Big Data and consumer behavior: imminent opportunities

Charles F. Hofacker
Florida State University, Tallahassee, Florida, USA
Edward Carl Malthouse
Department of Integrated Marketing Communications, Northwestern University, Evanston, Illinois, USA, and
Fareena Sultan
Northeastern University, Boston, Massachusetts, USA

Abstract
Purpose – The purpose of this paper is to assess how the study of consumer behavior can benefit from the presence of Big Data.
Design/methodology/approach – This paper offers a conceptual overview of potential opportunities and changes to the study of consumer behavior that Big Data will likely bring.
Findings – Big Data have the potential to further our understanding of each stage in the consumer decision-making process. While the field has traditionally moved forward using a priori theory followed by experimentation, it now seems that the nature of the feedback loop between theory and results may shift under the weight of Big Data.
Research limitations/implications – A new data culture is now represented in marketing practice. The new group advocates inductive data mining and A/B testing rather than human intuition harnessed for deduction. The group brings with it interest in numerous secondary data sources. However, Big Data may be limited by poor quality, unrepresentativeness and volatility, among other problems.
Practical implications – Managers who need to understand consumer behavior will need a workforce with different skill sets than in the past, such as Big Data consumer analytics.
Originality/value – To the authors' knowledge, this is one of the first articles to assess how the study of consumer behavior can evolve in the context of the Big Data revolution.

Keywords Big Data, Consumer behavior, Marketing analytics

Paper type Viewpoint

It has been said that the era of Big Data began at the point at which the cost of storing data dropped below the cost of deleting it. One need not take that saying literally to be impressed at the new sources and types of data sets available to marketers. Such data are increasingly available because more interactions with customers are taking place in social media, online and on mobile devices where all actions can be easily recorded. Consumers have become an "incessant generator of both structured, transactional data as well as contemporary unstructured behavioral data" (Erevelles *et al.*, 2015). Big Data are often characterized by three V's: volume, velocity and variety (Meta Group, 2001). Volume refers to the "bigness" property, while velocity refers to the rate at which the digital processes make Big Data even bigger. Variety refers to new formats and types of data. These are often data that are not in the rectangular form suitable for traditional statistical analysis, and that also often consist of words, images, video or other non-numeric consumer output.

In marketing, the main driver of the interest in Big Data is the potential usefulness of it for informing marketing decisions and executing marketing campaigns. Some authors have suggested that Big Data consumer analytics has significantly transformed the way marketing is conducted today (Erevelles *et al.*, 2015). *EMarketer* reports that in a 2013 survey of senior marketers in the USA, "85 per cent USA agency and brand executives said Big Data had yielded more than half of marketing initiatives when it came to increasing insights into consumer behavior" (emarketer.com, 2013). We will explore what Big Data are good for below, but for now we need only observe that much of what is described as Big Data is directly created from consumer behavior (CB), particularly in the social media environment. Such data should thus be of great interest to the discipline of marketing in general and CB researchers in particular. We will also explore the constraints operating on the use of such data by CB academics and marketing managers.

The current issue and full text archive of this journal is available on Emerald Insight at: **www.emeraldinsight.com/0736-3761.htm**

Journal of Consumer Marketing
33/2 (2016) 89–97

[ISSN 0736-3761]
[DOI 10.1108/JCM-04-2015-1399]

Received 16 April 2015
Revised 3 September 2015
17 November 2015
Accepted 20 November 2015

Journal of Consumer Marketing
Volume 33 · Number 2 · 2016 · 89–97

Big Data and consumer decision-making

To begin our discussion of the Big Data phenomenon and its intersection with CB, we use a narrational device that moves along the steps of the consumer decision-making process. In this, we will follow the structure given by Blackwell *et al.* (2005), but update it for a world in which more and more CB, and consumption itself, takes place or is immediately visible on social media (Figure 1). Our steps are problem recognition, search, alternative evaluation, purchase behavior, consumption, post-purchase evaluation and post-purchase engagement. Under the Blackwell *et al.* (2005) rubric, the exchange relationship between the customer and firm emerges from consumer problem-solving activities. We ask below, what sorts of Big Data get created from these activities and what research opportunities do they create? The research opportunities are enumerated below and summarized in below Table I.

Problem recognition

In the first stage of consumer problem-solving, the consumer sees a gap between what he or she has, or has experienced, and what he or she wants, or wants to experience. Companies can sense when this moment has arrived from a variety of sources including search queries, social media, addressable advertising (Blattberg and Deighton, 1991) and direct marketing response. Given the variety of data sources and types, we ask:

RQ1a. Which data sources are most sensitive for detecting consumer dissatisfaction?

RQ1b. How does the platform – mobile, online or social – moderate the consumer's expression of the gap?

Much discussion about products and issues with products takes place in social media environments that can be monitored. Such discussion can be used to identify ideas for new product and improvements. Bayus (2013), for example, shows how to extract new product ideas from online communities, while Chan *et al.* (2015) investigate how a customer's peer-to-peer network predicts the likelihood that they will generate a successful idea:

RQ1c. What are the key antecedents underlying consumer creativity as expressed in social media?

A change in transaction pattern can signal that the consumer has recognized a problem and is contemplating churn. As more human artifacts acquire IP addresses (IoT), companies will, for example, be able to detect that product usage is slowing or stopped. The IoT refers to devices that are connected to each other via the internet (Ashton, 2009). It is possible, for example, that IoT artifacts might upload nonverbal reaction to advertising, for example (Teixeira *et al.*, 2014). In general, the ability to sense trigger events (Malthouse, 2007) with this type of device, and legacy systems like the Web, is changing marketing practice:

RQ1d. What are the ways in which IoT artifacts can signal trigger events to marketers?

The consumer may also provide early warning signs on social media that there are problems with the relationship, and companies may have the opportunity to learn about the issues from the above sources and address them (Malthouse, 2007). The company should be interested in events that portend a change in the lifetime value of the customer, or other customers. Are customers complaining about the product, about service, or asking for certain features? E-word-of-mouth provides grist for the operations, HR and R&D managers' mill.

Qualitative research methods such as in-depth interviews and focus groups have traditionally been used to recognize problems. Such methods can possibly be improved by Big Data. For example, an examination of what is being said on social media about a brand could help prepare discussion outlines for face-to-face discussions with consumers.

Figure 1 Consumer problem-solving and example Big Data sources

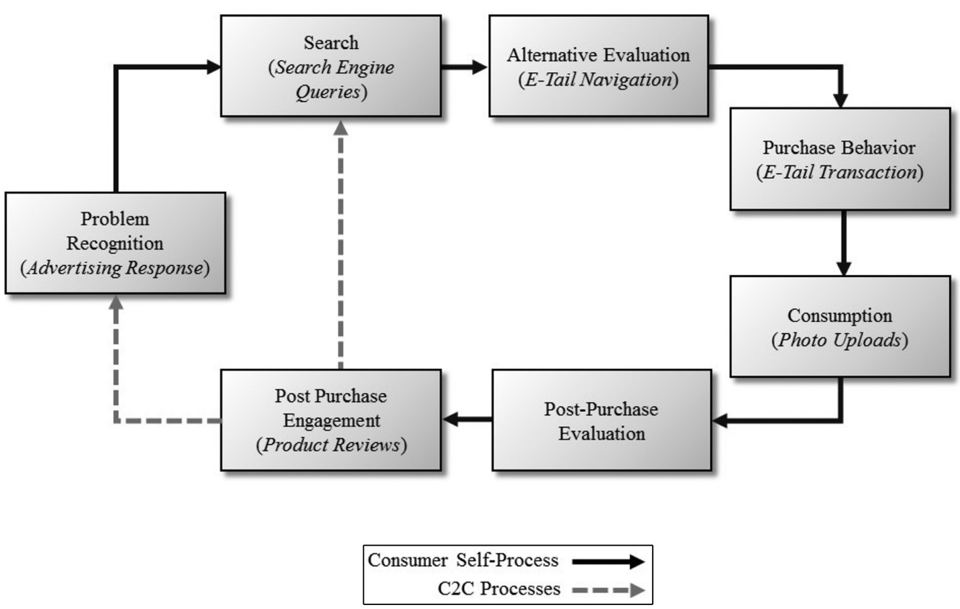


Table I Research questions

RQ1a	Which data sources are most sensitive for detecting consumer dissatisfaction?
RQ1b	How does the platform – mobile, online or social – moderate the consumer's expression of the gap?
RQ1c	What are the key antecedents underlying consumer creativity as expressed in social media?
RQ1d	What are the ways in which "The Internet of Things" or IoT artifacts can signal trigger events to marketers?
RQ2a	How does the platform – mobile, online or social – moderate the relationship between psychological characteristics and search behavior?
RQ2b	How does involvement change the type and amount of data created during consumer search?
RQ3a	How can researchers identify consumers' choice rules from Big Data sources?
RQ3b	How does the platform used for evaluation moderate the consumers' choice rules?
RQ4a	What behavioral and psychological phenomena can be inferred from the visual evidence of consumption found on social media?
RQ4b	How can researchers identify the level and type of brand engagement from the visual evidence of consumption found on social media and can this engagement be tied to marketing activities?
RQ4c	What is the relationship between engagement as revealed in text and engagement as revealed in photographs?
RQ4d	Can pleasure be inferred from IoT device output and be related to marketing activities and marketing outcomes?
RQ4e	Can arousal be inferred from IoT device output and be related to marketing activities and marketing outcomes?
RQ4f	Can dominance be inferred from IoT device output and be related to marketing activities and marketing outcomes?
RQ5a	Can we reliably distinguish between satisfaction (Oliver, 1980) and value (Zeithaml, 1988) on social media with large data sets?
RQ5b	Can we infer the subjective experience of satisfaction from IoT devices and can it be tied to marketing activities?
RQ5c	Can we infer the subjective perception of value from IoT devices and can it be tied to marketing activities?
RQ6a	How do review volume, variance and valence combine to create attitudes toward reviewed products?
RQ6b	How do social comparison processes drive review dynamics?
RQ6c	How do previous review volume, variance and valence impact new reviews?
RQ6d	What are the social psychological mechanisms underlying sender–receiver interaction?
RQ6e	What are the psychological mechanisms that underlie the expression of loyalty and how do they vary across platforms?
RQ7a	What cognitive, affective or behavioral variables, when experimentally manipulated, will moderate Big Data results?
RQ7b	Which firm actions inspired by data mining rather than theory will cause counterproductive CB?
RQ7c	How do psychological traits (e.g. involvement, innovativeness, locus of control, Big 5 personality or need for cognition) moderate market response to textual design elements?
RQ7d	How do psychological traits (e. g., involvement, innovativeness, locus of control, Big 5 personality or need for cognition) moderate market response to visual design elements?
RQ7e	How does privacy concern vary across IoT data and other location-specific data?

Search

In the traditional offline world, the consumer had trouble finding alternatives. In the digital world, the problem is having too many alternatives. The enriched search process now throws off digital data at every turn. In the past, a direct-to-consumer retailer would have a record of all items that were purchased, but now a retailer can record all of the search activity on its website and shopping app that leads up to the purchase as well, such as logs recording all activities on the site, including which items have been searched, clicked on, added to a shopping cart or wish list, abandoned, purchased, etc. It also knows which search terms attracted prospective customers from search engines, and whether it was a paid search term or an organic one. Klapdor *et al.* (2014) give a current review of search and propose ways to improve it:

RQ2a. How does the platform – mobile, online or social – moderate the relationship between psychological characteristics and search behavior?

Blackwell *et al.* (2005) emphasized the difference between routine problem-solving, limited problem-solving and extended problem-solving. Search effort is greater for the latter categories. Involvement is the prototypical moderator of the relationship between search and choice. We presume that high-involvement choice will necessarily throw off more data than low-involvement choice, and will involve different types of data. The amount and type of data will further vary depending on the product category and its mix of digital versus non-digital attributes. The riot of new data will add to the background complexity for managerial decisions:

RQ2b. How does involvement change the type and amount of data created during consumer search?

Alternative evaluation

The e-tailer will have data on alternative evaluations including consideration sets and inferred choice rules based on navigation sequences. In addition, the recommendation agents (Murray and Häubl, 2009) can utilize Big Data as input, including the choices of other shoppers (Ansari *et al.*, 2000), a technique known as collaborative filtering. Shopping cart abandonment (Kukar-Kinney and Close, 2010) can signal that customers are making price comparisons at other sites. Search behavior in general can provide additional hints as to the way the consumer is planning to evaluate the alternatives:

RQ3a. How can researchers identify consumers' choice rules from Big Data sources?

RQ3b. How does the platform used for evaluation moderate the consumers' choice rules?

Purchase behavior

We now have detailed digital records of what happens in many different types of purchase environments. Data sources include cameras in stores, mobile purchase activities on branded apps, scanners at checkout, direct marketing purchase responses online, online browsing and shopping carts, loyalty programs, 800 numbers and digital TV, among others.

Consumption

Consumers increasingly consume digitally. Media consumption is almost fully digital at this point in time (Netflix, Pandora, news sites, Kindles, iTunes). The offline part of our world is shrinking – think about how many selfies of people consuming something get uploaded to Facebook each day. Consider a restaurant customer who, before being seated, waits in the bar, checks in with Foursquare and enters the name of the craft beer he is drinking on the Untapped app. Once at the table, the waiter must wait while he uploads a photo of his hors d'oeuvres. This example illustrates the variety of Big Data formats, including transactional, location-based and pictorial. The latter in particular pose a challenge to managers and a research opportunity for CB scholars; however, Vilnai-Yavetz and Tifferet (2015) have already shown how to segment customers using Facebook profile pictures:

RQ4a. What behavioral and psychological phenomena can be inferred from the visual evidence of consumption found on social media?

RQ4b. How can researchers identify the level and type of brand engagement from the visual evidence of consumption found on social media and can this engagement be tied to marketing activities?

RQ4c. What is the relationship between engagement as revealed in text and engagement as revealed in photographs?

The "quantified self" and the "measured life" are popular names for how we allow more of our actions to be recorded. For example, Strava records cycling workout activity and uploads it automatically to Facebook, and wearable technologies such as watches and Fitbits record biometrics. Social media sites such as Ravelry allow knitters to post their projects, and report how many meters of yarn a member has consumed. The IoT will accelerate this trend, creating digital data from more types of consumption including that generated by using cars, vacuum cleaners, washing machines and refrigerators. Devices may reveal highly intimate physiological details about the wearer, possibly including the three classic responses pleasure-arousal-dominance (Hsieh *et al.*, 2014). Managers will need to find ways of using these new data sources to understand their customers, improve the execution of their marketing programs and make their products stickier:

RQ4d. Can pleasure be inferred from IoT device output and be related to marketing activities and marketing outcomes?

RQ4e. Can arousal be inferred from IoT device output and be related to marketing activities and marketing outcomes?

RQ4f. Can dominance be inferred from IoT device output and be related to marketing activities and marketing outcomes?

Post-purchase evaluation

Consumers evaluate the gap between their expectations and their consumption experience during and after consumption. This step by itself does not create Big Data but positive or negative gaps may be described online in reviews, tweets, shared photos, etc. We also wonder whether traditional post-purchase evaluation measures of satisfaction, commitment and attitudinal loyalty will be as relevant or necessary when the firm can monitor every interaction and use:

RQ5a. Can we reliably distinguish between satisfaction (Oliver, 1980) and value (Zeithaml, 1988) on social media with large data sets?

RQ5b. Can we infer the subjective experience of satisfaction from IoT devices and can it be tied to marketing activities?

RQ5c. Can we infer the subjective perception of value from IoT devices and can it be tied to marketing activities?

Post-purchase engagement

Product reviews are the prototypical Big Data exemplar. These exhibit all of the three V's mentioned in the opening section. Reviews and comments, and their consequences on others, take us full circle in Figure 1, back to problem recognition and alternative evaluation, as the one consumer's behavior becomes the antecedent for another's. With that in mind, we start by discussing the impact of reviews on the reader. This topic represents a great example of extant and potential Big Data research, including combining multiple data sets (variety) such as box office sales, critics' reviews, consumer reviews from different sources and so forth. Reviews have their own three V's: valence, volume and variance; all of which have been shown to impact the reader (Andrews *et al.*, 2015; Anderson and Lawrence, 2014; Cui *et al.*, 2012; Dellarocas *et al.*, 2007; Chen *et al.*, 2011; Clemons *et al.*, 2006):

RQ6a. How do review volume, variance and valence combine to create attitudes toward reviewed products?

It is interesting to note that one review writer can have an impact on later writers and that reviews are subject to intertemporal effects. These dynamic effects have been explored by Langley *et al.* (2014), Li and Hitt (2008) and Chen *et al.* (2011):

RQ6b. How do social comparison processes drive review dynamics?

RQ6c. How do previous review volume, variance and valence impact new reviews?

In addition to dynamics, the topic of product reviews necessarily encompasses to groups of consumers: readers and writers (King *et al.*, 2014). This leads us to suggest:

RQ6d. What are the social psychological mechanisms underlying sender-receiver interaction?

In addition to product reviews, the sources of post-purchase engagement are particularly numerous and include mobile apps, check-in platforms, Social TV and various forms of likes, blogs, retweets, forwards, posting pictures or videos about a product, comments made during public service exchanges and other forms of e-word-of-mouth (King *et al.*, 2014). The very meaning of loyalty in the digital era and the nature of the sorts of data it produces is a wonderfully dynamic CB topic worthy of continuing investigation. Thus we suggest:

RQ6e: What are the psychological mechanisms that underlie the expression of loyalty and how do they vary across platforms?

Exogenous factors

Detailed information can be integrated from weather sources, private list brokers, voting records, highway sensors monitoring traffic, government databases and numerous other sources. Varied Big Data can provide a more holistic view of the consumers' journey.

Problems associated with Big Data

We have tried to make the case that Big Data from online, mobile, but especially social media, can provide information complementary to traditional CB methods while also providing an advantage to marketers. We must also acknowledge and address various negative aspects.

Big Data come from the past

Big Data, by definition, are "Big" about the past. While historical accounts can often be important, prescriptions for the future are, by definition, more actionable. Theories and models provide such prescriptions. Big Data can be used to generate insights that inspire theoretical explanations, test theories and calibrate models, but have little value without a theory and/or model to provide explanation.

Big Data record what customers did, but not why

Observed behaviors are very useful to marketers, but constructs that have traditionally been important such as motivation and attitude can only be inferred from many of the Big Data sources we describe. One solution to this problem is to supplement Big Data sources with "Little Data" from more traditional research methods such as surveys.

This has the potential to change CB research in a fundamental way. In the past, a large amount of CB research has relied on survey samples and experiments, where consumers were asked about their attitudes, intentions and behaviors. While attitudes could be measured, the representativeness of such samples has often been questionable because of poor response rates and sampling frames, e.g. students, Mechanical Turk or marketing research panels. Big Data offer the possibility of having records of behavior for all current customers. The importance of understanding constructs such as motivations, cognitions, emotions, etc., will not diminish, but the question is whether such constructs can be inferred from behaviors. Another way to view the change is that in the past, attitudes and other constructs at time t were used to explain behaviors at time $t + 1$. In the Big Data world, behaviors at time t affect attitudes and other constructs at time $t + 1$. In the future, models that can capture the dynamic interchange between the two will be needed. Surveys will continue to be used, especially for understanding prospective customers, but the advantages of using records of customer behaviors are intriguing and create opportunities to extend CB research.

Big Data quality cannot be assumed

Having a database does not mean that it can be used for marketing purposes (Even *et al.*, 2010). Maintaining a clean database requires substantial effort, and the task of preparing a data set for analysis will often take longer than the analysis itself. The number of "data cleaning" books that have been published in the popular press confirms this point. For example, suppose that a firm learns a customer's age in years, or presence of children, at some point. It is rare for the firm to also record when the fields were populated, so the variables could be 10 years out of date. The length of time passed since being populated could vary across customers, and even across variables within a particular customer. Another common problem is that the same variable is recorded in multiple databases maintained by the organization, and when new information about the variable is known, not all versions of the variable are updated. Thus, there can be conflicting data and no way to know which version is more current.

Big data sets may not be representative

Marketers should not be impressed by the size of a data set alone, and should inquire about how the data were sampled and potential biases created by the sampling procedure. For example, a company's data may be detailed and numerous but only about long-term customers. In this case, there is the problem of survival bias. A second example might occur if customers self-select into certain data files or if managers pick consumers to include on some systematic basis (selection bias).

Let's take a closer look at the popular Big Data application known as sentiment analysis (Schweidel and Moe, 2014), which is based on opinions expressed online. How representative are these opinions? Social media should be thought of as the world's largest focus group, and be analyzed as such. For example, if 20 people are complaining loudly about some feature of a product on a review site, the company should treat this as an insight and investigate further. The company cannot infer, for example, the fraction of all targeted customers who have this complaint, which would require a more rigorous sampling plan. The company does not know if those complaining are customers, or if they are, in reality, employees at a competitor sabotaging the brand (Hu *et al.*, 2011, 2012; Chevalier and Mayzlin, 2006; Mayzlin *et al.*, 2014; Anderson and Simester, 2014; Li and Hitt, 2008; Anderson and Magruder, 2012). Nor can the company infer the effect of the complaints on sales. Many companies offer social-media tracking services and produce white papers showing how well chatter in social media correlates with outcomes such as unit sales. What is usually not publicized are situations when there is no relationship, and it is easy to cherry-pick examples to show a correlation. The same issues of representativeness should also be considered with customer reviews.

Big Data may not generalize

Another consideration of descriptive studies is the extent to which they generalize to other periods or situations, i.e. external validity. While one may have a complete census from some period, and the census may even be free of measurement errors, omitted variables and sampling errors, one cannot assume that the findings generalize.

Big Data may have omitted variables

Not all factors influencing consumer decisions will be recorded in Big Data sets, creating potential omitted variable biases. For example, exposure to customer reviews on a retailer's website provides a seemingly closed environment for studying their effects on purchase, but reviews (and other company-website-specific variables such as price and discounts) are not the only factors influencing the purchase decision. Other factors such as exposure to mass advertising, competitive actions (e.g. competitor's price and advertising) and individual differences (e.g. whether the consumer is influenced by others and is sensitive to price) are often unknown. There are research opportunities to bring data sets together to address the omitted-variable biases, or devise models that allow for heterogeneity:

RQ7a. What cognitive, affective or behavioral variables, when experimentally manipulated, will moderate Big Data results?

Big Data can be volatile

The value of certain types of Big Data is perishable and may vanish even in minutes. For example, knowing that a customer is in proximity to a store is generally not as useful after the customer moves to a new location, which could occur within seconds if the customer is driving. "In this world of real time data you need to determine at what point the data is no longer relevant to the current analysis" (Normandeau, 2013).

Big Data show associations but not causation

Big Data are often observational, where subjects are not randomly assigned to treatment levels. For example, customer relationship strategies often prescribe that firms invest more marketing resources (and thus contact points) in better customers than weaker ones. This creates a confound, or an endogeneity problem, where all high-value customers receive more marketing, and all low-value customers receive less. The level of marketing activity is confounded with customer quality, and it is impossible to disentangle separate effects unless the firm withholds some marketing from some better customers, and invests more marketing in some weaker ones. Such experiments may not be profitable and consequently firms resist them.

The danger with such correlational research, however, has always been in not understanding the causal relationship between the variables. In other words: if an organization does something (action/cause), what will happen (outcome)? Answering this question requires understanding the causal relationship between the actions and the outcomes. Of course, an alleged cause may be correlated with the outcome of interest, but the correlation could be due to omitted variables or even reversed causality. If the alleged cause does not affect the outcome, then changing it through a marketing action will not produce the desired change in the outcome variable, and the marketing resources spent on the action will be wasted. CB theory, by emphasizing underlying process, is one bulwark against causal ambiguity:

RQ7b. Which firm actions inspired by data mining rather than theory will cause counterproductive CB?

Consistently with the inductive process shown in Figure 2, research in many areas often starts with data mining and finishes with lab experiments to confirm theory. Here is an area where Big Data proponents and CB researchers might fully agree. While marketing scientists do their best to work around endogeneity with various mathematical tricks, managers use extensive A/B testing and CB researchers of course also perform experiments. The bigger the data set, the easier it is to find spurious correlation. Experimentation is one great antidote for the estimation bias created from missing variables. Another relevant form of logical inference is abductive reasoning, which proceeds from observation to a theory accounting for the observation that is the simplest and most likely explanation (Josephson and Josephson, 1995).

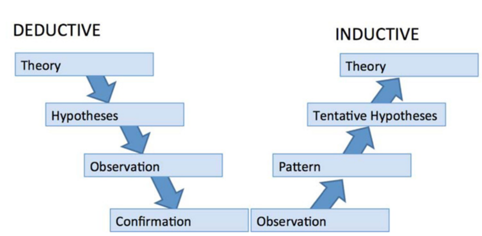
Despite the agreement on the value of experiments, Big Data proponents tend to design experiments atheoretically (Anderson, 2008) for online, mobile or promotional interaction. In some cases, a genetic algorithm creates new experimental conditions using a variety of textual or visual "genes" rather than the researcher using theory, or even thinking. We consider it an open question: Will CB researchers carefully designing experiments be able to keep up with managers exploring the behavior of millions of customers with hourly A/B experiments?

RQ7c. How do psychological traits (e. g., involvement, innovativeness, *locus* of control, Big 5 personality or need for cognition) moderate market response to textual design elements?

RQ7d. How do psychological traits (e. g., involvement, innovativeness, *locus* of control, Big 5 personality or need for cognition) moderate market response to visual design elements?

Big Data may address criticisms of the consumer research process that have been made over the past decades. For

Figure 2 Deduction versus induction



example, Paul (1983) critiques all phases of the consumer research process and identifies issues that should be addressed. For example, "In consumer research, overt behavior is seldom studied, even as a dependent variable to be predicted by interval events (p. 387)." He summarizes issues in establishing construct validity and concludes:

> [. . .] thus, many of the validity problems which currently plague consumer research may be reduced by investigating overt behavior [. . .] it is clear that at least baseline data on consumer overt behavior are needed (p. 388).

He also discusses issues with self-reported data collection methods. Big Data record overt behaviors and avoid such issues. He also questions the use of "laboratory experiments which focus on internal validity (p. 390)" because they eliminate the "complex dynamic process of human behavior". Field experiments enabled by online, media and mobile environments creating Big Data address this issue.

Another example comes from Arndt (1985) and Hunt (1983), who criticize marketing scholars for focusing too much on the "logic of justification", where the only scientific part of marketing research is model building and hypothesis testing. Arndt argues that "excluding the creative, hypothesis formation stages from science may make research in marketing the domain for conceptual auditors and controllers, driving out the visionaries and bold thinkers". The exploratory nature of many Big Data applications lends itself more to hypothesis formation than hypothesis testing.

Big Data and consumer privacy concerns/ethical issues

The collection of Big Data has the potential to worsen consumer privacy concerns. The convenience and relevance of personalization carry with them serious privacy concerns (Aguirre *et al.*, 2015). What's more, as we have described, there is an increasing variety of data sources and contexts. Further, the consumer is often not aware that data collection is taking place. The data sources include online navigation, and social media participation, but increasingly location data, data from mobile beacons and very intimate data that are generated in the IoT:

RQ7e. How does privacy concern vary across IoT data and other location-specific data?

Conclusions

Big Data grow bigger every second, every day. CB on social media represents a major driver of the phenomenon. The resulting exponential growth in the use of social media by the consumer and the growth of the IoT herald the 3 Vs: volume, velocity and variety. There are Facebook "likes" and Twitter "tweets", and check-ins, pins and posts:

> Social media has proven to be an endless fountain of such data. Each day over 350 million photos are uploaded to Facebook and over 500 million tweets are published (Marc Blinder, Director, Social and Strategic Marketing at Adobe).

There is information generated relevant to all stages of the consumer decision-making cycle, including what the consumer does, how it is done, where they consume, when they do it and with whom they consume.

We believe that academic and managerial dialogue between theory-based CB researchers and data mining researchers is necessary. A synthesis is desirable and may be possible. Only dialogue and cross-pollination will reveal whether this is so. We sense that after years of consensus on what progress means in our field, and how we achieve it, a new equilibrium between data and theory is emerging.

A substantial fraction of Big Data applications have either been descriptive (e.g. sentiment analysis and consumer reviews) or used to optimize tactics (e.g. recommendation systems, ad targeting and optimizing search keyword purchases) (Hennig-Thurau *et al.*, 2012; Chung and Rao, 2012). The implications of social media data on customer relationship management are discussed by Malthouse *et al.* (2013) and Ho *et al.* (2012). Such uses are valuable and will continue to grow, but there are also opportunities to develop more theory-oriented applications of Big Data. The observational nature of Big Data implies that resulting causal conclusions will have questionable internal validity. Consequently, many findings should be treated as exploratory, generating insights prompting theories and hypotheses, rather than confirmatory. Moreover, the fact that Big Data are often recorded across products and categories means that insights and theories can be easily examined with greater generality. The environments that enable Big Data, social, online and mobile, also enable rigorous tests, often with strong external validity. CB researchers should seize the opportunities to use Big Data for generating insights, theories and hypotheses, and the social, online and mobile environments to execute rigorous field tests. A caveat to this opportunity is that managers need to be cognizant of consumer privacy concerns that result from Big Data collection and utilization, even as they use it to deliver more relevant products and promotions to consumers. Only in such a case, can the opportunities resulting from the interaction between Big Data and CB be fully realized.

References

Aguirre, E., Grewal, D., Roggeveen, A.L. and Wetzels, M. (2015), "Personalizing online, social, and mobile communications: opportunities and challenges", Working Paper.

Anderson, C. (2008), "The end of theory: the data deluge makes the scientific method obsolete", *Wired Magazine*, 23 June.

Anderson, C. and Lawrence, B. (2014), "The influence of online reputation and product heterogeneity on service firm financial performance", *Service Science*, Vol. 6 No. 4, pp. 217-228.

Anderson, E.T. and Simester, D.I. (2014), "Reviews without a purchase: low ratings, loyal customers, and deception", *Journal of Marketing Research*, Vol. 51 No. 3, pp. 249-269.

Anderson, M. and Magruder, J. (2012), "Learning from the crowd: regression discontinuity estimates of the effects of an online review database", *Economic Journal*, Vol. 122 No. 563, pp. 957-989.

Andrews, M., Luo, X., Fang, Z. and Ghose, A. (2015), "Mobile ad effectiveness: hyper-contextual targeting with crowdedness", *Marketing Science*, in press.

Ansari, A., Essegaier, S. and Kohli, R. (2000), "Internet recommendation systems", *Journal of Marketing Research*, Vol. 37 No. 3, pp. 363-375.

Arndt, J. (1985), "On making marketing science more scientific: role of orientations, paradigms, metaphors, and puzzle solving", *Journal of Marketing*, Vol. 49 No. 3, pp. 11-23.

Ashton, K. (2009), "That 'internet of things' thing", *RFID Journal*, Vol. 22, pp. 97-114.

Bayus, B.L. (2013), "Crowdsourcing new product ideas over time: an analysis of the Dell IdeaStorm community", *Management Science*, Vol. 59 No. 1, pp. 226-244.

Blackwell, R.D., Miniard, P.W. and Engel, J.F. (2005), *Consumer Behavior*, 10th ed., South-Western College Publications.

Blattberg, R.C. and Deighton, J. (1991), "Interactive marketing: exploiting the age of addressability", *Sloan Management Review*, Vol. 33 No. 1, pp. 5-14.

Chan, K.W., Li, S.Y. and Zhu, J.J. (2015), "Fostering customer ideation in crowdsourcing community: the role of peer-to-peer and peer-to-firm interactions", *Journal of Interactive Marketing*, Vol. 31, pp. 42-62.

Chen, Y., Wang, Q. and Xie, J. (2011), "Online social interactions: a natural experiment on word of mouth versus observational learning", *Journal of Marketing Research*, Vol. 48 No. 2, pp. 238-254.

Chevalier, J. and Mayzlin, D. (2006), "The effect of word of mouth on sales: online book reviews", *Journal of Marketing Research*, Vol. 43 No. 3, pp. 345-354.

Chung, J. and Rao, V. (2012), "A general consumer preference model for experience products: application to internet recommendation services", *Journal of Marketing Research*, Vol. 49 No. 3, pp. 289-305.

Clemons, E., Gao, G.G. and Hitt, L.M. (2006), "When online reviews meet hyperdifferentiation: a study of the craft beer industry", *Journal of Management Information Systems*, Vol. 23 No. 2, pp. 149-171.

Cui, G., Lui, H.K. and Guo, X. (2012), "The effect of online consumer reviews on new product sales", *International Journal of Electronic Commerce*, Vol. 17 No. 1, pp. 39-58.

Dellarocas, C., Zhang, X. and Awad, N.F. (2007), "Exploring the value of online product reviews in forecasting sales: the case of motion pictures", *Journal of Interactive Marketing*, Vol. 21 No. 4, pp. 23-45.

Erevelles, S., Fukawa, N. and Swayne, L. (2015), "Bid data consumer analytics and the transformation of marketing", *Journal of Business Research*, in press.

Even, A., Shankaranarayanan, G. and Berger, P.D. (2010), "Managing the quality of marketing data: cost/benefit tradeoffs and optimal configuration", *Journal of Interactive Marketing*, Vol. 24 No. 3, pp. 209-221.

Hennig-Thurau, T., Marchand, A. and Marx, P. (2012), "Can automated group recommender systems help consumers make better choices?", *Journal of Marketing*, Vol. 76 No. 5, pp. 89-109.

Ho, T.-H., Li, S., Park, S.E. and Max, Z.J. Shen (2012), "Customer influence value and purchase acceleration in new product diffusion", *Marketing Science*, Vol. 31 No. 2, pp. 236-256.

Hsieh, J.-K., Hsieh, Y.C., Chiu, H.C. and Yang, Y.R. (2014), "Customer response to web site atmospherics: task-relevant cues, situational involvement and pad", *Journal of Interactive Marketing*, Vol. 28 No. 3, pp. 225-236.

Hu, N., Bose, I., Gao, Y. and Liu, L. (2011), "Manipulation in digital word-of-mouth: a reality check for book reviews", *Decision Support Systems*, Vol. 50 No. 3, pp. 627-635.

Hu, N., Bose, I., Koh, N.S. and Liu, L. (2012), "Manipulation of online reviews: an analysis of ratings, readability, and sentiments", *Decision Support Systems*, Vol. 52 No. 3, pp. 674-684.

Hunt, S. (1983), *Marketing Theory: The Philosophy of Marketing Science*, Irwin, Homewood, IL.

Josephson, J.R. and Josephson, S.G. (1995), *Abductive Inference: Computation, Philosophy, Technology*, Cambridge University Press, Cambridge.

King, R.A., Racherla, P. Bush, V.D. (2014), "What we know and don't know about online word-of-mouth: a review and synthesis of the literature", *Journal of Interactive Marketing*, Vol. 28 No. 3, pp. 167-183.

Klapdor, S., Anderl, E.A., von Wangenheim, F. and Schumann, J.H. (2014), "Finding the right words: the influence of keyword characteristics on performance of paid search campaigns", *Journal of Interactive Marketing*, Vol. 28 No. 4, pp. 285-301.

Kukar-Kinney, M. and Close, A.G. (2010), "The determinants of consumers' online shopping cart abandonment", *Journal of the Academy of Marketing Science*, Vol. 38 No. 2, pp. 240-250.

Langley, D.J., Hoeve, M.C., Roland Ortt, J., Pals, N. and van der Vecht, B. (2014), "Patterns of herding and their occurrence in an online setting", *Journal of Interactive Marketing*, Vol. 28 No. 1, pp. 16-25.

Li, X. and Hitt, L.M. (2008), "Self-selection and information role of online product reviews", *Information Systems Research*, Vol. 19 No. 4, pp. 456-474.

Malthouse, E.C. (2007), "Mining for trigger events with survival analysis", *Data Mining and Knowledge Discovery*, Vol. 15 No. 3, pp. 383-402.

Malthouse, E.C., Haenlein, M. Skiera, B., Wege, E. and Zhang, M. (2013), "Managing customer relationships in the social media era: introducing the social CRM house", *Journal of Interactive Marketing*, Vol. 27 No. 4, pp. 270-280.

Mayzlin, D., Dover, Y. and Chevalier, J. (2014), "Promotional reviews: an empirical investigation of online review manipulation", *American Economic Review*, Vol. 104 No. 8, pp. 2421-2455.

Meta Group (2001), *3D Data Management: Controlling Data Volume, Velocity and Variety*, Gartner.

Murray, K.B. and Häubl, G. (2009), "Personalization without interrogation: towards more effective interactions between consumers and feature-based recommendation agents", *Journal of Interactive Marketing*, Vol. 23 No. 2, pp. 138-146.

Normandeau, K. (2013), "Beyond volume, variety and velocity is the issue of big data veracity", *Inside BigData*, available at: http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/ (accessed 15 April 2015).

Oliver, R.W. (1980), "A cognitive model of the antecedents and consequences of satisfaction decisions", *Journal of Marketing Research*, Vol. 22 No. 4, pp. 460-469.

Paul, P.J. (1983), "Some philosophical and methodological issues in consumer research", in Hunt, S. (1983), *Marketing Theory: The Philosophy of Marketing Science*, Irwin, Homewood, IL.

Schweidel, D.A. and Moe, W.W. (2014), "Listening in on social media: a joint model of sentiment and venue format choice", *Journal of Marketing Research*, Vol. 51 No. 4, pp. 387-402.

Teixeira, T., Picard, R. and El Kaliouby, R. (2014), "Why, when, and how much to entertain consumers in advertisements? A web-based facial tracking field study", *Marketing Science*, Vol. 33 No. 6, pp. 809-827.

Vilnai-Yavetz, I. and Tifferet, S. (2015), "A picture is worth a thousand words: segmenting consumers by Facebook profile images", *Journal of Interactive Marketing*, Vol. 32, pp. 53-69.

Zeithaml, V.A. (1988), "Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence", *Journal of Marketing*, Vol. 52 No. 3, pp. 2-22.

Further reading

Bleier, A. and Eisenbeiss, M. (2015), "The importance of trust for personalized online advertising", *Journal of Retailing*, Vol. 91 No. 3, pp. 390-409.

Dellarocas, C. (2006), "Strategic manipulation of internet opinion forums: implications for consumers and firms", *Management Science*, Vol. 52 No. 10, pp. 1577-1593.

Eisenbeiss, M., Cornelißen, M., Backhaus, K. and Hoyer, W.D. (2014), "Nonlinear and asymmetric returns on customer satisfaction: do they vary across situations and consumers?", *Journal of the Academy of Marketing Science*, Vol. 42 No. 3, pp. 242-263.

Corresponding author

Charles F. Hofacker can be contacted at: chofack@business.fsu.edu