School of Information Systems, Singapore Management University, 80 Stamford Road, 178902, Singapore
[b] School of Information Systems, Singapore Management University, 80 Stamford Road, 178902, Singapore
[c] Division of Business Administration, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul 140-742, South Korea

Keywords:
Analytics
Big data
Computational social science
Data analytics
Interdisciplinary research
Managerial decision-making
Paradigm shift

ABSTRACT

The era of big data has created new opportunities for researchers to achieve high relevance and impact amid changes and transformations in how we study social science phenomena. With the emergence of new data collection technologies, advanced data mining and analytics support, there seems to be fundamental changes that are occurring with the research questions we can ask, and the research methods we can apply. The contexts include social networks and blogs, political discourse, corporate announcements, digital journalism, mobile telephony, home entertainment, online gaming, financial services, online shopping, social advertising, and social commerce. The changing costs of data collection and the new capabilities that researchers have to conduct research that leverages micro-level, meso-level and macro-level data suggest the possibility of a *scientific paradigm shift* toward *computational social science*. The new thinking related to empirical regularities analysis, experimental design, and longitudinal empirical research further suggests that these approaches can be tailored for rapid acquisition of big data sets. This will allow business analysts and researchers to achieve frequent, controlled and meaningful observations of real-world phenomena. We discuss how our philosophy of science should be changing in step with the times, and illustrate our perspective with comparisons between earlier and current research inquiry. We argue against the assertion that theory no longer matters and offer some new research directions.

> “To get value from big data, ‘quants’ or data scientists are becoming analytic innovators who create tremendous business value within an organization, quickly exploring and uncovering game-changing insights from vast volumes of data, as opposed to merely accessing transactional data for operational reporting.”
> [Randy Lee, Vice President, Aster Center for Data Innovation, Teradata [81]]

> “The best way to engage in … data-driven marketing is to gather more and more specific information about customer preferences, run experiments and analyses on the new data, and determine ways of appealing to [casino game] players' interests. We realized that the information in our database, couple with decision science tools that enabled us to predict individual customer's theoretical value to us, would allow us to create marketing interventions that profitably addressed players' unique preferences.”
> [Gary Loveman, CEO and President of Caesar's Entertainment [70]]

> “Each methodology has its strengths and weaknesses. Each approach to data has its strengths and weaknesses. Each theoretical apparatus has its place in scholarship. And one of the biggest challenges in doing interdisciplinary work is being [able] to account for these differences, to know what approach works best for what question, to know what theories speak to what data and can be used in which ways.”
> [Danah Boyd, Senior Researcher, Microsoft; Research Assistant Professor, New York University [16]]

1. Introduction

With the rapid advances in technology, business interactions involving consumers and suppliers now generate vast amounts of information, which make it much easier to implement the kinds of data analytics that

* Corresponding author. Tel.: +82 2 2077 7907.
E-mail addresses: mrchang@smu.edu.sg (R.M. Chang), rkauffman@smu.edu.sg (R.J. Kauffman), yokwon@sm.ac.kr (Y. Kwon).
[1] Tel.: +65 6808 5227.
[2] Tel.: +65 6828 0929.

Gary Loveman, current CEO of Caesar's Entertainment, discussed in a 2003 *Harvard Business Review* article on data mining [70]. Today, this is referred to as the *big data revolution* in the popular press, and viewed as creating challenges and opportunities for business leaders and interdisciplinary researchers. The world's volume of data doubles every eighteen months, for example, and enterprise data are predicted to increase by about 650% over the next few years [45,54]. Today, most firms have more data than they can handle, and managers recognize the potential for value, but the promise of big data still has not been realized, according to the leading academic [35,78] and business media sources [38,79].[3] The potential arises from the use of data to support the way organizations operate and serve their stakeholders. A recent article in *MIT Sloan Management Review* [62] described the use of big data by an Atlanta-based public school, for example. High school graduation rates increased due to better-informed policy decisions that were based on the application of advanced analytics capabilities to student performance data. Likewise, organizations now are embedding analytics in their operations to support data-intensive strategies.

1.1. The emergence of big data

A recent McKinsey report has referred to *big data* as "data sets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" [71].[4] Such data come from everywhere: pictures and videos, online purchase records, and geolocation information from mobile phones. Big data are not just about sheer volume in terabytes though. Other important aspects have been emphasized in addition to *volume*, including *variety, velocity* and *value* [76]. Big data may be unstructured too: examples are text with social sentiments, audio and video, click streams, and website log files. Such data may flow in real-time streams for analysis, which can enable a firm to maximize business value by supporting business decisions in near to real-time. This new trend in decision support is evocative of what we saw in the 1990s with the emergence of data mining, and the new emphasis on data with a large number of dimensions and much higher complexity (e.g., spatial, multimedia, XML and Internet data). Most of the data sets were "one off" opportunities, rather than data that had become available due to systemic and technological advances.

Considerable challenges are present in the quest to capture the full potential of big data. The shortage of analytics and managerial talent is a significant and pressing problem, for example. *CIO Magazine* [72] and the Corporate Executive Board [79] have reported that it is difficult for firms to find the right people. The U.S. alone is reported to face a shortage of 140,000 to 190,000 people with deep analytical skills, as well as 1.5 million managers and analysts to make effective decisions [71]. (See Fig. 1.)

1.2. Toward computational social science

New perspectives in social science are now tracking the developments in big data. For example, *computational organization science* has broadened researchers' perspectives on social, organizational and policy systems, by adopting computational models that combine social science, computer science, and network science [22]. Other related developments have occurred, including the emergence of computational social science and e-social science [37,63]. *Computational social science* involves interdisciplinary fields that leverage capabilities to collect and analyze data with an unprecedented breadth, depth, and scale. Computational modeling approaches now can predict the behavior of sociotechnical systems, such as human interactions and mobility, that were previously not studied with one-time snapshots of data for very many people [83]. We see a *paradigm shift* in scientific research methods — and one that prompts new directions for research. A useful perspective in this context is attributable to Runkel and McGrath [75], who characterized research methodologies based on three goals: *generality*, *control* and *realism*. They distinguished between their *obtrusiveness* and *unobtrusiveness* for the subjects of research.

With emerging collection techniques for big data sets, there seem to be fundamental changes that are occurring related to research methods, and the ways they can be applied too [58]. In e-business, for example, the contexts include social networks, blogs, mobile telephony, and digital entertainment. The new approaches we see are based on more advantageous costs of data collection, and the new capabilities that researchers have to create research designs that were hard to implement before. The research contexts include human and managerial decision-making, consumer behavior, operational processes, and market interactions. The result is a change in our ability to leverage research methodology to achieve control and precision in measurement, while maintaining realism in application and generality in theory development.

We will discuss the causes of the paradigm shift, and explore what it means for decision support and IS research, and more broadly, for the social sciences. How can we take advantage of big data in our research? What new perspectives are needed? What will the new research practices look like? What kinds of scientific insights and business value can they deliver in comparison to past research? And what research directions are likely to be especially beneficial for the production of new knowledge?

Section 2 reviews traditional methods for research and discusses the key factors that are creating the basis for a paradigm shift. Section 3 describes the new paradigm in the era of big data, and how it relates to decision support, IS and social science research. Section 4 assesses how the research has been changing, through the use of a set of specific comparisons between research that was conducted before and after the emergence of new methods associated with big data. Section 5 offers some new research directions, and section 6 concludes.

2. How are big data supporting a research paradigm shift?

The move to computational social science in the presence of big data involves a Kuhnian *scientific paradigm shift* [60]. We will provide a background on the traditions of research inquiry, and then examine the driving forces for the paradigm shift, and why access to large stores of data is speeding the process.

2.1. Traditions of research inquiry

Churchman [27] characterized research with a set of different *inquiring systems*. They involve methods, procedures and techniques to describe and explain behavior, test hypotheses, assess causality, and

[3] A recent article in MIS Quarterly by Chen et al. [25] identified a total of 560 published articles in three areas in the IS literature, based on keywords such as business intelligence, business analytics and big data. They stated that 243 are big data-related, with 84% appearing only in 2007 or later. Also included is a table indicating that, among the major IS journals, Decision Support Systems has published the most business intelligence and business analytics articles (a total of 41), followed by Communications of the AIS (19), the Journal of Management Information Systems (12), Management Science (10), and Information Systems Research (9). This information suggests that Decision Support Systems is an important outlet for the kinds of research described in this article. The journal ranks first for publishing business intelligence and business analytics articles between 2000 and 2010. Electronic Commerce Research and Applications devoted a special issue of ten articles on "The Role of Business Analytics in E-Commerce in early 2012, and MIS Quarterly also published a "Business Intelligence" special issue in 2012, with five articles.

[4] According to Reuters, the oil company Chevron accumulates 2 terabytes of data a day from operations [48]. The Large Hadron Collider, the largest particle accelerator, generates 40 terabytes per second when it operates — about 15 petabytes annually. Online content providers generate large amounts of data too. Twitter creates more than 7 terabytes, while Facebook creates 10 terabytes, and Google produces 24 terabytes of data every day just from its search operations. Data set size for academic research is growing exponentially too. About 20% of *Science* authors now use data sets exceeding 100 gigabytes and 7% use 1 terabyte-plus data sets, for example [82].

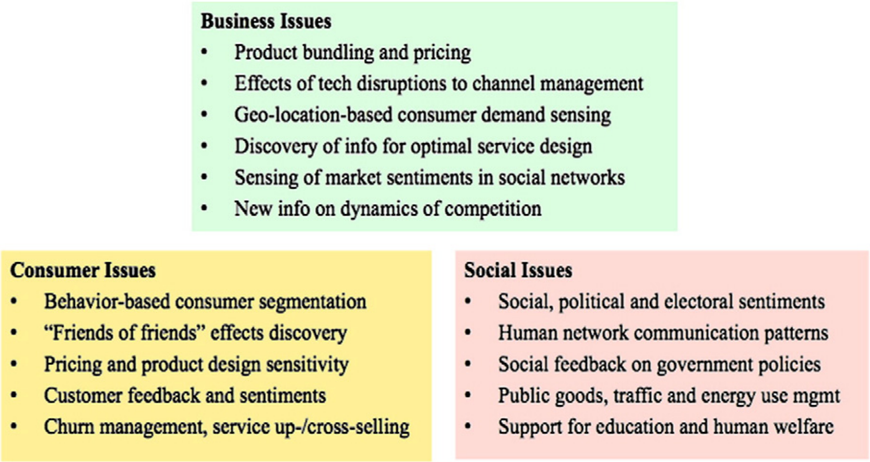

Business Issues
Product bundling and pricing
Effects of tech disruptions to channel management
Geo-location-based consumer demand sensing
Discovery of info for optimal service design
Sensing of market sentiments in social networks
New info on dynamics of competition
Consumer Issues
Behavior-based consumer segmentation
"Friends of friends" effects discovery
Pricing and product design sensitivity
Customer feedback and sentiments
Churn management, service up-/cross-selling
Social Issues
Social, political and electoral sentiments
Human network communication patterns
Social feedback on government policies
Public goods, traffic and energy use mgmt
Support for education and human welfare

Fig. 1. Business, consumer and social issues that big data can address.

establish new truths. Runkel and McGrath [75], in contrast, include theory-building, sample surveys, judgment tasks, lab experiments, field experiments and studies, and simulations.[5] (See Fig. 2.)

2.1.1. Runkel and McGrath's three-horned dilemma has diminished in intensity in the presence of big data

Today's methodologies are different in terms of how data are collected, the intrusiveness of the procedures, and the degree to which each methodology applies to situation-specific versus universally-observable behavior. The strengths they provide include achieving *generality*, *control* and *realism* in the research designs. Runkel and McGrath [75] argued that these strengths cannot be maximized simultaneously with any single research methodology though. Instead, choosing one method takes advantage of a key strength, while the leverage the others offer may be lost. This is a classic *three-horned dilemma* for research methodology. For example, consider the control of a lab experiment, versus the realism in a field study or a computer simulation of a process, versus formal theory-building uncontaminated by the complexities of the real world.

The cost and efficacy of using traditional research methodologies have been affected by the emergence of the Internet. New methods, such as online experiments from economics, online judgment tasks from cognitive science, online surveys from sociology, the capture of full repositories of business transaction data in IS, and data stream filters from computer science are being used today to answer research questions that were previously impractical or impossible to answer, due to the inaccessibility of the data. In addition, since it is relatively easier to establish subject pools to which multiple control and stimulus conditions can be applied in online social networks or in mobile phone systems, research designs can be implemented that allow researchers to find the data that meet a set of pre-defined experimental conditions – the "needles in the data haystack" – without ever taking subjects into a laboratory setting to conduct an experiment [21]. The new paradigm seems to provide a way of addressing, if not entirely resolving Runkel and McGrath's dilemma.

2.2. The driving forces of change in the paradigm shift

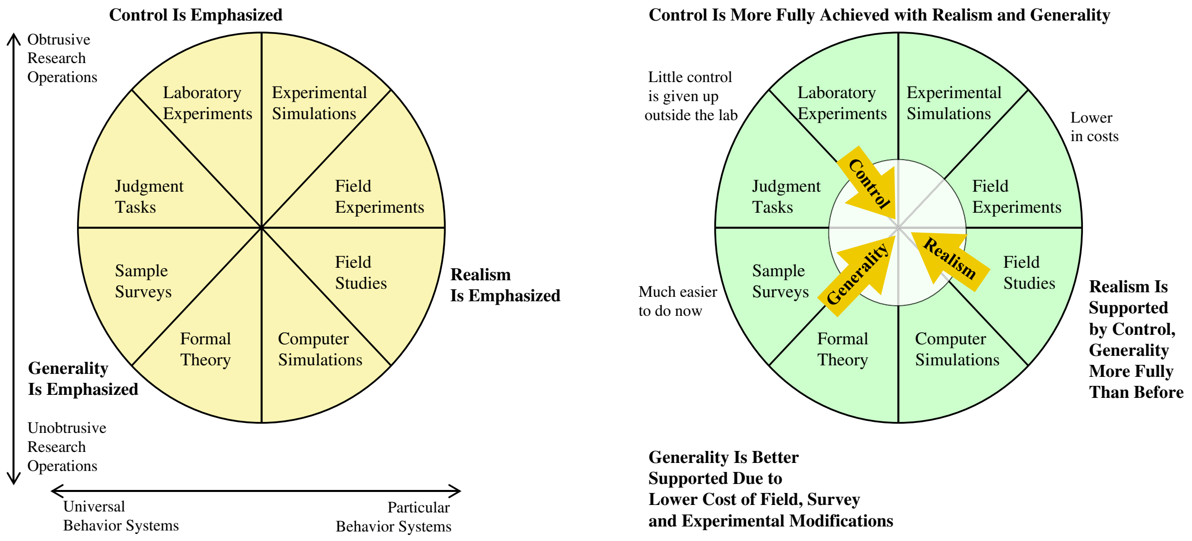
The paradigm shift for computational social science overall, and for decision support and IS research, is based on several distinct forces. They include technological changes, the convergence of disciplines, and the availability of new tools and solution approaches for data analytics. They cover new business practices driven by senior management perspectives on the value of data analytics in decision support, changes in market competition, and what's required for firms to differentiate themselves. (See Fig. 3.)

2.2.1. Technological change plays a role

IT has improved the efficiency and effectiveness of organizations since the original implementations of decision support systems (DSS) in the mid-1960s. Since then, new technologies have been developed, and new categories of information systems (IS) have emerged, including data warehouses, enterprise resourcing planning (ERP), and customer relationship management (CRM) systems. The technologies have evolved and become commoditized and less expensive, and have delivered continuous increases in storage and computing power. Moreover, the Internet revolution has had a drastic impact on business and society, including the rise of social communication through e-mail, instant messaging, voice over Internet protocol (VoIP) services, video conferencing, and other means.

Blogs and wikis have changed the patterns of communication between organizations and the communities that surround them too. They have affected how individuals communicate— including in their homes, in their companies at work, and on vacation. User-generated content has become a source of creative interpersonal interactions for people and valuable economic exchange. Analytics can be used to surface useful information from the raw data that these different settings and technologies generate, and support decision-making like never before. A representative example of the new architectures and technologies, and various tools required for creating business intelligence with big data is Hadoop. It is the most popular open source software framework that supports the use of very large data sets spread across distributed applications in larger systems [71]. Another example is NoSQL databases, which are used to store very large sets of data in a way that supports flexible manipulation of unstructured data. Large data stores often need massively parallel processing so queries can be made

[5] Friedman and Sunder [43, p. 3], in their 1990s book on *Experimental Economics: A Primer for Economists*, wrote: "Traditionally, observations from naturally occurring economic phenomena were the only source of data to stimulate revision of theory. If data relevant to an economic proposition could not be captured from naturally occurring conditions, then the proposition went without benefit of empirical refinement. In recent years, experimental methods have given economists access to new sources of data and have enlarged the set of economic propositions on which data can be brought to bear."


Control Is Emphasized
Obtrusive Research Operations
Laboratory Experiments
Experimental Simulations
Judgment Tasks
Field Experiments
Sample Surveys
Field Studies
Formal Theory
Computer Simulations
Realism Is Emphasized
Generality Is Emphasized
Unobtrusive Research Operations
Universal Behavior Systems
Particular Behavior Systems
Control Is More Fully Achieved with Realism and Generality
Little control is given up outside the lab
Lower in costs
Control
Generality
Realism
Much easier to do now
Realism Is Supported by Control, Generality More Fully Than Before
Generality Is Better Supported Due to Lower Cost of Field, Survey and Experimental Modifications

(a) Runkel and McGrath's three-horned dilemma for research methods

(b) The new paradigm's changes

Note. The three-horned dilemma of research methodology is: generality of findings across actors; control and precision of measurement; and the realism of the context. There are four setting types: naturaal (field experiments and field studies), contrived (lab experiments and experimental simulations) and unaltered settings (judgment tasks and sample surveys), and also settings in which no data collection occurs (computer simulation and formal theory).

Fig. 2. Traditional research methodologies and the new research paradigm with big data.

effectively across multiple machines simultaneously for faster processing [42]. They also require in-memory computational capabilities at the server level, so it's possible to do real-time analysis of streaming data. Major software vendors are already providing technical solutions, and advances in large-scale data analytics are tied to advances in the technologies, and will support innovations in managerial and organizational decision-making too.

2.2.2. Interdisciplinary convergence intensifies the changes

Yoffie [87] observed another aspect of the recent technological revolution: previously distinct technologies have converged to unify a variety of functions that are demanded by people in business and society. Yoffie called this *digital convergence*. Examples include wireless phones and cameras, TVs and hand-held devices, and database and data analytics tools. Accompanying these changes with technology, there is also a convergence in the study of leading research problems through interdisciplinary scientific interactions.

In some of the areas where *interdisciplinary convergence* is occurring, there seems to be a special need for new ways to deal with the emergence of big data. This motivates, for example, convergence among the disciplines of marketing, economics and computer science, where fresh insights on consumer behavior and product design provide a new basis for competitive advantage. Our key observation is this: the large data sets that describe the inner workings of complex social systems that researchers would like to investigate cannot be adequately understood from a single disciplinary perspective. For example, a huge amount of data is available for the study of group and individual behavior in sociology, but it cannot be analyzed without data mining and machine learning methods from computer science. The same can be said of the centrality of machine learning, econometrics and mechanism design for advances in keyword search advertising, where creating deep insights on ad slot valuation and positioning, and the effective prediction of consumer responses are central to business search advertising success.[6] Interdisciplinary convergence offers diverse motivation for new research questions and theoretical perspectives relevant in all of the phases of the research process – from problem and theory formulation – to research model and methods selection – all the way to data analysis and interpretation of the results.

2.2.3. Competitive strategy and the new data analytics support the change

Data analytics are a matter of competitive strategy for many businesses and organizations today [30]. More sophisticated analytics solutions are being developed to solve complex problems based on access to and capabilities for processing large amounts of data [76].[7] In particular, improvements in exploratory and predictive analytics capabilities have made it possible for research to be more data-driven, but in a way that is highly beneficial and deeply insightful. They no longer carry the "baggage of the past" — that such approaches are somehow anathema to

[6] An example is keyword auctions research, where major advances have been made in modeling to support improved mechanism design. For example, generalized second price position auctions have recently been identified as not being equivalent to the Vickrey–Clarke–Groves mechanism in the keyword auction context [6,40]. In addition, attention has been given to: controllable aspects of keyword auctions for keyword search mechanism design [26]; the application of historical information as an auction design aid [67]; strategic bidding behavior [39]; and refinement of Google and Yahoo!'s approach to user click-through weights on keyword-related ad positions to improve profitability [68]. We have also seen the emergence of computational methods for the estimation and recovery of keyword auction bidder valuation distributions, theory-based structural econometric model applications, and counterfactual experiments to estimate the impacts of different formats and modified competitive environments on keyword auctions [7]. Reiss and Wolak [74, p. 4281] have pointed out that when "there is little economic theory on which to build, the empiricist may instead prefer to use nonstructural or descriptive econometric models. … if there is a large body of relevant economic theory, then there may significant benefits to estimating a structural econometric model …"

[7] According to a 2011 market forecast by IDC [55], the main data analytics vendors included Oracle, SAP, IBM, Microsoft, SAS, Teradata, Informatica, Microstrategy, FICO and Adobe. IDC expects the business analytics market to grow at an 8.9% compound rate during the 2011 to 2015 period. It defines the market in terms of size, momentum, diversity, reliance, and its relationship with the broader business analytics software market.

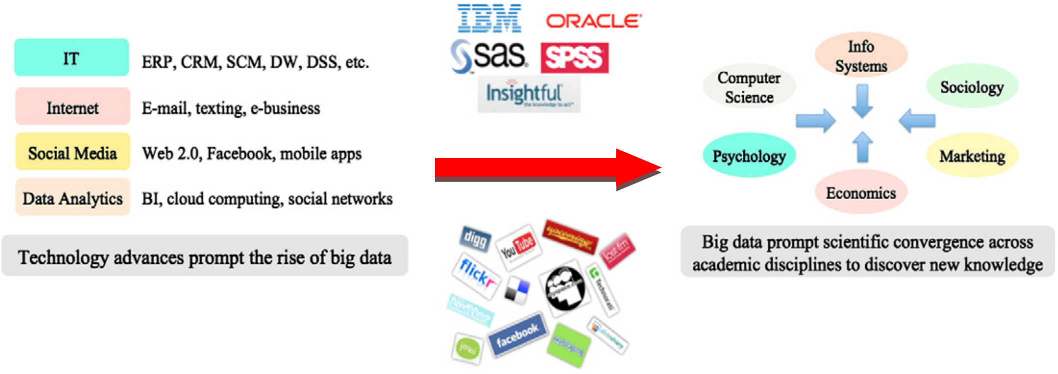

IT
ERP, CRM, SCM, DW, DSS, etc.
Internet
E-mail, texting, e-business
Social Media
Web 2.0, Facebook, mobile apps
Data Analytics
BI, cloud computing, social networks
Technology advances prompt the rise of big data
IBM
ORACLE
SAS
SPSS
Insightful
Info Systems
Computer Science
Sociology
Psychology
Marketing
Economics
Big data prompt scientific convergence across academic disciplines to discover new knowledge

Fig. 3. Forces creating pressure toward computational social science with big data.

good practice in social science. The current generation of data analytics methods provides capabilities for interpreting various types of streaming data, such as video and GPS data. They also permit managers to embed analytical capabilities in business processes, which can support monitoring and analysis of business processes and activities in near real-time. Another complementary method is data visualization, which offers managers a chance to make sense of large data sets. They can discover patterns and visual associations that stand out and provide information for decision support [85].

2.2.4. New business practices and organizational environments speed the process

Another driver affecting the rapidity of adoption of business data analytics in organizations is the creation of new business practices, with the idea that "technology is the service" [77]. As a result, more firms are redeveloping their organizational cultures to support new approaches to understanding their businesses through large data sets. This requires a firm's senior managers, employees, and its strategic partners to work together to understand the potential of business data analytics. This way, they will increase the likelihood of reaping the advantage that many authors, pundits and observers have suggested await the market leaders.

3. The new paradigm: computational social science with big data

We next discuss the details of the paradigm shift in research, as a by-product of the identified forces.

3.1. Definitions of key terms to describe the paradigm shift

We begin with a series of definitions that support the development of this discussion. (See Table 1.)

Today, we can offer a more in-depth description of the world, as it might be viewed from the perspective of computational social science with the capacity to collect large amounts of data to describe the variety of activities and events that are occurring within it. The real world is a *complex adaptive system* with a collection of entities and their interactions [23,80]. In such a system, the entities are continuously adapting to an ever-changing environment.[8] Within this dynamic system, a series of events that arise based on the actions and reactions of the entities occur in different contextual, spatial and temporal settings. These typically can be observed and quantified, though some aspects may be unobservable due to the limitations of the data capture methods. The events generate useful data that can be collected on the behavior of and interactions between entities that an analyst wishes to study. Figuring out what drives the events that are observed, and developing analytical approaches to reliably predict them are where new organizational competencies for decision-making and competitive advantage begin.[9]

3.2. The macro–meso–micro data spectrum

Thinking about data and research settings this way supports a fuller understanding of the real world as an *action space* for much more well-informed decision-makers to create value. It prompts the identification of empirical regularities and the patterns of people in the midst of characteristic behavior that presage decisions that occur on the value threshold of their willingness to act or transact. The main difficulty that many researchers have faced is the shortage of data describing the real world they wish to study. They have also had to deal with limitations with the application of the most effective research methods that aid in the discovery of new knowledge. Data for social science research now are more plentiful though, so the methods to use for key problems will offer dramatically more leverage for useful insights.

Advanced technologies now support researchers' efforts to acquire data to identify interesting empirical regularities in complex real-world systems to support high value decision support. The *data spectrum* available in the new decision support spans the largest scale to the smallest possible scale. (See Table 2.) One end of the data spectrum – the most aggregated scale or the *macro level* – involves tracking patterns of interactions between nations or economies for different sectors of industry over a period of time, or tracking inter-regional immigration. *Meso-level* societal-scale data can be used to track the behavior of individuals in a population of a place, such as the mobile communication records for a single service provider's customers in a city or society.[10] Another example of meso-level data that can be collected is from tracking the communications and behaviors of all users of a social network.

At the other end of the data spectrum, managers can craft new decision support capabilities from *micro-level* data related to a set of individuals' interactions with their relatives in shared phone plans, or the patterns of textual expression in blog posts over a period of time. Even more detailed *nano-level* or *atomic-level* data might arise in the

[8] An example is the emotional reactions of consumers to products they purchase or experience they gain in the consumption of various services. Although a lot of information can be captured about their transactions for decision support, current tools for gauging consumer reactions don't yet include the capability to collect affective data on their emotional responses. So the best that can be done is to collect data that proxy for the information associated with unobservable latent variables. Emerging research is exploring consumer image and video data for product-related emotional sensing in experimental settings.

[9] At the Living Analytics Research Centre of Singapore Management University, we refer to this information about the behavior, action, interaction and event trajectories of the entities as *digital traces*, especially when they apply to people in the roles of consumers, Twitter tweet creators, blog posters and other kinds of social network participants.

[10] An example related to data set size is the number of communications a person in a social network has with other people versus the number of additional communications that any single communication spawns. There are similarities in this insight to Georg Cantor's number theory, and which supports our intuition for gauging the size of countable and uncountable sets.

Table 1
Definitions of terms.

Terms	Definitions and examples
Entity	Key units of analysis or object of research interest, including: • *People*: decision-makers, users, consumers, voters, and social network participants • *Products*: digital music, movies, software, clothing, books, etc. • *Services*: online advertising, SMS services, airline ticketing, banking transactions • *Devices*: POS scanners, RFID tags, GPS locators, mobile phones, cable TV set-top boxes
Data	Attributes, features, actions, events identified to describe the entity in various kinds of settings: • *Contextual*: chatting online, buying products in a store, making actions in an online game • *Spatial*: consumption of a product at a physical location, use of software at home, banking transactions in a neighborhood in the city • *Temporal*: Things happening now or in the past, or at many points in time over a specified period, can extend into the future
Data Agency	Institutions that collect and store relevant data in different settings: individuals, consultants, firms, schools, IT services providers, network operators, non-profit organizations

Note. We distinguish between *data agency* and *data-collecting agents*, which use software to collect digital data [58].

Table 2
The micro–meso–macro data spectrum.

Type	Definition
Micro-data	The least aggregated level of data in very large data sets, resulting from the capture of technology-mediated human, social, machine and physical settings; in this category, *atomic data* reflect the least aggregated relevant level for data capture, such as tweets in Twitter analytics for an individual person, channel changes in digital set-top box viewership pattern tracking for a viewer, blog posts by individuals, sensed activities by digital sensors in the so-called "Internet of Things", or clickstreams in Internet session tracking for a user, etc.
Meso-data	The mid-level of data aggregation in very large data sets, resulting from the collection of data that are a level-up from micro-data in terms of the kinds of information that is captured. So, for example, while capturing the specifics of a user's tweets may constitute micro-data, examining the extent of their tweeting and re-tweeting behavior over time, and in response to specific issues or the tweets of other users will produce different kinds of information. The data are still very large in the sense of potentially being society-wide in their coverage.
Macro-data	The most aggregated level of data in large data sets describes: the market, regional or geographic area; the industry or economy sector; or the national and international levels. Macro-data on patterns of electricity use among the cities and towns of a country, or of the long-distance phone calling patterns of people in different regional area codes, and so on are aggregated at that level.

assessment of human emotional reactions based on physiological proxies when affective computing assessments are used.[11] These are essentially indicative of the lowest level, most disaggregated data in which measurement is relevant for a given analysis setting. In neuroscience and physioanalytics, for example, data collection approaches include magnetic resonance imaging and brain scans for human decision-making under uncertainty [34], and the measurement of affective responses in new survey research designs [84]. The data sets are likely to be of similar orders of magnitude, even though they exhibit different degrees of granularity related to phenomena that researchers are studying.

We also consider *attribute relevance* and *attribute completeness* to describe the entities in a research setting. The typical attitudes that we see expressed among managers and researchers are that "more attributes are better," but also that "capturing much more information becomes prohibitively costly." These attitudes may or may not be tied in with business value, but current data acquisition methods will make the costs of not getting this right much less expensive. With greater access to data and the attributes of the relevant entities, researchers will be able to leverage *data stratification* cheaply for multidimensional analysis, beyond the standard data hyper-cubes in online analytic processing (OLAP) in data warehousing.

Another characteristic of data for computational social science research is whether it comes in stocks or flows. A *data stock* usually can be acquired as a data set from an existing database, and it may be updated over different periods of time. Quarterly corporate earnings performance, or retirement investment portfolio values changes year to year are good examples. When the updating is continuous and occurs in real-time or near to real-time it is a *data flow*. Clickstreams from a two-way cable TV set-top box that tracks viewer use of the remote control handset, and sequences of advertising views when an Internet user is conducting a search are examples. Choosing which approach – data stocks or data flows – for acquiring data gives researchers the agility to study complex real-world systems in thoughtful ways that can be tailored more carefully to match the critical desiderata for effective decision-making.

3.3. The interdisciplinary research approach

When conducting empirical research with data available now from online auctions, social network communications, mobile phone networks, gaming platforms, search engines, and blogs, researchers need to be more open to exploration from the viewpoints of multiple disciplines. Extracting meaningful information from a large data set may require methodological and interdisciplinary flexibility to create business intelligence for decision support from it. Someone who has an expertise in sociology and psychology may be unaware of the insight-surfacing opportunities that exist to work with the same data using data mining and machine learning techniques, for example. Different knowledge from statistics and economics will be useful too. The data may no longer exhibit the small sample properties that many people in the social sciences have been trained to work with. Also user behavior in social networks may be correlated, which requires new methods to make sense of the causal relationships that are present in the observed stimulus–response loops. Thus, it is appropriate to embrace *methodology-focused collaboration* to develop an interdisciplinary, multi-perspective approach that will yield the most interesting results for the new decision support and e-commerce issues, and other social science problems of interest.[12]

Another consideration involves the kinds of theoretical explanations of interest. Extensive data sets that cover consumer demographics, product and service usage, and managerial decisions to redesign product bundles [8,9,41], for example, are open to theoretical and modeling assessment from the perspective of multiple disciplines. The impetus might be to understand the behavioral economics of consumer reactions to service prices or market share competition among firms. Or it

[11] There are many contexts that produce micro-data. An example in mobile phone use is the checks that service providers make by pinging mobile phones to identify their geolocations, or by tracking the changing locations of an individual's mobile phone as she moves through a shopping mall with femtocell sensors that can triangulate location and speed of movement to predict a new location several seconds later. In cable TV services, micro-data arise through tracking a consumer's channel-changing behavior via a two-way set-top box. In package shipments, micro-data are created on their changing spatial locations over time.

[12] We caution readers not to confuse our view on multiple perspectives in research with how it is treated by others. For example, *critical realism* and the boundary between the natural and the social world can be studied with a blend of quantitative and qualitative methods [5]. *Multimethodologies* point out the strengths, weaknesses and use cases for the qualitative and quantitative methods [73]. These perspectives, associated with the socio-technical tradition of research, emphasize research as a process to establish the *truth* about different technology-related phenomena of the real and social worlds, in which mixed quantitative and qualitative approaches are appropriate. Such approaches can be used to address why people in groups adopt mobile phone texting under different circumstances, or how blogs encourage people to express themselves with extreme forms of communication outside the norms of courteous and considerate interpersonal exchange.

may come from trying to discover the efficacy of new product designs intended to capture the new revenues from customers without promoting cross-product revenue cannibalization. To bring these diverse concerns together requires *theory-focused collaboration* involving expertise in marketing, psychology, design and economics too.[13]

A third reason is the problems themselves and the *measurement innovation collaborations* that they motivate. The study of organizational and human communications and interactions in social networks is a case in point — apart from any application area that the communications relate to (seller reputations, service quality, etc.). Relevant theory comes from sociology, communications, cognitive science and social psychology. But how the connections among participants are formed is just as much a matter of network economics, graph-theoretic operations research, computer science and accounting. So much of what is happening can be described by making relevant measurements in settings that can be abstracted as graphs, or understood through new metrics dashboards.[14]

But what measurements? And with what basis in theory? In social network-related research, we can distinguish among many different things that need to be measured so they are understood better. These include the strength of human reactions and communications, the position and distances between the entities in a network, the costs of communicating, changes in social sentiments, changing interactions among corporations and people in the marketplace, and so on. For every different measurement problem, there may be different disciplines of measurement knowledge to leverage. So there needs to be an effort to bring together people whose capabilities complement one another, and can bring new measurement approaches to support the creation of surprising discoveries and deep new insights.

3.4. Research methodology

We next will discuss how the new paradigm has been affecting the three-horned dilemma of research methods. We also will link our experience to the current era with what occurred in the 1980s and 1990s, as computers and data analytics were beginning to mature. The idea is that each method has a different strength in terms of achieving the three research goals of generality, control, and realism. So it isn't possible in principle to maximize the power of all three simultaneously: tradeoffs are involved.

But to what extent is this still true? The new approaches suggest we can now study similar phenomena to what was studied in the past, and new phenomena that have not been fully described in published research, with much greater control than before in more realistic settings. We also can do this without creating any disturbances in the processes under observation. Interdisciplinary big data-focused research methods are making it possible to address all of the dimensions in a simultaneously more effective way.

3.4.1. Big data blur the distinction between field experiments and laboratory experiments

With the emerging capabilities to collect data sets from a variety of real world contexts, the field itself has become the researcher's new behavioral research lab. We no longer have to architect artificial settings in the lab to affect control or forgo realism by mimicking the real world in a synthesized setting. Instead, we can capture information that represents the fundamental elements of human actions and interactions – clickstreams, tweets, user opinions, auction bids, consumer choices, and social network exchanges – directly from the real world as *digital traces* of human behavior. By using appropriate instrumentation for data collection, we can take advantage of situations that are naturally experimental or create randomized treatments to distill the conditions under which certain behaviors and unique outcomes are observed.

3.4.2. Societal-scale data support greater realism in survey and experimental research

Runkel and McGrath [75] suggested that judgment tasks and sample surveys support higher generalizability by decontextualizing real-world settings. This goal can be accomplished through the use of big data. Survey data often involve subjective responses, and they will complement what we can learn from objective factual data that are available in empirical research designs. What we cannot learn from objective data we may be able to gain insights on with participants' subjective responses. Further, we are no longer restricted by the number of subjects in surveys and experiments due to the widespread use of online social media, mobile phones, e-markets, and other new contexts. We can include many more participants and population-level data collection, which support greater generality through decontextualization.

3.4.3. The new methods affect how we can work best with theory

Social science research has emphasized development of theoretical models and testable hypotheses. Research with data at terabyte or petabyte scale challenges how we work with assumptions and theories. In a 2008 article in *Wired* magazine, Anderson [1] boldly suggested that big data might lead to the “death of theory.” Although his statement is clearly hyperbole and begs a reality check, consider this: Google has mostly depended only on numbers and applied mathematics in matching ads to user search contents. Its Adwords has supported a remarkable level of profitability that doesn't require a full understanding of the underlying theory.

Anderson's argument seems to be that we no longer have to build theoretical models, since no semantic relationships or causal analyses are required any longer. Instead, research methodologies in the new paradigm can put more emphasis on better data and better analytical tools, rather than better and more elaborate theories. We think this is similar to technical analysis and fundamental analysis in the financial market context. The former looks for patterns to establish price trends to predict valuation; the latter seeks underlying structure and causes to provide explanations for why some level of valuation is appropriate. They work hand in hand to help financial market investors gain more knowledge of their world.

Theory, in our view, still ought to remain central in the development of even more robust and interesting research insights. One possible way to create theory more effectively is by iterating between the data discovery and theory-building approaches. We think of this as a sort of *grounded theory* development approach, only it is supported by big data, and it involves the use of predictive analytics [74].

3.4.4. Greater control of research designs is no longer intrusive

Traditional lab experiments attempt to implement research designs that can eliminate noise and establish causality to achieve for more meaningful results. In many cases, the experiments are designed to have subjects choose one of two possible alternates, or to ask them to provide information on their preferences. This intrusiveness can lead to lower precision in a controlled experiment, as exemplified by the

[13] Consumer preferences and market shares for specific products and services have been difficult to study in empirical research, though there are many theoretical and empirical approaches that have been developed for these purposes. This is because consumer choice models [53,69], market share attraction models [29], and cross-overs between the two [61] require a lot of data to map out the range of consumer preferences and the set of firms that are competing for shares in the marketplace.

[14] A current example of a research effort focuses on *strategic public information and communication enhancement* (SPICE) and conceptualizes a new suite of metrics for social media-based communication connecting government and public organizations to residents in a densely populous urban environment. The innovations leverage measurement approaches from financial accounting and cost accounting. For example, social sentiments can be measured in terms of: stocks and flows of social sentiments; the acceleration or deceleration of changing sentiments; and the random and co-integrated temporal drift of different kinds of social sentiments. Other measurable aspects include the identification of baseline social sentiments, the variance in the changes from baseline that are observed, and the intertemporal probabilities of swings in social sentiment in response to market and social events that act as disruptors to how people feel about various issues. This joint research is being conducted by Singapore's Agency for Science, Technology and Research (A*STAR), Northwestern University and Singapore Management University.

effect of forced choices in most academic and industry studies of consumer preferences and decision-making [33]. The computational social science paradigm with big data can diminish the intrusiveness by using a subset of the data that are available. For example, researchers are now able to understand subjects' preferences from the digital traces of their behavior without directly asking them questions. Thus, an important consideration in research design is choosing which data to use in view of changing contextual conditions. The new approaches support greater control for the experimenter and greater precision in data collection, and thus will be able to produce much richer and valuable capabilities for decision support.

The reader should not falsely conclude, however, that big data are a panacea for resolving all aspects of the three-horned dilemma. There is a lot of noise in real-world environments that may make it difficult to work with very large data sets. Researchers often point out that being successful involves figuring out some “tricks” to work with big data also. We also are not indemnified against modeling and research design errors. These things require more attention, and we will discuss them more fully later in this article.

3.5. Strategies for observation

A unique aspect of the new paradigm is that researchers now have more control over the *timing of observations* they can leverage in social science. Traditionally, one round of a survey might cost between US $5000 and US$25,000 to reach a subject pool of sufficient size for a given research design. Much traditional research in social science has been cross-sectional in its coverage of a unit of analysis at a single point in time, single units of analysis with a sequence of observations over time, and many targeted subjects or entities at a given level of analysis in longitudinal observational designs.

3.5.1. More frequent data collection is possible

The real world often creates situation in which the researcher is able to take advantage of changing conditions: the acceleration of online game competition, electronic market demand changes, new service delivery mechanisms in mobile telephony, and so on. As a result, it may be appropriate to collect data frequently, including following up with online survey response requests after every customer transaction is made with a firm.

Empirical studies for which it was only possible to capture data on a one-time basis at a high cost can now be redesigned so it's possible to harvest the same kinds of data repeatedly – and even quite frequently – with greater fidelity and richness at a fraction of the prior cost. This opens up the study of a range of geospatial and spatio-temporal contexts that involve the spatial movement of consumers in stores, medical patients in their homes and workplaces, delivery vehicles and public transit buses, and so on.

3.5.2. Appropriate data granularity must be defined by the research

Depending on the questions, a macro-, meso- or micro-level of *data granularity* may be appropriate. In most cases, the lowest level – micro-level or atomic-level data – may be desired to create the highest resolution for understanding unique contextual behavior in a research setting. This kind of data may represent a *behavioral time-series of human actions*, a more precise way of describing their digital traces. In Internet surfing, for example, clickstreams provide evidence for what a person is doing online; and in TV viewing, detailed viewership records of specific channels with timestamps can be collected. Micro-data can be aggregated, and this will support the study of higher-level issues too.

3.5.3. Iterative research designs create new power

Research designs include one-time to daily or even hour-by-hour experiments in environments where the patterns of human micro-behavior are relevant. This requires a pre-specified experimental approach in a natural context for the study of the phenomenon of interest — the technological, individual, organizational, network, systemic or market context. A researcher can perform iterative collection of data in chosen time intervals, and then apply different metrics over time. The process can be further refined in fresh ways to take advantage of the changing empirical regularities or trends that are observed in the system that has produced the data. It also creates the opportunity to conduct different kinds of sensitivity analysis through sequential variation in controlled experiments, as well as repeating, cyclical and even fractal patterns that are not easily discerned in other ways.

Overall, these strategies for observation support pre-theoretical analysis and explanation of the empirical regularities that arise for different phenomena in different settings. These span theory and model-based hypothesis testing in experiments, to secondary data via empirical designs, to the use of online surveys, simulations and assessments embedded in real-world product and service settings.

4. A comparison of examples of traditional and new paradigm research

A telltale indicator of the changes that are occurring is when we can identify research that introduces fresh research questions associated with longstanding problems and issues that can be studied in ways that were not possible before, and with new depth and insight from the findings. We next explore three representative areas of research that now involve the use of big data and analytics for business, consumer and social insights: Internet-based selling and pricing; social media and social networks; and electronic market and auction mechanisms. We will make comparisons between two representative studies and some related works for each. For the representative studies, one uses more traditional research approaches and does not use big data-driven designs versus another that reflects more current research that fits into the computational social science paradigm. In addition to considering the contracts between the business problems and research questions asked, we also consider the differences with the data and research methods, and the new kinds of findings that result. (See Table 3.)

4.1. Internet-based selling and pricing

We have seen a lot of interest expressed by IS researchers and their interdisciplinary collaborators in the area of Internet-based selling and pricing [66]. For example, research on the airline industry [24,25] has evolved because more detailed data at various levels of analysis have been able to be obtained through the Internet, supplementing traditional sources. One such source is the U.S. Department of Transportation's OD1A 10% sample of all airline tickets written for all of the origin–destination city-pairs in the United States. To illustrate the contrast between more traditional but still sophisticated research approaches and current work, we consider Duliba et al. [36], who used an extensive but not historically large data set before web-based online airline booking systems became available. We compare it with the work of Granados et al. [48], who used airline ticket booking micro-data for the online and offline channels.

The scope of research problems that can be addressed in the presence of newly available big data has expanded to include issues and questions that were not able to be studied with data available in the 1980s and 1990s. Of interest then, for example, was to estimate the business value benefits of computerized reservation systems (CRS) and their effects on air carrier market shares at the city-pair level and revenue and load factor performance at the aggregate national level. More recently, the newly-available data have permitted analysis of much more detailed price elasticity dynamics across the online and offline channels, blending the multiple market levels of analysis (city-pair and regional comparisons), and the impacts of different product attributes, especially price premium sensitivity for service upgrades [24,28,49,50]. In addition to the OD1A Department of Transportation data used by Duliba et al. [36] in the earlier research, the authors were able to gain access to proprietary data on the locations and density of

Table 3
Contrasts among research questions, data analyzed, and methods applied with the shift to big data.

Research	Questions	Data	Methods	Contrasts
• Internet-based selling and pricing				
Duliba et al., *Organization Science*, 2001 [36]	What are the benefits of com-puterized reservations systems (CRS) ownership in airline industry? Can we determine the city-pair origin–destination market share and national level of airline performance impacts?	4 years for 72 city-pairs and 12 years for national level. Travel agencies with CRS, departing flights, advertising expenses, revenue passenger miles, market share, airline fares. Acquired from U.S. Dept. Transportation.	Measured airline performance at city-pair and national levels; multinomial logit market share model for city-pair analysis; industry model for national level analysis. Focused on logical causality with relevant tests for robustness.	Research qs. Current work considers more refined and detailed questions that were not able to be studied with the data in the 1980s and 1990s; examination of channel effects supports specification of precise hypotheses, for which new theoretical implications can be derived. Data. Micro-data with details of airline passenger sales data can be used to estimate price elasticites in offline, online channels, under different intermediary-led information transparency conditions. Proxies. Direct measures and effect estimation possible using econometric methods via better proxy measures.
Granados et al., *Information Systems Research*, 2012 [51]	What are the differences in price elasticity of demand between the online and offline channels in airline competition? What are the implications for pricing? Multi-channel and transparency strategy with IT?	2.21 million records on airline ticket sales in the online and offline channels; access to a global distribution system via a leading airline; enabled coverage of multiple airlines and online travel agents.	Measured airline performance at city-pair and national levels; multinomial logit market share model for city-pair analysis; industry model for national level analysis. Focused on logical causality with relevant tests for robustness.	
• Social media and social networks				
Brown and Reingen, *Journal of Consumer Research*, 1987 [17]	What are roles of tie strength, homophily in macro (across groups) or micro (in small groups) word-of-mouth (WOM) processes? Do stronger ties in groups support brand referrals?	Traced who-told-whom social networks for 3 piano teachers, 67 piano students, and 118 participants in the social network. Involved the capture of WOM brand referrals.	Interview data rising from verbal reports; two-phase methods involving cross-check on interview results; relational analysis of WOM, hypothesis tests, including t and χ^2 tests, regression analysis.	Research qs. Social influence now can be assessed by designing viral features into social networks; no longer necessa-ry to ask what users perceive about word-of-mouth through interviews; data support direct observation. Data. Individual users' behaviors can be directly observed, indicating whether they adopted a product or service because of peer influence. Interpretation of empirical results. Drivers of peer influence can be distinguished from correlates with peer influence in social networks now.
Aral and Walker, *Management Science*, 2011 [3]	Can we separate correlation from causation to measure peer influence in social networks? Can firms add viral features to products to facilitate WOM effects? Peer influence?	Observation of user behavior, access to peer influence in product adoption via Facebook; 1.4 million friends of 9,678 Facebook users as social network experimental subjects.	Experiment analysis with randomized trials and randomized treatments, inside-outside experimental design, and robustness checks; event history analysis, hazard modeling, before- and-after event analysis.	
• Electronic market and auction mechanisms				
Hausman, *American Economic Review*, 1981 [52]	Can we use observed market demand to derive the compensated demand curve, through which actual consumer demand can be measured? What is the level of consumer surplus in traditional markets?	Market demand data can be used to predict the unobserved compensated demand curve; validation is based on using the two-good case; extension is to the many-good case with one price change	Quasi-indirect utility and quasi-expenditure functions; specification of utility function after the observation of market demand curve; the shortcomings of Marshallian demand curve approximation	Research qs. The level of consumer surplus can be now measured through field experimentation, in contrast to the dominant approach of analytical modeling in prior research. Data. Detailed data for online auction activities can be collected using bidding agent software,
Bapna et al., *Information Systems Research*, 2008 [12]	What is the level of consumer surplus generated in eBay auctions? How can consumer surplus be estimated in this important context? How robust will the empirical results be based on the estimation procedures used, and the nature of its underlying assumptions?	Sample of 4,514 eBay auctions in 18 major categories from 2004 and 2006: bid price, willingness-to-pay, bid placement time, number bidders; automated data collection using software agent.	Field experiments with randomly- drawn validation sample; comparison between experimental and field data; mean value for the total consumer surplus estimate; linear regression for robustness check, for similarity check between the 2004 and 2005 data sets.	

CRS terminals in travel agencies in city-pairs. In contrast, Granados et al. [48] gained direct access to an airline industry global distribution system (GDS) that brought together information on ticket bookings from multiple airlines. This allowed them to collect millions of data points from the online and offline channels.

Another contrast between the studies involved the data analysis methods used. Duliba et al. [36] applied a multinomial logit market share model for city-pair analysis and an industry production model for the national level analysis. They showed that the presence of agency CRS terminals was associated with greater airline city-pair market share and also with higher levels of aggregate airline performance. The authors used relatively indirect measures to assess the impact of IT investments on the airlines' corporate performance and produce meaningful results. Today, it is possible to employ better proxy variables and more direct measures of the effects that are central to the research inquiry through the use of big GDS data. Granados et al. [49–51] used sales data to assess the effects on price elasticity of demand in the presence of digital intermediaries and intensifying city-pair route competition, and how the effects varied across the online versus offline markets. The authors used a natural experiment in which it was possible to do matched assessments of differently intermediated competitive settings. To deal with endogeneity issues arising from city-pair route price-setting and the associated service production costs that are private for most competitive firms, the authors were able to construct instrumental variables in their econometric analysis from publicly-available data sources.

The argument that we made earlier regarding the impacts of the new paradigm for computational social science research is borne out in the airline industry price elasticity of demand research: the strength of Runkel and McGrath's [75] three-horned dilemma seems to have diminished. Our capability to capture so much data with so little intrusion on operations has enhanced the efficacy of the research methods that can be used to come closer to the truth without the traditional burden of the trade-offs in power.

4.2. Social media and social networks

The burgeoning research on social media and social network during the past five years reflects some of the most exciting new impacts of IT on relationships among people in society, as well as significant innovations for how new methodologies and interdisciplinary perspectives can be used [4,31,32,44,46,47]. In the time before the emergence of social networking services and web-based interactions became popular, it was not easy to directly measure the strength of the relationship between two individuals. Yet a related research question that has been of interest to marketers and sociologists over the years is: "How do strong ties or homophily (similarities between people) impact word-of-mouth processes?" A traditional approach to answering this question, demonstrated in the research of Brown and Reingen [17], is to interview individuals in a network. In contrast, as shown by Aral and Walker [3], observation of the behavior of individuals in an online social network such as Facebook now allows the above question to be answered without directly asking individuals. New data analytics methods also have created the capability to separate correlation from causation to measure peer influence in social networks much better.

While Brown and Reingen [17] performed hypothesis tests using two-phase methods involving cross-checks on their interview results for 118 participants in a social network, Aral and Walker [3] used a randomized field experiment with 9167 experimental users and 1.4 million friends on Facebook.[15] They also assessed social network effects when viral elements were embedded into product marketing campaigns on the Internet, and found that passive broadcast viral features caused a greater increase in peer influence and social contagion than active personalized viral features.

Again, we can see the changes in the research related to word-of-mouth effects made possible by the emergence of social networking systems. We now have the available means for direct observations of how users behave in social networks, the capacity to attack new research questions with new interdisciplinary research designs and approaches, such as graph-theoretic models and the analysis of events, to understand hidden user behavior [13,88,89].

4.3. Electronic market and auction mechanisms

Research on markets and auctions has received considerable attention from researchers over the years. Economists have modeled and estimated consumer demand and surplus in traditional markets with mathematical models and econometrics analysis. New kinds of data on the behavior of e-market participants are now more visible on the Internet, and able to be collected through the use of screen-scraping and agent-based data collection approaches [10,11,56,57], which has opened up many new opportunities for the study of electronic markets and auctions mechanisms. Sites such as uBid and eBay have been the most popular and widely used, and consequently the most studied with computational social science methods.

An example of research that estimated the level of consumer surplus in traditional markets is by Hausman [52], who applied quasi-indirect utility and quasi-expenditure functions, and used market demand data to parameterize the unobserved compensated demand curve. In the more current eBay context, Bapna et al. [12] asked a different, but related question: How much consumer surplus can eBay auction participants extract per transaction and in aggregate? They also assessed the extent to which the estimated auction surplus might be skewed, due to issues with the data and the assumptions applied to it. Their research involved 4,514 eBay auctions from 2004, and used experimental controls for their data selection. Their analysis suggested a conservative estimate of average surplus per auction was on the order of US$4, and more than US$7 billion in aggregate consumer surplus. They also were able to check the robustness of the results against 2005 data, an extraordinary step made possible by the changed costs of data collection. This shows how big data research design supports more refined results and insights.

Similar to the other research examples we have discussed, research on e-markets and online auctions now can take advantage of Internet-based and other sources for large data sets. From such sources, researchers can bring together representative sets of observations to identify and probe different kinds of "needle-in-the-haystack" conditions that can be analyzed to produce significant statistical results and support a more complete understanding with more meaningful insights than were possible before.

5. Research guidelines and practical considerations

We next discuss several new directions for research that have become possible.

5.1. Research directions with the new paradigm

Contextual information is a key aspect of the new kinds of information available due to the big data revolution. We propose:

- **Research Direction 1. Explore the impacts of knowledge about contextual awareness.** *Newly available contextual information can be collected inexpensively, and supports the analysis of new levels of contextual awareness on the part of consumers in many commercial settings. New research questions can be explored that have not been considered*

[15] The literature to date has mostly focused on the application of *network structure randomization* and *treatment randomization designs*. The former involves experimental manipulation of a social network's structure, and where an agent is assigned to participate in it. The latter focuses on manipulation of other aspects of the treatments that different participants receive in a social network. For this, the participants are already situated in specific locations in the network.

before, which creates the potential for new knowledge in the business, consumer and social insights arena.

Contextual information provides important clues to understanding why different outcomes occur in different settings. For example, it may be relevant to explore a spatio-temporal setting (e.g., a consumer leaving a coffee shop in a mall) or a sequence of settings over time (e.g., mobile location information to show the shops a consumer visited in a shopping mall on Saturday morning versus Wednesday afternoon). These provide especially rich settings in which to discover consumer insights that have never been identified before. Traditionally, seasonal demand effects offered useful contextual information in aggregate. Today though, information can be collected at the level of the individual (e.g., customers visiting an online shop or a bricks-and-mortar seller). We also can study consumers' instantaneous reactions to time and location-specific purchase incentives, or responses to tweets in a social network context involving political topics, candidates and elections, or user responses to instant deals in group-buying. New research can focus on contextual factors and triggers that create consumer decision responses.

To exploit data on contextual awareness, researchers should consider how such information can support *extreme personalization* or *hyperdifferentiation* in product and service relationships. We propose:

- **Research Direction 2. Reassess the business value of personalization and extend the study to consumer and firm co-informedness and hyperdifferentiation.** *The extraordinary amount of data that can be captured on consumers today creates dramatic changes in the basis for value maximization from the consumer's and the vendor's standpoints. Innovative experimental approaches create new means for vendors to tease out the new dimensions for product and service designs, and new product and service bundling designs that are likely to emerge as a result. Affective computing-based data collection, decision monitoring, location-based data, and sequential response data can support new kinds of research in these areas.*

The era of terabytes and petabytes of data also affords inexpensive collection of micro-level data to yield insights on human behavior. One outcome is that individual-level information has the potential to transform the quantitative study of consumer decision-making, which will offer new directions for software and machine-based decision support. We expect to see new and more in-depth research on pre-adoption and post-adoption usage behavior for all kinds of things — social media, software-based decision aids, websites, data storage services, and so on. Researchers will be able to move beyond *purchase intention* to *purchase observation* research in the presence of sales patterns for hyperdifferentiated products and services, and produce more valuable research findings in the process. New levels of consumer informedness are likely to affect consumer web surfing behavior and the action sequences that lead to purchases. Recent approaches involving neuroimaging and affective computing offer other new ways to explore individual diversity and the empirical regularities of emotional reactions through big data [34,82].

A third aspect is related to data that are produced when events occur in various settings. Some examples are: transactions in electronic markets; exchanges in social and mobile phone networks; first consumer contacts with a company's website; viewing initiation in digital entertainment settings; and so on. We suggest the following research direction on how big data can support this kind of research:

- **Research Direction 3. Pursue big data studies that focus on patterns, explanations and predictions of events in business, consumer and social insight settings.** *New means for digital sensing of events in various settings are available through technological advances that can produce behavioral and economic information at the individual, group, network, firm and societal levels. Event-focused analysis opens up the possibility of the application of new research methods to discover knowledge that may otherwise not be acquired also.*

Events in different business, consumer and social settings are indicative of critical transitions that are often worthwhile to study closely. For example, these occur when a consumer's online search results in a purchase transaction, or when the occurrence of social events lead to changes in social sentiments, or when pricing changes result in the changed sales performance of a firm. Decisions have been made. The study of events that have occurred in the past or are likely to happen motivates natural and controlled experiments. We also can obtain an understanding of where events occur; for example: where in a social network; where in a physical network of stores; where in the geospatial patterns of people's usage of mobile phones; or where a Twitter user originates a message to her followers. These kinds of observations open up the use of methods for analysis that involve interdisciplinary knowledge. These include: event history analysis to understand the genesis, likelihood and timing of observable events [15,59]; spatial econometrics and data mining to assess how space interacts with time in the daily patterns of consumer purchase behavior [2,14]; and count data modeling and stochastic counting processes to bring greater analytical sophistication to our assessment of the simplest elements of observable phenomena in data, such as the count, frequency and variation in density of events [20,86].

- **Research Direction 4. Explore market and societal-level impacts with new data-sensing and social media capabilities.** *New data collection and data-sensing mechanisms in market and social media contexts open up opportunities for conducting research to observe and understand the impacts of human behavior in more aggregate market and societal settings. Examples are group-buying and social commerce, electronic auctions and digital markets, credit card transaction volumes, transportation and traffic flows, and social sentiments.*

During the past few years, there has been much research conducted using social media websites, especially Twitter and Facebook, but also job posting websites, such as Monster, LinkedIn, Indeed, CareerBuilder, and SimplyHired. From these kinds of websites, there are many opportunities to collect data that permit the assessment of more aggregate level issues in society and in the economy. The social media sites report on social sentiments among people in online communities, as well as their preferences, choices and social relationships. The job posting sites provide social clues on the state of labor and unemployment, and reflect aggregate-level information in terms of secular shifts in demand for a variety of skills and human capital in the marketplace. In addition, it may be possible to glean information from such sources as credit card services providers, phone services vendors, and digital entertainment providers on basic societal-level information, including people's spending behavior, creditworthiness, and sensitivity to changes in the economy. Research on aggregate behavior – the market and society levels – will provide a new basis with societal-scale data sets to identify how to sense the need for and improve social services, such as municipal rapid transit, highway traffic problems, and the quality of urban housing and public services. Governments will play an important role in fostering the disclosure of large data sets in the pursuit of open innovation, and as a means to improve social welfare for the general public.

Underlying all of the other research directions that we have mentioned is the issue of information privacy, which has become more critical, as people suffer from the negative impacts of clickstream data mining, mobile phone location tracking, social network connection observation, and the general loss of privacy associated with digital life. We propose:

- **Research Direction 5. Conduct complementary research on information privacy issues related to big data sets in parallel with investigations of business, consumer and social insights.** *The rise of big data and the new vulnerabilities that it creates prompt the exploration of information privacy issues that affect people as individuals, in groups, in relation to private and public organizations, and with respect to government and society as a whole.*

There are multiple directions that such research can take. Important research, for example, can be done on the creation of a new codex for appropriate processes in research projects that involve the capture of extensive data on individual behavior, communication, transactions and other kinds of social interactions. As it stands today, in many research settings that we are familiar with, the individual – as a social network participant, or a consumer, or in some other role – is vulnerable to the capture of data that she doesn't know is being targeted. Sometimes this is occurring with the best of intentions. For example, a credit card company tracks a consumer's purchases to identify her merchants and shopping patterns so as to be able to shut down her account in an instant if there are hints that fraudulent transactions are being made. University-based institutional review boards for research with human subjects still have not come to grips with the technological complexities and implications of data collection from mobile phones, social networks, and other online settings in which micro-level data can be obtained though.

5.2. Practical considerations

To appropriate the benefits of using very large data sets for data analytics research, additional effort is required, and there also are practical issues to deal with that involve limitations and feasibility issues.

5.2.1. Difficulties with data collection, cleaning and reliability

Collecting and cleaning large data sets is never easy. Advanced, scalable data processing and programming skills for analytics software are required (e.g., DB SQL, SAS, Matlab, Stata, R, etc.). Data come from multiple sources and data scientists must match and integrate large tables of data to make them useful. Real-world business data tend to be messy, with inconsistencies, built-in biases, and errors. So analysts need to use special care and a lot of efforts to prepare and understand the contents of the data. Summary statistics offer useful clues to check whether the data are consistent and logical. The data are likely to be noisy as well, so it is appropriate to identify the appropriate scope and volume of data that are needed to support the estimation or computation of various models and deliver meaningful statistical results.

5.2.2. Organizational relationship development

Data collected in business contexts typically are proprietary, and subject to non-disclosure agreements. Fortunately, organizations today seem to have become aware of the strategic importance of investing in business, consumer and data insight analytics-based decision-making. Changing attitudes toward the practicality of working with large data stores have encouraged academic researchers and industry practitioners to join forces to understand people and the organizations, communities, and markets in which they live and work. The new imperative in data analytics is *value co-creation* [90], which encourages many organizations to bring people with different skills together to jointly pursue a high-value agenda of digital innovation for leading-edge information strategy.

5.2.3. Data acquisition and information security

The acquisition of large data sets for joint academic and industry research from a sponsor is a process that requires care and foresight to ensure that nothing goes wrong. Porting data between an organizational sponsor and a university-based research center gives rise to a range of issues. Data need to be anonymized or aggregated so that no personally-identifying information is made available. They have to be sent from one institution to another with full controls to minimize risk and errors. Also, even after data have been ported to secure servers in a research institution, project researchers still must be careful to prevent any data leakages, inappropriate uses, or losses of control. To implement the necessary protections is a costly and challenging job, which requires attention to data handling process design, the specification and implementation of training and procedures so employees know their responsibilities, periodic auditing to ensure the research and data server environments are properly protected, and problem-identification processes that identify the potential for information security events to arise. Our experience with this process suggests best practice is not a matter of how individuals work with data. Instead, it is how their organization encourages and structures the work of its research faculty, staff, students and other collaborators to ensure safe data and information handling. Organizational competencies support the development of staff competencies and competitive advantage in developing skills with data analytics, as well as data acquisition and information security safeguards for data at scale. These things are likely to become the new hallmarks of the capabilities of leading research universities.

6. Conclusion

The excitement is high around the new opportunities that big data make available in research. We have emphasized the importance in the role it plays to diminish the three-horned dilemma in computational social science research. This change is a paradigm shift that enables us to study a wider range of issues in time and context with unprecedented control and new real-world insights. Still, the challenges for conducting this kind of research are significant, as our discussion of practical considerations suggests.

Beyond the new research approach that we have described remain the same kinds of issues that make creativity in research a never-ending pursuit, and a distant image on the horizon of new knowledge. Big data offer us new ways to proceed, but the sharp thinking that leading researchers seem to demonstrate about how to craft insightful experiments, distill useful explanations and rework theoretical interpretations from intuition to knowledge is often nothing short of miraculous. As Jonah Lehrer [64,65], author of the “Frontal Cortex” blog at *Wired* magazine, has written, a key quality of the kinds of people who are successful with creativity in the range of activities they pursue is the psychological trait called “grit.” Grit is about persistence, work and iteration much more than it is about discovery, since it describes the effort and the process that we must engage in to be successful in research. With big data and the methods that can be used to leverage it, nothing will come for free. We also should not confuse changes in the costs of methodology for discovery with the typically high expenses of achieving insights.

We must remind the reader that results with data alone rarely tell the full story that managers need to hear. For managers to successfully interpret what the data have to say based on sophisticated analytics often requires relatively simple background information. This might include, for example: what a firm's competitors were doing when a firm introduced new prices for its digital telephony and entertainment services; how e-banking transaction patterns changed at physical automated teller machine locations, when a bank introduced new functionality through its Internet banking or mobile phone banking channels; or how the creation of Wikipedia material changed in the face of endogenous and exogenous shocks.

Big data, in our view, are powerful – powerful to an extreme in some ways – but they also complement simple and unsophisticated managerial knowledge about what is actually going on in the real-world setting. We advocate “walkabout” observation used in social science research for identifying key informants in the research contexts that we study with big data. We need to have their help so we can understand the simple aspects and complexities of the research settings, and key events that have occurred or are due to happen that make it possible to conduct insightful natural and fully-specified experiments.

Although we have focused on how the presence of big data changes the paradigm for computational social science research, the insights obtained from the new approaches are key to more effective data-driven decision-making by managers. Brynjolfsson et al. [18] have pointed out that firms that emphasize decision-making based on business analytics have higher performance in productivity, asset utilization, return on equity and market value. Since big data are now everywhere and most firms can acquire it, the key to competitive advantage is to accelerate managerial decision-making by providing managers with implementable guidelines for the application of data analytics skills in their business processes. This takes us beyond the "diamonds in the data mine" to the rigor of best practices in data analytics for the business world [70]. Humans still will design the processes, and insights can be best discovered with a combination of machine-based data analysis and the intuition of people [19].

Acknowledgments

The ideas in this article were presented as they were developed at: the 2011 Workshop on E-Business in Shanghai, China; the 2012 International Conference on Information Management in Kaohsiung, Taiwan; the 2012 Wuhan International Conference on E-Business in China; the China University of Geoscience; and included as discussion papers for the 2012 Workshop on Analytics for Business, Consumer and Social Insights, and the 2012 International Conference on Electronic Commerce in Singapore. We acknowledge the people who provided us with comments at the conferences, workshop and universities we visited, and for ongoing discussions with colleagues Darren Choo, Peter Cook, Pedro Ferreira, Steve Fienberg, Kim Huat Goh, Timothy Goh, Jia Jia, Kwansoo Kim, Youngsoo Kim, Adrian Koh, Ramayya Krishnan, Prakash Kumar, Hoong Chuin Lau, Dongwon Lee, Ee-Peng Lim, Dan Ma, Rakesh Menon, Steve Miller, Archan Mishra, Andrew Ngai, Srini Reddy, Mike Sherman, Jialie Shen, Yesha Sivan, Sushila Shivpuri, Priyanka Singh, Insoo Son, Raj Srivastava, Rahul Telang, Chuck Wood, Irene Xu, Dongsong Zhang, Michael Zhang, Peiran Zhang, Jing Zhao, Yinping Yang, Alejandro Zentner, Vladimir Zwass and the doctoral students in IS 705 on Research Foundations for IS and Management Research at the School of Information Systems at Singapore Management University. Interactions with faculty, research colleagues and senior administration colleagues related to the "Analytics for Business, Consumer and Social Insights" interdisciplinary Area of Excellence at Singapore Management University were helpful too. This research was supported by the Singapore National Research Foundation under the International Research Centre @ Singapore Funding Initiative, administered by the Interactive Digital Media Program Office (IDMPO).

References

[1] C. Anderson, The end of theory: the data deluge makes the scientific method obsolete, Wired (July 16 2008).

[2] L. Anselin, Spatial Econometrics: Methods and models, Kluwer, Dordrecht, Netherlands, 1988.

[3] S. Aral, D. Walker, Creating social contagion through viral product design: a randomized trial of peer influence in networks, Management Science 57 (9) (2011) 1623–1639.

[4] S. Aral, D. Walker, Identifying social influence in networks using randomized experiments, IEEE Intelligent Systems 26 (5) (2011) 91–96.

[5] M. Archer, R. Bhaskar, A. Collier, T. Lawson, A. Norrie, Critical Realism: Essential Readings, Routledge, London, UK, 1988.

[6] S. Athey, G. Ellison, Position auctions with consumer search, Quarterly Journal of Economics 126 (3) (2011) 1213–1270.

[7] S. Athey, D. Nekipelovb, A Structural Model of Sponsored Search Advertising Auctions, Working paper, Department of Economics, Harvard University, Boston, MA, 2009.

[8] Y. Bakos, E. Brynjolfsson, Bundling information goods: pricing, profits, and efficiency, Management Science 45 (12) (1999) 1613–1630.

[9] Y. Bakos, E. Brynjolfsson, Bundling and competition on the Internet, Marketing Science 19 (1) (2000) 63–82.

[10] R. Bapna, P. Goes, A. Gupta, Replicating online Yankee auctions to analyze auctioneers' and bidders' strategies, Information Systems Research 14 (3) (2003) 244–268.

[11] R. Bapna, P. Goes, A. Gupta, Y. Jin, User heterogeneity and its impact on electronic auction market design: an empirical exploration, MIS Quarterly 28 (1) (2004) 21–43.

[12] R. Bapna, W. Jank, G. Shmueli, Consumer surplus in online auctions, Information Systems Research 19 (4) (2008) 400–416.

[13] S. Bhattacharjee, R.D. Gopal, K. Lertwachara, J.R. Marsden, R. Telang, The effect of digital sharing technologies on music markets: a survival analysis of albums on ranking charts, Management Science 53 (9) (2007) 1359–1374.

[14] R.S. Bivand, E.J. Pebesma, V. Gómez-Rubio, Applied Spatial Data Analysis with R, Springer, New York, NY, 2008.

[15] J.M. Box-Steffensmeier, B.S. Jones, Event History Modeling: A Guide for Social Scientists, Cambridge University Press, New York, NY, 2004.

[16] D. Boyd, Big data: opportunities for computational and social sciences, blog post, Danah Boyd Apophenia. available at www.zephoria.org April 17 2010.

[17] J.J. Brown, P.H. Reingen, Social ties and word-of-mouth referral behavior, Journal Consumer Research 14 (3) (1987) 350–362.

[18] E. Brynjolfsson, L.M. Hitt, H.H. Kim, Strength in numbers: how does data-driven decision-making affect firm performance? Working Paper, Sloan School of Management, MIT, Cambridge, MA, April 22 2011.

[19] E. Brynjolfsson, A. McAfee, Race Against the Machine, Digital Frontier, Lexington, MA, 2011.

[20] A.C. Cameron, P.K. Trivedi, Regression Analysis for Count Data, 2nd ed. Cambridge Univ. Press, New York, NY, 2013.

[21] D.T. Campbell, J.C. Stanley, Experimental and Quasi-Experimental Designs for Research, Houghton Mifflin, Boston, MA, 1963.

[22] K.M. Carley, Computational organization science: a new frontier, Proceedings of the National Academy of Sciences 99 (Supplement 3) (2002) 7257–7262.

[23] B. Castellani, F.W. Hafferty, Sociology and Complexity Science: A New Field of Inquiry, Springer-Verlag, Berlin, Germany, 2009.

[24] R.K. Chellappa, R.G. Sin, S. Siddarth, Price formats as a source of price dispersion: a study of online and offline prices in the domestic U.S. airline markets, Information System Research 22 (1) (2011) 83–98.

[25] H.C. Chen, R.H.L. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact, MIS Quarterly 36 (4) (2012) 1165–1188.

[26] J. Chen, D. Liu, A.B. Whinston, Auctioning keywords in online search, Journal of Marketing 73 (4) (2009) 125–141.

[27] C.W. Churchman, The Design of Inquiring Systems, Basic Books, New York, NY, 1971.

[28] E.K. Clemons, I.H. Hann, L.M. Hitt, Price dispersion and differentiation in online travel: an empirical investigation, Management Science 48 (4) (2002) 534–549.

[29] L.G. Cooper, M. Nakanishi, Market Share Analysis: Evaluating Competitive Marketing Effectiveness, Kluwer, Boston, MA, 1988.

[30] T.H. Davenport, J.G. Harris, Competing on Analytics: The New Science of Winning, Harvard Business Press, Boston, MA, 2007.

[31] C. Dellarocas, C.A. Wood, The sound of silence in online feedback: estimating trading risks in the presence of reporting bias, Management Science 54 (3) (2008) 460–476.

[32] C. Dellarocas, X.M. Zhang, N.F. Awad, Exploring the value of online product reviews in forecasting sales: the case of motion pictures, Journal of Interactive Marketing 21 (4) (2007) 23–45.

[33] R. Dhar, I. Simonson, The effect of forced choice on choice, Journal Marketing Review 40 (2) (2003) 146–160.

[34] A. Dimoka, P.A. Pavlou, F.D. Davis, NeuroIS: the potential of cognitive neuroscience for Information Systems Research, Information Systems Research 22 (4) (2011) 687–702.

[35] C. Doctorow, Big data: welcome to the petacentre, Nature 455 (2008) 16–21.

[36] K.A. Duliba, R.J. Kauffman, H.C. Lucas, Appropriating value from computerized reservation system ownership in the airline industry, Organization Science 12 (6) (2001) 702–728.

[37] W.H. Dutton, E.T. Meyer, Experience with new tools and infrastructures of research: an exploratory study of distance from, and attitudes toward e-research, Prometheus 27 (3) (2009) 233–238.

[38] Economist, Data, Data Everywhere, Special Report on Managing, Information, February 25 2010.

[39] B. Edelman, M. Ostrovsky, Strategic bidder behavior in sponsored search auctions, Decision Support Systems 43 (1) (2007) 192–198.

[40] B. Edelman, M. Ostrovsky, M. Schwarz, Internet advertising and the generalized second-price auction: selling billions of dollars worth of keywords, The American Economic Review 97 (1) (2007) 242–259.

[41] A. Elberse, Bye-bye bundles: the unbundling of music in digital channels, Journal of Marketing 74 (3) (2010) 107–123.

[42] EMC Data Science Community, Data science revealed: a data-driven glimpse into the burgeoning new field, DataMiningBlog.com. available at www.dataminingblog.com/analytics-and-data-science-reports February 5, 2012.

[43] D. Friedman, S. Sunder, Experimental Methods: A Primer for Economists, Cambridge Univ. Press, New York, NY, 1994.

[44] R. Garg, M.D. Smith, R. Telang, Measuring information diffusion in an online community, Journal of Management Information Systems 28 (2) (2011) 11–38.

[45] Gartner, Post event brief, Gartner IT Infrastructure, Operations and Management Summit 2009, Orlando, FL. available at www.gartner.com June 23–25 2009.

[46] A. Ghose, S.P. Han, An empirical analysis of user content generation and usage behavior on the mobile Internet, Management Science 57 (9) (2011) 1671–1691.

[47] A. Ghose, P.G. Ipeirotis, B. Li, Designing ranking systems for hotels on travel search engines by mining user-generated and crowd-sourced content, Marketing Science 31 (3) (2012) 493–520.

[48] S. Ghosh, Analysis: crunching big data more than a byte-sized bet, Reuters, June 8 2011.

[49] N. Granados, A. Gupta, R.J. Kauffman, Designing online selling mechanisms: transparency levels and prices, Decision Support Systems 45 (4) (2008) 729–745.

[50] N. Granados, A. Gupta, R.J. Kauffman, Information transparency in business-to-consumer markets: concepts, framework, and research agenda, Information Systems Research 21 (2) (2010) 207–226.

[51] N. Granados, A. Gupta, R.J. Kauffman, Online and offline demand and price elasticities: evidence from the air travel industry, Information Systems Research 23 (1) (2012) 164–181.

[52] J.A. Hausman, Exact consumer's surplus and deadweight loss, Amer. Econ. Rev. 71 (4) (1981) 662–676.

[53] D. Hensher, J. Rose, W.H. Greene, Applied Choice Analysis: A Primer, Cambridge University Press, London, UK, 2005.

[54] IDC, Digital data to double every 18 months, worldwide marketplace model and forecast, Framingham, MA. available at www.idc.com May 2009.

[55] IDC, Worldwide CRM applications market forecast to reach $18.2 billion in 2011, up 11% from 2010, according to IDC, December 22 2011. (Framingham, MA).
[56] R.J. Kauffman, S.T. March, C.A. Wood, Design principles for long-lived Internet agents, Intelligent Systems in Accounting, Finance and Management 9 (4) (2000) 217–236.
[57] R.J. Kauffman, T.J. Spaulding, C.A. Wood, Are online auction markets efficient? An empirical study of market liquidity and abnormal returns, Decision Support Systems 48 (1) (2009) 3–13.
[58] R.J. Kauffman, C.A. Wood, Revolutionary research strategies for e-business: a philosophy of science view in the age of the Internet, economics. Chapter 2 in: R.J. Kauffman, P.P. Tallon (Eds.), Economics, Information Systems, and Electronic Commerce: Empirical Research, M. E. Sharpe, Armonk, NY, 2009, pp. 31–62.
[59] D.G. Kleinbaum, M. Klein, Survival Analysis: A Self-Learning Text, 3rd ed. Springer, New York, NY, 2012.
[60] T.S. Kuhn, The Structure of Scientific Revolutions, University of Chicago Press, Chicago, IL, 1962.
[61] K. Lau, A. Kagan, G. Post, Market share modeling within a switching regression framework, Omega 25 (3) (1997) 345–353.
[62] S. Lavelle, E. Lesser, R. Shockley, M.S. Hopkins, N. Kruschwits, Big data, analytics and the path from insights to value, Sloan Management Review 5 (2) (2011) 21–32.
[63] D. Lazer, A.S. Pentland, L. Adamic, S. Aral, A.L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, M.V. Alstyne, Life in the network: the coming age of computational social science, Science 323 (5915) (2009) 721–723.
[64] J. Lehrer, Imagine: How Creativity Works, Houghton Mifflin Harcourt, New York, NY, 2012.
[65] J. Lehrer, Imagine: How Creativity Works, Speech, Science and Tech. Series, BookTV, CSPAN2, Washington, DC, 2012.
[66] D. Levy, D. Lee, H. Chen, R.J. Kauffman, M. Bergen, Price points and price rigidity, The Review of Economics and Statistics 93 (4) (2011) 1417–1431.
[67] D. Liu, J. Chen, Designing online auctions with past performance information, Decision Support Systems 42 (3) (2006).
[68] D. Liu, J. Chen, A.B. Whinston, Ex ante information and the design of keyword auctions, Information Systems Research 21 (1) (2010) 133–153.
[69] J.J. Louviere, D.A. Hensher, J.D. Swait, Stated Choice Methods: Analysis and Applications, Cambridge University Press, Cambridge, UK, 2000.
[70] G. Loveman, Diamonds in the data mine, Harvard Business Review 8 (15) (2003) 109–113.
[71] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A.H. Byers, Big data: the next frontier for innovation, competition, and productivity, McKinsey Global Institute Report, New York, NY, May 2011.
[72] D. McCafferty, Storage Slideshow: The Big Data Conundrum, CIO Insight, November 9 2010, November 9 2010.
[73] J. Mingers, Combining IS research methods: towards a pluralist methodology, Info. Sys. Res. 12 (3) (2001) 240–259.
[74] P.C. Reiss, F.A. Wolak, Structural econometric modeling: rationales and examples from industrial organization. Chapter 64 Handbook of Econometrics vol. 6A (2007).
[75] P.J. Runkel, J.E. McGrath, Research on Human Behavior: A Systematic Guide to Method, Holt, Rinehart and Winston, New York, NY, 1972.
[76] P. Russom, Big Data Analytics, Best Practices Report, Fourth Quarter, The Data Warehouse Institute, Renton, WA, September 18 2011. (available at tdwi.org).
[77] S. Shivpuri, Women's Place In Financial Services T & O, Remarks Made in a Public Forum, School of Information Systems, Singapore Management University, Singapore, November 14 2011.
[78] G. Shmueli, O. Koppius, Predictive analytics in IS research, MIS Quarterly 35 (3) (2011) 553–572.
[79] P. Spenner, Beware the big data hype, Forbes. available at www.forbes.com November 9 2011.
[80] H. Tanriverdi, A. Rai, N. Venkatraman, Reframing the dominant quests of information systems strategy research for complex adaptive business systems, Info. Sys. Res. 21 (4) (2010) 822–834.
[81] Teradata, Big data-driven online marketing analytics top business impact, press release, Dayton, OH. available at www.marketwatch.com September 19 2011.
[82] M. Twombly, Introduction to challenges and opportunities, Science 331 (6018) (2011) 692–693.
[83] A. Vespignani, Predicting the behavior of techno-social systems, Science 325 (5939) (2009) 425–428.
[84] C.J. Westland, The adoption of affective information technology in survey research, Information Technology and Management 12 (4) (2011) 387–408.
[85] C. White, Using big data for smarter decision making, IBM, Yorktown Heights, NY, 2011.
[86] R. Winkelmann, Econometric Analysis of Count Data, 5th ed. Springer, New York, NY, 2008.
[87] D.B. Yoffie, Competing in the Age of Digital Convergence, Harvard Bus Press, Boston, MA, 1997.
[88] X. Zhang, C. Wang, Network positions and contributions to online public goods: The case of Chinese Wikipedia, Journal of Management Information Systems 29 (2) (2012) 11–40.
[89] X. Zhang, F. Zhu, Group size and incentives to contribute: a natural experiment at Chinese Wikipedia, The American Economic Review 101 (4) (2011) 1601–1615.
[90] V. Zwass, Co-creation: toward a taxonomy and an integrated research perspective, International Journal of Electronic Commerce 15 (1) (2010) 11–48.

Ray M. Chang is a Research Scientist at the Living Analytics Research Centre, and an Adjunct Faculty in the School of Information Systems at Singapore Management University. He previously served as a Visiting Scholar at the Desautels Faculty of Management at McGill University in Montreal, Canada. He received his Ph.D. from Pohang University of Science and Technology in South Korea, and worked for several years as an R&D analyst and manager at SK Telecom. His research interests include business analytics and business intelligence, online social networks, open-source software communities, IT innovation and diffusion, and IT market strategy. His research appears in *MIS Quarterly* and *Information Systems Research*, with others across the IS, operations, and telecommunications fields.

Robert J. Kauffman is a Lee Kuan Yew Faculty Fellow for Research Excellence, and Professor of Information Systems at the School of Information Systems at Singapore Management University. He also serves as Associate Dean for Research, and Deputy Director of the Living Analytics Research Center. He recently was a Distinguished Visiting Fellow at the Center for Digital Strategies of the Tuck School of Business, Dartmouth College. He has received awards in multiple disciplines for his research contributions.

YoungOk Kwon is an Assistant Professor in the Division of Business Administration at the College of Economics and Business Administration, Sookmyung Women's University, Korea. She received the Ph.D. degree in Information and Decision Sciences from the Carlson School of Management, University of Minnesota. Her research interests include knowledge discovery and data mining, personalization technologies, business intelligence, and human decision-making. Her research has been published in *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Intelligent Systems*, *INFORMS Journal on Computing*, and presented at a number of computer science and information systems conferences.