BIG AND LEAN IS BEAUTIFUL: A CONCEPTUAL FRAMEWORK FOR DATA-BASED LEARNING IN MARKETING MANAGEMENT

Emre Soyer, Koen Pauwels and Steven Seggie

ABSTRACT

While Big Data offer marketing managers' information that is high in volume, variety, velocity, and veracity (the 4Vs), these features wouldn't necessarily improve their decision-making. Managers would still be vulnerable to confirmation bias, control illusions, communication problems, and confidence issues (the 4Cs). The authors argue that traditional remedies for such biases don't go far enough and propose a lean start-up approach to data-based learning in marketing management. Specifically, they focus on the marketing analytics component of Big Data and how adaptations of the lean start-up methodology can be used in some combination with such analytics to help marketing managers improve their decision-making and innovation process. Beyond the often discussed technical obstacles and operational costs associated with handling Big Data, this chapter contributes by analyzing the various learning and decision-making problems that can emerge once the 4Vs of Big Data have materialized.

Keywords: Decision-making; marketing analytics; lean start-up; managerial learning; innovation

> Data is a lot messier and noisier than people want to acknowledge.
>
> Nate Silver (2016)

> There's no shortage of widely held myths around Big Data. Perhaps the most dangerous is that Big Data knows all and that it obviates the need for human judgment.
>
> Phil Simon (2016)

Marketing in a Digital World
Review of Marketing Research, Volume 16, 63–83

ISSN: 1548-6435/doi:10.1108/S1548-643520190000016006

As the many books and chapters written on the topic can attest, gaining actionable insights from data has always been messy (e.g., Davenport & Harris, 2007, Pauwels, 2014, Simon, 2015). The current attention to Big Data and its volume, velocity, variety and veracity (4Vs) may give the impression of "solving" many of the issues in human judgement and decision-making identified by, for example, Ariely (2010) and Kahneman (2011). In contrast, Simon (2016, opening quote) and several others are warning about such Big Data "myths." Translating such concerns to positive recommendations would involve (1) identification of possible problems that come with using Big Data in managerial learning and decisions and (2) innovative strategies to overcome the possible problems and harness the positive power of Big Data.

Current literature on Big Data range from its definition to its various sources, from privacy and ethical issues to methods of analysis, and from machine learning to its future prospects in management research (George, Haas, & Pentland, 2014; Isaac & Dixon, 2017, O'Neil, 2016). Most relevant to our purpose, Seggie, Soyer, and Pauwels (2017) examined how Big Data could be combined with lean start-up methods to aid with business model evolution for large firms by catalyzing innovation and alleviating Big Data-related issues. The authors argue that legacy firms are under unprecedented threat from start-ups and to respond to this competition; these legacy firms could successfully leverage resources they have (specifically Big Data) while building on the lean start-up methodology to evolve business models for effective innovation.

The main premise of the arguments put forward by Seggie and colleagues is that Big Data have the potential (more than any "smaller" data) to exacerbate confirmation bias, problems with communication, and illusions of control and that an adaption of the lean start-up methodology can alleviate these biases, at least during the process of business model evolution. Specifically, legacy firms can mitigate confirmation bias through the use of Big Data, which would be primarily used to come up with hypotheses that can be then tested through experimentation, rather than using the results of analysis to directly reach conclusions. Subsequently, the build-measure-learn loop provides agility and ambidexterity to the incumbent firms allowing them to challenge any illusions of control that will result from Big Data analyses. Finally, for communication problems, the use of innovation accounting allows complex messages to be broken down into simpler metrics and processes that are much easier to communicate.

In this manuscript, we extend this approach to specifically focus on the marketing analytics component of Big Data and how adaptations of the lean start-up methodology can be used in some combination with such analytics to help marketing managers improve their decision-making and innovation prowess. We also expand the traditional 3Vs of Big Data: volume, velocity, and variety to include a fourth V, veracity, in the context of marketing management. Finally, instead of detailing the often discussed technical obstacles and operational costs associated with handling Big Data (i.e., data collection, storage, revision), we contribute by analyzing the various learning and decision problems that can emerge once the 4Vs of Big Data have materialized.

BENEFITING FROM BIG DATA IN MARKETING MANAGEMENT

Much of the research on Big Data has looked at the increased volume, velocity, and variety of data in recent years (e.g., Gartner, 2012, McAfee & Brynjolfsson, 2012). With this increase in the 3Vs, managers have come to expect better learning and decision processes with regard to the marketing function (Laney, 2001; McAfee & Brynjolfsson, 2012). Many also contend that there is a fourth V to be considered: veracity (Forrester, 2012). This is a key characteristic that further catalyzes learning from Big Data, as without the truthfulness and accuracy of the information, the value of the lessons learned through 3Vs would be considerably reduced (Lukoianova & Rubin, 2014; Raghupathi & Raghupathi, 2014). Particularly at the present time, with the increased amount of and awareness of fake news, resources should be targeted towards ensuring the veracity of the data (Lazer et al., 2018; Vosoughi, Roy, & Aral, 2018).

The promise of Big Data analytics in the marketing context involves managers receiving real-time customer insights that were not available to them in a pre-Big Data environment. For example, marketing managers are now able to engage in better forecasting and planning as a result of being able to analyze the latest customer trends in a far timelier manner than was previously possible (Du, Xie, & Schroeder, 2009). They are also more able to customize and personalize products for consumers and engage in segmentation at the micro-level (Kumar & Shah, 2011; Natter, Mild, Wagner, & Taudes, 2008). When it comes to key decisions that marketing managers have to make, such as pricing or promotion adjustments, Big Data analytics have helped them make real-time decisions (Kannan, Pope, & Jain, 2009; Silva-Risso & Ionova, 2008). Other areas where marketing managers have benefited from access to Big Data include the measurement of the return on marketing investment (e.g., Kumar, Bhaskaran, Mirchandani, & Shah, 2013; Pauwels & Reibstein, 2010; Skiera & Abou Nabout, 2013), the identification and addition of new and prospective customers (e.g., Natter, Ozimec, & Kim, 2015), and learning from previous marketing strategies to identify and design new and successful marketing strategies (e.g., Danaher, Roberts, Roberts, & Simpson, 2011; Fischer, Albers, Wagner, & Frie, 2011).

These recent and rapid developments have inevitably led to an explosion in the number of online blogs and opinion chapters that serve the purpose of generalizing these benefits to a wide range of marketing practices beyond those studied and observed.[1] Moreover, global survey initiatives such as the CMO survey AU:4 suggest that over the last decade, marketing managers are increasingly requesting and using Big Data analytics in a wide variety of functions.[2]

Such analyses, surveys, and generalizations might lead managers to believe that Big Data constitutes a safe and secure tool to obtain and communicate marketing insights (Lilien, Roberts, & Shankar, 2013; Seggie et al., 2017). Yet, it's important to acknowledge that the advantage of analytics does not stem from "the size of the data, but how you use it" (Pauwels, 2014). This is an idea that often appears to be overlooked in the clamor for bigger and bigger data and more complex statistical analyses. However, with Big Data "companies risk magnifying the impact of underlying inaccuracies and errors and falling into a big bad data trap" (Saleh, Lenhard, Goldon, & Opolon, 2015). Interestingly, the latest CMO survey for the first time reported *less* use of *marketing* analytics in actual decision-making (from 42% in February to 36% in August 2018) and its *lower contribution* to company performance (from 4.1 to 3.5 on a seven-point scale). Hence, the question becomes: could the existence of 4Vs of Big Data not only fail to contribute positively, but actually deteriorate decision processes related to marketing and sales?

DECEIVED BY BIG DATA: KEY LEARNING CHALLENGES

There are two major assumptions that are at the foundation of the multiple benefits marketing management stands to gain from Big Data in terms of learning and decisions.

The first is that the 4Vs accurately label the information that a manager possesses. That is, the available data are indeed large, fast, multidimensional, and truthful. Most of the extant literature that deals with the problems of Big Data questions the legitimacy of this assumption and discusses the adverse effects of not achieving one or more of these Vs (Chen, Chiang, & Storey, 2012; Fan, Lau, & Zhao, 2015; Kaisler, Armour, Espinosa, & Money, 2013; Kwon, Lee, & Shin, 2014; Raghupathi & Raghupathi, 2014). While this is an important issue to consider, in this chapter, we go beyond this assumption to ask "could Big Data lead to biased learning *because* it actually is high in volume, variety, velocity, and veracity"? In other words, does "Big Data" appear to provide the unquestionable "truth," even though it is also subject to the judgment and decision biases extensively documented for "small data"?

This brings us to a second assumption that is often ignored by managers and analysts: Is there a good match between learning and target settings? For data to be useful, there must be a good or approximate match between these two (Hogarth, Lejarraga, & Soyer, 2015). Specifically, the learning setting is the environment from which the data materialize. The resulting insights help decision-makers evaluate a wide variety of aspects of a given situation. The target setting, however, is where these insights are applied as decisions, predictions, and resulting marketing strategies. The analyses and decisions are often built on the assumption that these two settings match, at least approximately, so that correct lessons learned in one are aptly applied in the other (Tversky & Kahneman, 1973). Such a match is akin to the notion of external validity in an experimental setting: What one finds in the lab (learning setting) would apply in a naturally occurring situation outside of the lab (target setting) (Rothwell, 2005). However, if there is too big of a mismatch between the learning settings and target settings, data-based findings can lead to systematic biases.

Consider, for example, a marketing manager striving to evaluate the return on investment of a certain advertising campaign. If the Big Data the marketing manager has access to come from a representative sample of consumers, whose tastes and behavior are sticky in the short-to-medium term, and if the complex insights obtained and the inherent uncertainties are well understood by the manager who will act on them, then it is most likely that the 4Vs would help improve Big Data-based decisions.

However, if (1) the collection or analysis of Big Data involves censors that hide certain aspects of the actual situation, (2) there are inherent uncertainties or changes in consumer behavior or managers' objectives in the short-to-medium term, and/or (3) managers are not statistically sophisticated enough to translate the various nuances of complex analyses into actionable lessons, then the 4Vs might lead to misperceptions and exacerbate the biases rather than eliminating them. More knowledge generated in this fashion could actually make managers more confident about their unwise intuitions.

In fact, managers in environments that censor observations, feature high degrees of uncertainty, and are governed by naive decision-makers, tend to learn the wrong lessons from the information available to them (Feiler, Tong, & Larrick, 2013; Hoch & Schkade, 1996; Soyer & Hogarth, 2012). The fact that these lessons are Big Data-approved would further calcify the biases, making them potentially irreversible.

A significant discrepancy between what data teach and the setting where those lessons are applied may develop as a result of both external factors (availability and/or measurement issues) and as a result of internal mechanisms that data scientists or managers can influence (data gathering methods, selection of analysis, presentation of insights). That is, Big Data provide decision-makers with a diverse set of tools for self-deception.

To put Big Data analytics in marketing management into a learning framework, we discuss below the 4Cs (confirmation, control, communication, and confidence) that would stem from the 4Vs and lead to a mismatch between the learning and target settings.

Table 1 illustrates our learning framework. In particular, we argue that a combination of the traditional 3Vs of Big Data (variety, velocity, and volume)

Table 1. Conceptual Framework: 4Vs of Big Data, 4Cs of Learning Challenges, Traditional Methods to Mitigate the Challenges, and the Advantages Offered by the Lean Start-up Methodology.

Features of Big Data (4Vs)	Learning Challenges (4Cs)	Traditional Remedies	Lean Start-up Methodology
Volume *Large amounts*	Confirmation *Biased search and analysis*	Blindness to and falsification of hypotheses	Using Big Data to develop hypotheses
Velocity *Real-time*	Control *Biased predictions*	Regularly updating analyses to reveal shifts and trends	Build-measure-learn loop
Variety *Different types*	Communication *Biased perceptions*	Transparent presentation of statistical insights	Frequent and rapid feedback on perceptions
Veracity *Truthful*	Confidence *Inability to debias*	Statistical literacy and risk intelligence	Calibration through simpler and frequent experimentation

may enable unjustified *confirmations* of theories, illusions of *control* over future trends or outcomes, and biases due to *communication* problems (3Cs). The fourth V (veracity), however, acts as a catalyst that falsely validates the generated insights, leading decision-makers to more easily ignore the various issues with 3Cs and build excessive *confidence* in their newly found data-based intuition. While these challenges could occur in all types of data analysis, Big Data stand to exacerbate them and make decision-makers more susceptible to resulting illusions and biases, a notion that is referred to as "Big Data hubris" (Lazer, Kennedy, King, & Vespignani, 2014).

Crucially, by recognizing these challenges and their specific natures, managers can design the appropriate mechanisms to avert them. Later in this work, we offer solutions that are specific to each learning challenge and then propose an approach that combines Big Data analytics and lean start-up methodology to effectively alleviate them in marketing settings. However, the first step to mitigating such issues is becoming aware of them. Hence, we start by defining the 4Cs and revealing the major reasons behind them.

Confirmation: Biased Search and Analysis

When managers have access to data that feature the volume, velocity, variety (3Vs), they may find it easier to check if their hypotheses regarding strategies indeed hold. Unfortunately, decision-makers often disregard evidence that would invalidate their assumptions (Klayman, 1995). And because they firmly believe in the veracity of the evidence they obtained, such information-rich analytics would reduce the chances that the resulting convictions would be later scrutinized through recurrent analyses or alternative testing methods.

To complicate things further, there are two complementary ways in which the 4Vs of Big Data can lead to such false confirmations (Baron, 2000). First, the managers can collect data in a way that fits their agendas, so that different analyses on that same evidence yield to certain predestined insights. For example, the sales force is often compensated mainly on sales targets and thus has an incentive to overstate customer price sensitivity so that the company will reduce price and increase sales volumes. In several companies known to the authors, managers allowed the sales force to both design and administer surveys on what customer value most. While low price comes out as the top priority, the authors' conjoint analysis and modeling of historic data showed instead that quality, convenience, and billing characteristics were more important to customers.

Such confirmatory practices would create a mismatch between the learning and target settings: The strategies built on the lessons learned from such selected-yet-massive data would not fit the reality of the situation.

Second, the statistical analyses employed can systematically ignore or distort disconfirming evidence. Hence, although managers may be operating with representative data, the statistical assumptions made and tools selected by the analysts may depict a picture that is biased toward the initial assumptions of the final decision-makers. In particular, when the compensation and interests of those who conduct the research and analyses depend on the satisfaction of marketing managers who employ them, the findings will be more likely to confirm and build on those managers' hopes and dreams. This tendency would be exacerbated with Big Data, as there would be more opportunities to highlight confirming evidence and obscure disconfirming details.

In most situations, confirmatory search and analyses may, in fact, be simultaneously in play (Baron, 2000). As a result, two managers analyzing the same market through the same dataset about the same problem they wish to solve can end up confirming two different, even opposite, hypotheses. Both may, in fact, be biased yet now confident enough to act on their insights that, if tested objectively, could not be replicated. Simmons, Nelson, and Simonsohn (2011) found that such confirmatory search and analyses were widespread in many peer-reviewed and ultimately published academic findings, mainly due to a publish-or-perish work culture. Analogously, insider accounts by professional data scientists hint to their possible existence in marketing analytics.[3]

Control: Biased Predictions

Through the 3Vs of Big Data, managers would expect to better predict future consumer trends and behavior based on evidence from recent and remote past. The validity of this expectation, however, would depend on how much and in which ways the past (the learning setting) resembles the future (the target setting). In particular, if the environment is highly unpredictable and prone to paradigm shifts, then an analysis that perfectly fits the past is not of much use when a manager is trying to foresee potential prospects and opportunities. Moreover, the veracity of the data and the results of that analysis can further amplify the subsequent illusions of control over the future trends and behavior.

There are two major reasons why insights from Big Data may be rendered fast obsolete. First, the goals of the company or the department may change. In such a case, while the learning setting stays constant, the targets of the manager shift, causing a potential bias-prone mismatch. Knowledge, hence, needs to be refreshed accordingly through new modes of mining and analyses based on the existing data (Fang, Sheng, & Goes, 2013; Peters, Chen, Kaplan, Ognibeni, & Pauwels, 2013).

Second, the environment where data come from (the learning setting) may incorporate a high degree of uncertainty, in that major changes in the market may be driven by unpredictable yet high-impact tail events, also called Black Swans (Makridakis, Hogarth, & Gaba, 2009; Taleb, 2007). These would severely and suddenly reduce the shelf life of available Big Data and marketing analytics to managers.

For example, if a new product, such as digital photography or movie streaming, disrupts the whole market, marketing strategies within that domain may change drastically within a matter of weeks, if not days (Christensen, Raynor, & McDonald, 2016). Less dramatically, the established relationship between price and quality (the fair value line) may change in intercept, slope, and range with changing customer tastes and/or competitive entry (Pauwels & D'Aveni, 2016).

While under the 4Vs the information flow to marketing managers would still be fast (velocity), the insight analytics derived from that information is misguided if the source of that information is now obsolete (e.g., customers and sellers of the disrupted market).

Big Data (as opposed to any other type of smaller or more specific data) involves large investments made to obtain the 4Vs, which may convince managers cling onto the existing structure and analyses, leading them to use the data way beyond its (unpredictable) expiration date (Banerjee, Bandyopadhyay, & Acharya, 2013).

Communication: Biased Perceptions

The 3Vs of Big Data may indeed provide analysts with a much better capacity to analyze and produce relevant findings, which would help managers devise appropriate marketing strategies. Yet this process requires those statistical findings to be translated to a format upon which managers can effectively act. If there's a wide gap between the statistical sophistication of analysts and managers, however, then certain crucial elements of the analysis may be lost in translation. This would happen even if the data are collected and analyzed in an unbiased fashion. Some aspects of the insights would be misperceived in a way that leads to systematic biases due to communication problems. And the veracity (fourth V) of the underlying data may further render managers oblivious to such mistakes in their perceptions, making the problem even harder to identify and mitigate.

In particular, the resulting mismatch between analysis and perception can once again generate illusions of control over future outcomes, this time purely due to communication issues rather than shifts in goals or disruptions in the market. Research shows, for example, that standard presentations of results stemming from even simple statistical analyses may hide uncertainties from decision-makers, even when these are statistically sophisticated (Armstrong, 2012; Hogarth & Soyer, 2011; McCloskey & Ziliak, 1996; Soyer & Hogarth, 2012; Taleb & Goldstein, 2007; Taleb & Goldstein, 2012; Ziliak, 2012). In such a case, communications would lead managers to overgeneralize data-based insights and be confused about the relative importance of various determinants of the dependent variables under scrutiny. In fact, communication issues have been found to hinder learning from Big Data in a wide range of domains, such as customer loyalty (Cooil, Keiningham, Aksoy, & Hsu, 2007), financial investment (Kaufmann, Weber, & Haisley, 2013), and climate change (Budescu, Broomell, & Por, 2009).

Confidence: Inability to Debias

The 3Vs of Big Data, if not considered appropriately, can help analysts deceive managers, or worse, help managers deceive themselves. A crucial issue is then to make sure that managers become aware of the potential mismatches due to confirmatory tendencies, desires of control, and miscommunications. Arguably, they would be more skeptical of analytics and the resulting insights if sometimes the underlying information proves to be inaccurate. This could serve as a trigger to remind the managers that they should review and potentially revise parts of the process.

However, if Big Data contain almost always accurate information (fourth V), decision-makers can potentially confuse such veracity with value and refrain from questioning the process. This sort of halo effect would reduce the chances that mismatches due to the 3Cs of Big Data we discussed above are likely to be caught and mitigated. Hence, paradoxically, the 4Vs of Big Data may mislead managers for a longer period of time. For instance, the analyses of online metrics have demonstrated they are more volatile than purchase metrics, while traditional survey metrics are less volatile (Pauwels & van Ewijk, 2013). While online metrics accurately explained same-week sales (short-term), they created so much noise in sales prediction months in the future (long-term) that forecast accuracy was improved dropping them from the set of predictors (ibid).

MITIGATING THE LEARNING CHALLENGES OF BIG DATA

Once managers recognize and identify the reasons behind the main learning challenges that the 4Vs of Big Data can cause, they can design mechanisms to mitigate them. Research in diverse fields has made many behavioral approaches and statistical tools available to managers, each with their specific strengths and weaknesses. Below, we first review several prominent and practice-oriented mechanisms that aim to increase the match between learning and target settings. We then argue in the following section that combining Big Data analytics with lean start-up methods would actually be a viable alternative for marketing managers to harness the major strengths of these more traditional mechanisms, while avoiding many of their weaknesses.

A/B Testing

Marketing (analytics) has a long history of hypothesis testing in the form of A/B (or split-run) experiments (e.g., Brian, 2000). Subjects are randomly assigned to either "A" (the control group, typically the current practice) or "B" (the treatment group), and we compare their reaction based on previously agreed upon metrics, such as sales and profits of clicks. In online environments, "the *modification* could be a new feature, a change to the user interface (such as a new layout), a back-end change (such as an improvement to an algorithm that, say, recommends books at Amazon), or a different business model (such as an offer of free shipping)" (Kohavi & Thomke, 2017, emphasis added). Examples include the color of Bing search results and having a new tab automatically open when an internet user clicks on a link (Kohavi & Thomke, 2017).

As evident from the yearly Conference on Digital Experimentation, the technical sophistication of such A/B testing has made great leaps forward, including the creation of "Ghost Ads" to efficiently measure a baseline in comparing digital ad effectiveness (Johnson, Lewis, & Nubbemeyer, 2017). Companies such as Microsoft, Amazon, Booking.com, Facebook, Google, and Microsoft conduct more than 10,000 online controlled experiments annually, and the total benefits of these many "tiny changes" run in the millions of dollars (Kohavi & Thomke, 2017).

So if "any company that has at least a few thousand daily active users can conduct these tests" (Kohavi & Thomke, 2017), why are digital companies still reluctant to experiment even with easy-to-change actions such a prices (Ariely, 2010), let alone business models? As Pauwels (2014) shows across companies and continents, managers may be fine with experimenting with tactical changes (such as colors and interfaces mentioned above), but typically avoid experiments with more strategic changes. The perceived risk is often too high (e.g., Wiesel et al., 2011) and senior managers will typically aim to steer the experiment toward the desired action. Moreover, managers in the "control group" A need to fully trust that their lower metrics (e.g., lower sales in their territory) would not be held against their career. If the treatment is generally believed to be beneficial, no manager wants to be in the control group – just as patients don't want to be in the control group of the medical trial of a promising new drug. The key difference is that patients are randomly assigned and have no way of influencing which treatment they get nor how the results will be interpreted.

Blindness to and Falsification of Hypotheses

If the knowledge of hypotheses of managers leads to *confirmatory* search, then one approach would be to ensure that market analysts don't know the exact hypotheses to be tested. For instance, to uncover the true drivers behind bank branch performance, analysts should be given branches with different performance, but without being told which is which.

While this would arguably increase the match between the learning and target settings, managers might resist the idea that analysts should be blind to their hypotheses. This is because, the results of less biased insights come at the cost of lower speed of analysis and less precise insights into the details of the hypotheses. Therefore, marketing managers would be better off taking the resulting insights of blind analyses as merely an initial step to build on – much like confirmatory factor analysis builds on exploratory factor analysis. As a practical example, evidence shows that US courts give more evidence weight to surveys where the researcher does not know the hypothesis being tested (Borek & Oza, 2018).

Another related approach would be to strive to actively refute the hypothesis instead of attempting to confirm it. That is, managers could look for evidence that confirms the opposite (or an alternative) hypothesis, as if that was the original hypothesis. For instance, Schmittlein and Morrison (2003) tried to disprove the hypothesis that in vitro fertilization providers offered money back guarantees in order to induce prospective parents to forego cheaper and less invasive procedures. When they ruled out several reasonable alternative explanations, they could state their case with increased and warranted confidence.

There are two important details to consider when conducting such a disconfirming analysis. First, given the potential conflicts of interest, this approach might require the marketing manager to declare to the analyst the opposite hypothesis as if it is the actual one. In this way, any confirmation bias now would work toward finding evidence that would prove the initial belief wrong. Second, such an analysis should be considered as a complementary tool to reveal where and how the initial beliefs of the managers can be erroneous and need to be updated. While this may seem a redundant or inefficient task at first glance, it would provide managers with clues as to what can go wrong with the strategies they are considering implementing. In fact, innovative firms like Pixar incorporate these redundancies to their processes to make sure that faulty assumptions are identified and eliminated before further (and often irreversible) investments are made on their projects (Catmull, 2008). However, the fact that Pixar is hailed as one of the few companies that could adopt such a rich communication scheme proves that it may be hard to implement in many organizations.

Regularly Updating Analyses to Reveal Shifts and Trends

To avoid illusions of *control* over the future based on knowledge of the past, a logical remedy would be frequent updating of the analysis to reveal shifts and trends. The former can catch the paradigm shifts and environmental changes that pose threats to the organization and its position in the market, while the latter can uncover opportunities as analysts extrapolate the trend in customer and competitor behavior. Such threats and opportunities can have a clear focus, such as the industry's fair value line, that is, the quantified and visualized relationship between product quality components and prices in the market.

Using easily available data, Pauwels and D'Aveni (2016) establish the primary quality dimension and the intercept, slope, and explanatory power of the fair value line at different times. They derive key insights from comparing shifts and trends in these fair value line components, such as (1) fair line elevation (erosion), including a higher (lower) willingness-to-pay for the baseline product in the industry; (2) line steepening (flattening), indicating customer are willing to pay more (less) for higher performance on the primary quality dimension; and (3) line tightening (blurring), indicating the primary quality dimension explains more (less) in products' prices. Eventually, a different quality component takes over as the new primary quality dimension and a new fair value line form, as did so several times in the automobile, sweetener, and electronics industries. Fig. 1 shows an example of fair value line elevation and flattening in the US market of pick-up trucks.

While important and recommendable, a weakness of this approach is that the regular updating of the analysis often remains passive. The analyst and decision-maker tend to wait for new data to come in and uncover trends or shifts in the market. However, the velocity of Big Data would be positively instrumental in this context, as rapid inflow of information would both motivate and facilitate regular updates.

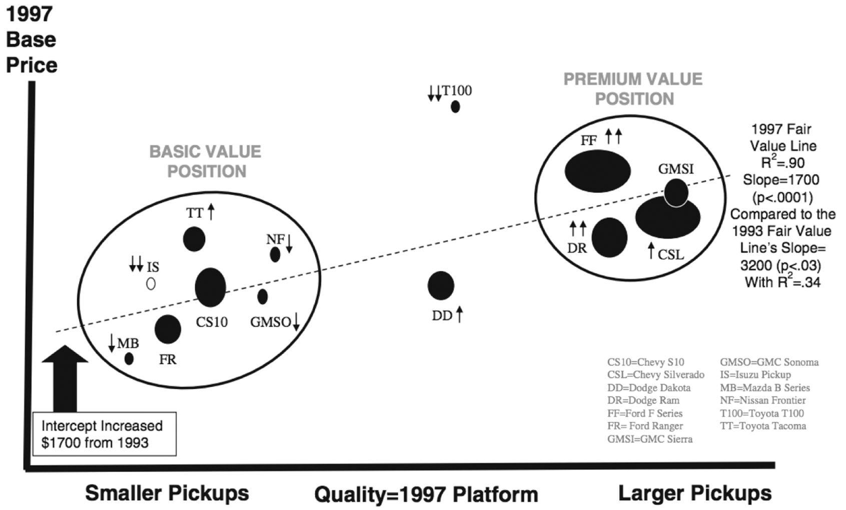

Fig. 1. Fair Value Line Elevation and Flattening for Pick-up Trucks 1993–1997 (Reproduced with First Author's Permission from Pauwels & D'Aveni, 2016).

Transparent Presentation of Statistical Insights

To mitigate mismatches due to *communication* of Big Data-generated insights between analysts and marketing managers, presentation modes need to be carefully designed. Graphical visualizations, for example, have been shown to help decision-makers better perceive uncertainties (Hanssens & Pauwels, 2016; Spiegelhalter, Pearson, & Short, 2011). In fact, plots may be designed to reveal individual observations, making sure that risks become more transparent (Hogarth & Soyer, 2013). For instance, for a large company marketing consumer durables, the heatmap in Fig. 2 visualizes expected profits (from high in green color to low in red color) from changing, respectively, price (*X*-axis) and TV advertising (*Y*-axis) (Pauwels, 2014).

A drawback of such visualizations is that only a few variables (two in Fig. 2) can be effectively visualized at the same time. For situations involving multiple factors and dimensions, simulations can be used to translate Big Data analysis into decision-friendly insights (Hogarth & Soyer, 2015). These tools let decision-makers enter their inputs to an interface and subsequently observe, summarize, and interact with the outcomes estimated by the sophisticated analyses conducted on the underlying Big Data. Experimental findings suggest that decision-makers make more accurate probabilistic judgments in situations involving uncertainties and complexities compared to those who have only access to descriptive statistics on the same analyses (Bradbury, Hens, & Zeisberger, 2014; Hogarth & Soyer, 2011; Sterman, 2011).

Fig. 3 shows slide bars that help decision-makers observe the predicted profit impact of changing multiple variables at the same time – adding product age, online advertising, and distribution share. More sophisticated dashboards with

Heat Map of the Interaction of Two Marketing Variables on Profits

TV advertising in thousands of $ \ Price in $	10	15	20	25	30	35	40	45	50	55	60	65	70	75
0	0.02	1.04	1.92	2.64	3.22	3.65	3.93	4.06	4.04	3.87	3.56	3.09	2.47	1.71
250	0.65	1.68	2.56	3.28	3.86	4.29	4.57	4.70	4.68	4.51	4.19	3.73	3.11	2.35
500	1.25	2.27	3.15	3.87	4.45	4.88	5.16	5.29	5.27	5.10	4.79	4.32	3.70	2.94
750	1.79	2.81	3.69	4.41	4.99	5.42	5.70	5.83	5.81	5.64	5.33	4.86	4.24	3.48
1000	2.28	3.30	4.18	4.91	5.48	5.91	6.19	6.32	6.30	6.13	5.82	5.35	4.73	3.97
1250	2.72	3.74	4.62	5.35	5.92	6.35	6.63	6.76	6.74	6.58	6.26	5.79	5.18	4.41
1500	3.11	4.13	5.01	5.74	6.32	6.74	7.02	7.15	7.13	6.97	6.65	6.18	5.57	4.80
1750	3.45	4.48	5.35	6.08	6.66	7.09	7.37	7.50	7.48	7.31	6.99	6.52	5.91	5.14
2000	3.74	4.77	5.65	6.37	6.95	7.38	7.66	7.79	7.77	7.60	7.28	6.82	6.20	5.44
2250	3.99	5.01	5.89	6.62	7.19	7.62	7.90	8.03	8.01	7.84	7.53	7.06	6.44	5.68
2500	4.18	5.21	6.08	6.81	7.39	7.81	8.09	8.22	8.21	8.04	7.72	7.25	6.64	5.87
2750	4.32	5.35	6.23	6.95	7.53	7.96	8.24	8.37	8.35	8.18	7.86	7.40	6.78	6.02
3000	4.42	5.44	6.32	7.05	7.62	8.05	8.33	8.46	8.44	8.27	7.96	7.49	6.88	6.11
3250	4.46	5.49	6.36	7.09	7.67	8.10	8.38	**8.51**	8.49	8.32	8.00	7.54	6.92	6.15
3500	4.46	5.48	6.36	7.09	7.66	8.09	8.37	8.50	8.48	8.31	8.00	7.53	6.91	6.15
3750	4.40	5.43	6.30	7.03	7.61	8.04	8.32	8.45	8.43	8.26	7.94	7.48	6.86	6.09
4000	4.30	5.32	6.20	6.93	7.50	7.93	8.21	8.34	8.32	8.15	7.84	7.37	6.75	5.99
4250	4.14	5.17	6.04	6.77	7.35	7.78	8.06	8.19	8.17	8.00	7.68	7.22	6.60	5.84
4500	3.94	4.97	5.84	6.57	7.15	7.57	7.85	7.98	7.97	7.80	7.48	7.01	6.40	5.63
4750	3.69	4.71	5.59	6.32	6.89	7.32	7.60	7.73	7.71	7.54	7.23	6.76	6.14	5.38
5000	3.38	4.41	5.29	6.01	6.59	7.02	7.30	7.43	7.41	7.24	6.92	6.46	5.84	5.08

Fig. 2. Visualizing the Profit Impact of Two Decision Variables in a Heatmap (Reproduced with Permission from Pauwels, 2014).

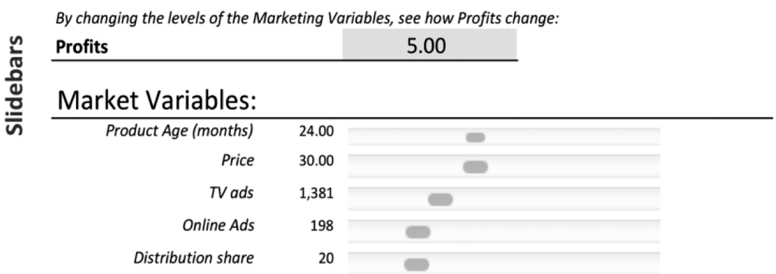

Fig. 3. Interactive Simulation for Predicted Performance Effect from Changing Levels of Several Marketing Variables (Reproduced with Permission from Pauwels, 2014).

drill-down menus enable users to observe and act on the relation between their decisions and relevant outcomes (Pauwels, 2014). These dashboards can also help with data veracity by integrating Big Data with "small data", for example, traditional surveys from a representative sample of consumers or qualitative insights from observation. Pauwels and van Ewijk (2013) show across 15 industries that such small data are less correlated (than online Big Data) with brand sales in the short run (same week) but predict brand sales better in the long run (several months out).

The most recent advancement in this regard are hybrid algorithms that not only show what the decision-maker requested or is familiar with but also show data-generated novel insights that can complement or even contradict current views (Pauwels, 2017). Such algorithms are already popular in movie recommendation systems (showing a mix of movies your profile suggest but also a few that are outside of the comfort zone) and should be adapted to advise marketing decision-makers.

There are, however, several issues associated with all these approaches. First, they all require managers to realize that they may be misperceiving and misinterpreting the insights generated by analysts. Yet if the feedback to their decisions is not immediate and/or there are uncertainties involved, they may not be able to easily identify the mismatches between what they learn and the target environment where they'll act. Second, the visualizations and simulations we described above need to be built and correctly used. This can often be costly endeavors and hence considered redundant, especially if managers are not aware of the extent of the miscommunication problem in the first place. Third, managers may need to be statistically sophisticated to appropriately use the more advanced communication tools available to them, but then gaining further statistical literacy may prove to be problematic (more on this below).

Statistical Literacy and Risk Intelligence

Statistical literacy is defined as "the ability to interpret, critically evaluate, and communicate about statistical information and messages" (Gal, 2002). It is considered a prerequisite for competent decision-making in many relevant domains, including medicine, law, and management (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007; Schield, 2011). While they may still be susceptible to biases, research suggests that decision-makers who are statistically sophisticated would be more willing to gain further experience and obtain feedback to better calibrate their confidence intervals around their predictions and decisions (Hogarth & Soyer, 2011). Evans (2012) refers to this ability to successfully regulate one's *confidence* in the face of uncertainty as risk intelligence.

However, a transformation that leads to high levels of statistical literacy and risk intelligence may prove unrealistic in marketing. First, it would undoubtedly be a costly and slow undertaking that would yield its benefits mainly in the long run (Wallman, 1993). Second, many marketing professionals may be attracted to the field due to its creative and artful stimulation of the more intuitive "System 1" thinking (Kahneman, 2011), giving less priority to statistical sophistication.

In fact, in a recent presentation, the Chief Marketing Officer of Sonos showed that creative marketers are often fearful of Big Data, given its power to shut down breakthrough ideas, those which typically don't test very well a priori (Pauwels, 2018). Hence, she argues for data to be used primarily (1) to inspire deviating from the norm and (2) to evaluate ideas after they are aired – to demonstrate their sales and profit impact. Indeed, when hiring analysts, Sonos specifically looks for humility and advocates for bringing up data insights only when needed and in a way that everyone can understand.

COMBINING BIG DATA WITH LEAN START-UP METHODOLOGY IN MARKETING MANAGEMENT

The lean start-up methodology is a scientific approach to conducting business, intended to help start-ups grow their businesses in the fastest and most efficient manner possible. It is an iterative methodology that relies on managers creating hypotheses and running multiple experiments to promptly test these. This continuous experimenting, with the use of minimal resources, would ultimately lead to a viable business model (Blank, 2013). We hypothesize that this methodology can also be used to mitigate the learning challenges from Big Data for marketing managers in an effective and efficient manner and to reduce the mismatch between learning and target settings. Crucially, we focus on the notion that lean approaches can help mitigate the weaknesses and difficulties associated with the more traditional solutions we reviewed above.

Rather than have incumbent firms adopt the lean start-up methodology, we argue instead that incumbent firms should combine the lean start-up methodology with Big Data analytics to aid with decision-making. Incumbent firms are different from start-ups regarding organizational structures, decision-making, resources, and so on. As such, it would be reckless for us to suggest that they adopt a methodology that is designed for start-ups. That said, in the area of marketing analytics, the adaptation of the lean start-up methodology by incumbent firms would lead to better decision-making and help to counter many of the problems that arise from the issues of confirmation, control, communication, and confidence.

Lean Start-up Methodology and Confirmations

The lean start-up methodology involves the continuous creation and testing of hypotheses through experiments. Hence, this approach is closely linked to research methods that actively look for disconfirming evidence regarding decision-makers' prior beliefs. Business education across the world provides managers with perspectives and tools that would guide them in their learning through experimentation.

Unfortunately, however, it is still the case today that many marketing departments in large firms do not conduct experiments (Ariely, 2010; Pfeffer & Sutton, 2006; Schrage, 2014). A new role of Big Data could then be to help develop hypotheses that can then be tested through experiments, rather than driving the decision directly. For example, if a firm wants to develop hypotheses regarding the pricing of their product/service to a particular customer segment, then Big Data analysis can be used to develop the hypothesis (e.g., *H1*: customers in the home shopping segment are willing to pay 10 dollars for next day delivery). In this way, firms use historical evidence to develop hypotheses about a specific component of the business model (in this instance pricing) and then use experiments to test the hypotheses stemming from the analysis.

Through such a combination of Big Data and lean start-up methodology, marketing managers would avoid coming to false conclusions as a result of statistical analyses. In particular, there would be less need for resorting to blinded testing or falsifications that require intricate communication policies. This is because, even if managers may prefer to look for supporting evidence from Big Data, the analysis under a lean start-up regime would only be used to produce hypotheses to be tested. So even if Big Data analysis and confirmation bias lead to the false conclusions, these would be turned into hypotheses that would promptly be disconfirmed (Seggie et al., 2017).

Lean Start-up Methodology and Control

A main factor that led to the development of lean start-up methodology was the unpredictable environment where start-ups have to operate. The methodology is designed in a way that entrepreneurs can take small yet concrete steps as they build their businesses, without needing to rely on strong convictions that may turn out to be false and subsequently lead to costly-yet-irreversible investments. For example, they may develop hypotheses about the price that certain customer segments are prepared to pay for a particular value offering. They would then proceed to quickly and roughly testing the hypotheses before implementing any long-term strategy, ensuring that their intuition eventually matches their reality. In this way, start-ups systematically reduce the potential negative effects of risk and uncertainty.

Prior to the adoption of lean start-up principles, entrepreneurs often engaged in a stealth form of product development (Villano, 2013). That is to say, they would develop their product or service in secret and spend substantial amounts of money to develop a finished product that the consumer may or may not be interested in buying given the price point. Compared the lean start-up way of launching a product, this is exceptionally risky because it does not allow the start-up to regularly test the different components of the business model with potential customers before launch.

The lean start-up methodology is thus envisioned as an iterative process of development that generally follows a so-called build-measure-learn loop (Ries, 2011). In this process, start-ups first strive to develop what is referred to a minimum viable product (MVP). The MVP is "a product with the minimum set of features that solves the problem of the customer" (p. 159, Seggie et al., 2017). This MVP is then tested with customers and the feedback gathered from the customers is used to improve the MVP, which is then tested with customers again.

For marketing managers, Big Data can once again primarily be a source of ideas to be tested, rather than insights on which decisions would be based. In particular, data can help managers design an initial marketing strategy, an MVP, to be tested in a limited and controlled capacity. These tests, in turn, would further generate data, which would then be used to design further ideas and tests. By adopting such a combination of methods, marketing managers would not passively wait for data to emerge, which would be then exploited to belatedly reveal shifts and trends. Instead, they would be actively exploring how these are evolving, potentially even contributing to those evolutions themselves.

Lean Start-up Methodology and Communication

There are two ways in which the lean start-up methodology helps alleviate issues due to communication of statistical findings in marketing management. First, the ideas tested tend to succeed each other in small steps, ensuring that the differences between results are more likely to be correctly understood. For instance, if there is extensive data on a previous marketing campaign and managers need to assess the implications of a potential update, then a lean approach would dictate that multiple tests be conducted on small updates, allowing decision-makers to easily perceive the relative effects of different changes.

Second, the experimentation under lean start-up methodology happens by default rapidly and frequently. This leads managers to obtain regular and immediate feedback on their assessments, predictions, and decisions, thereby ensuring that they become promptly aware of any systematic misperceptions due to communication methods. Under such a lean regime, often costly and analysis-specific visualization tools may become unnecessary, as the level of complexity of the statistical analyses remains low.

Lean Start-up Methodology and Confidence

Mitigating confirmation, control, and communication issues by combining Big Data analytics with lean start-up methodology would consequently help marketing managers calibrate their confidence accurately on the consequences of their predictions and decisions. Moreover, they would get the minimum required statistical knowledge on the job, as they apply experimental methods and explore the information that rapid and frequent testing generates. While using Big Data directly to inform decisions might offer a wide variety of opportunities to become overconfident, using it to generate testable hypotheses would once again constrain self-deceptions to levels that are warranted by the environment in which managers operate.

CONCLUSION

In this chapter, we ask (1) when and why the 4Vs of Big Data can lead to learning biases in marketing management? (2) What type of specific remedies can be employed to mitigate these and what would be the strengths and weaknesses of these? And (3) how could combining Big Data analytics with lean start-up methodology alleviate the learning challenges of marketing managers in an effective way?

We recommend that firms make use of multiple methods for managing potential biases. Starting with confirmation bias, we begin from a general position that firms should not use Big Data to reach conclusions but instead should use the results of analyses to guide hypothesis building and testing. We argue that it is beneficial for firms to incorporate and adapt some of the components of the lean start-up methodology to encourage firms to conduct experiments that can test the hypotheses generated by Big Data analysis. In this way, firms can address the potential dangers of confirmation bias where managers use Big Data to actively confirm their own biases. In addition, we argue that these experiments would ideally be conducted with the analyst not knowing the exact hypothesis to be tested. In this way, we can ensure that managers are not engaging in confirmatory search but instead truly engaging in objective testing.

For the issue of the illusion of control, we suggest that managers engage in constant updating of analysis to ensure that changes and also trends are captured. In this way, managers will be cognizant of the threats and also the opportunities that are likely to come to bear in the market.

To avoid communication problems, we argue that resources of both time and effort need to be applied to the design of the presentation of the results of the analyses. We argue for the use of graphical visualizations such as plots, simulations, and heatmaps to help the decision-makers better interact with the results of the Big Data analysis. Finally, we argue that managers require a degree of statistical literacy to better understand and interact with Big Data analysis.

To sum up, our suggestions are not a panacea for all the threats that accrue through the misuse of Big Data. However, – if analysts and marketing managers in firms engage in more thoughtful use of Big Data analytics in line with suggestions outlined above, then they will be able to get better insights and make better decisions.

NOTES

1. Including: www.smartdatacollective.com/benefits-big-data-in-marketing; www.digitaldoughnut.com/articles/2016/august/how-does-big-data-benefit-marketing; www.cleverism.com/best-uses-big-data-marketing
2. cmosurvey.org/
3. https://nathanbrixius.wordpress.com/2014/09/11/confirmation-bias-in-data-science/

REFERENCES

Ariely, D. (2010). Why businesses don't experiment. *Harvard Business Review*, *88*(4). AU:5

Armstrong, J. S. (2012). Illusions in regression analysis. *International Journal of Forecasting*, *3*, 689–694.

Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential? *Vikalpa*, *38*(4), 1–12.

Baron, J. (2000). *Thinking and deciding*. Cambridge University Press. AU:6

Blank, S. (2013). Why the lean start-up changes everything. *Harvard Business Review*, *91*(5), 63–72.

Borek, C., & Oza, A. (2018). Consumer surveys in trademark infringement litigation: FIJI vs VITI case study. In N. Mizik & D. M. Hanssens (Eds.), *Handbook of marketing analytics*. Edward Elgar Publishing. AU:7

Bradbury, M. A., Hens, T., & Zeisberger, S. (2014). Improving investment decisions with simulated experience. *Review of Finance*, *19*(3), 1019–1052.

Brian, C. (2000-02-27). *The A/B test: Inside the technology that's changing the rules of business wired business*. Wired.com. Retrieved from https://www.wired.com/business/2012/04/ff_abtesting/

Budescu, D. V., Broomell, S., & Por, H. H. (2009). Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological Science*, *20*(3), 299–308.

Catmull, E. (2008). How Pixar fosters collective creativity. *Harvard Business Review*. AU:8

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 1165–1188. AU:9

Christensen, C. M., Raynor, M. E., & McDonald, R. (2016). What is disruptive Innovation. *The Encyclopedia of Human-Computer Interaction*, 2.

Cooil, B., Keiningham, T. L., Aksoy, L., & Hsu, M. (2007). A longitudinal analysis of customer satisfaction and share of wallet: Investigating the moderating effect of customer characteristics. *Journal of Marketing*, *71*(1), 67–83.

Danaher, P. J., Roberts, J. H., Roberts, K., & Simpson, A. (2011). Applying a dynamic model of consumer choice to guide brand development at Jetstar Airways. *Marketing Science*, *30*(4), 586–594.

Du, J., Xie, L., & Schroeder, S. (2009). PIN optimal distribution of auction vehicles system: Applying price forecasting, elasticity estimation, and genetic algorithms to used-vehicle distribution. *Marketing Science*, *28*(4), 637–644.

Evans, D. (2012). Risk intelligence. In *Handbook of risk theory* (pp. 603–620). Springer. AU:10

Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research*, *2*(1), 28–32.

Fang, X., Sheng, O. R. L., & Goes, P. (2013). When is the right time to refresh knowledge discovered from data? *Operations Research*, *61*(1), 32–44.

Feiler, D. C., Tong, J. D., & Larrick, R. P. (2013). Biased judgment in censored environments. *Management Science*, *59*(3), 573–591.

Fischer, M., Albers, S., Wagner, N., & Frie, M. (2011). Dynamic marketing budget allocation across countries, products, and marketing activities. *Marketing Science*, *30*(4), 568–585.

Forrester. (2012). *The big deal about big data for customer engagement business: Leaders must lead big data initiatives to derive value*. Retrieved from http://www.forrester.com/The+Big+Deal+About+Big+Data+For+Customer+Engagement/fulltext/-/E-RES72241

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25.

Gartner. (2012). *Big data*. Retrieved from http://www.gartner.com/it-glossary/big-data/

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321–326.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96.

Hanssens, D. M., & Pauwels, K. H. (2016). Demonstrating the value of marketing. *Journal of Marketing*, *80*(6), 173–190.

Hoch, S. J., & Schkade, D. A. (1996). A psychological approach to decision support systems. *Management Science*, *42*(1), 51–64.

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science*, *24*(5), 379–385.

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus nontransparent description. *Journal of Experimental Psychology: General*, *140*(3), 434.

Hogarth, R. M., & Soyer, E. (June 2013). A picture's worth a thousand numbers. *Harvard Business Review*, 26.

Hogarth, R. M., & Soyer, E. (2015). Using simulated experience to make sense of big data. *MIT Sloan Management Review*, *56*(2), 49.

Isaac, W., & Dixon, A. (2017). How Big Bag data could make policing worse. *Fast Company*, May 11. Retrieved from https://www.fastcompany.com/40419894/how-big-bad-data-could-make-policing-worse

Johnson, G. A., Lewis, R. A., & Nubbemeyer, E. I. (2017). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, *54*(6), 867–884.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In *System sciences (HICSS), 46th Hawaii international conference* (pp. 995–1004). IEEE. AU:11

Kannan, P. K., Pope, B. K., & Jain, S. (2009). Pricing digital content product lines: A model and application for the National Academies Press. *Marketing Science*, *28*(4), 620–636.

Kaufmann, C., Weber, M., & Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management Science*, *59*(2), 323–340.

Klayman, J. (1995). Varieties of confirmation bias. In *Psychology of learning and motivation* (Vol. 32, pp. 385–418). Academic Press.

Kohavi, R., & Thomke, S. (2017). The surprising power of online experiments. *Harvard Business Review*, *95*(5), 74–82.

Kumar, V., Bhaskaran, V., Mirchandani, R., & Shah, M. (2013). Creating a measurable social media marketing strategy: Increasing the value and ROI of intangibles and tangibles for hokey pokey. *Marketing Science*, *32*(2), 194–212.

Kumar, V., & Shah, D. (2011). Uncovering implicit consumer needs for determining explicit product positioning: Growing prudential annuities' variable annuity sales. *Marketing Science*, *30*(4), 595–603.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387–394.

Laney, D. (2001). *3D data management: Controlling data volume, velocity and variety*. META Group Research Note, 6, 70.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., et al. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. AU:12

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205.

Lilien, G. L., Roberts, J. H., & Shankar, V. (2013). Effective marketing science applications: Insights from the ISMS-MSI practice prize finalist papers and projects. *Marketing Science*, *32*(2), 229–245.

Lukoianova, T., & Rubin, V. L. (2014). Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online*, *24*(1), 4–15.

Makridakis, S., Hogarth, R., & Gaba, A. (2009). *Dance with chance: Making luck work for you.* Oneworld publications.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, October, 3–9.

McCloskey, D. N., & Ziliak, S. T. (1996). The standard error of regressions. *Journal of Economic Literature*, *34*(1), 97–114.

Natter, M., Mild, A., Wagner, U., & Taudes, A. (2008). New tariffs at tele. ring: The application and impact of an integrated segmentation, targeting, and positioning tool. *Marketing Science*, *27*(4), 600–609.

Natter, M., Ozimec, A. M., & Kim, J. Y. (2015). ECO: Entega's profitable new customer acquisition on online price comparison sites. *Marketing Science*, *34*(6), 789–803.

Pauwels. (2017). *The future of market research*, December 22. Retrieved from https://analyticdashboards.wordpress.com/2017/12/22/the-future-of-market-research/

Pauwels. (2018). *The why behind the buy: Mad men + mad science*, April 14. Retrieved from https://analyticdashboards.wordpress.com/2018/04/14/the-why-behind-the-buy-mad-men-mad-science/

Pauwels, K. H. (2014). *It's not the size of the data, it's how you use it: Smarter marketing with analytics and dashboards*. New York, NY: AMACOM.

Pauwels, K., & D'Aveni, R. (2016). The formation, evolution and replacement of price–quality relationships. *Journal of the Academy of Marketing Science*, *44*(1), 46–65.

Pauwels, K., & Reibstein, D. (2010). Challenges in measuring return on marketing investment. In *Review of marketing research* (pp. 107–124). Emerald.

Pauwels, K., & van Ewijk, B. (2013). Do online behavior tracking or attitude survey metrics drive brand sales? An integrative model of attitudes and actions on the consumer boulevard. *Marketing Science Institute Working Paper Series*, *13*(118), 1–49.

Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., & Pauwels, K. (2013). Social media metrics – A framework and guidelines for managing social media. *Journal of Interactive Marketing*, *27*(4), 281–298.

Pfeffer, J., & Sutton, R. I. (2006). *Hard facts, dangerous half-truths, and total nonsense: Profiting from evidence-based management*. Harvard Business Press.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, *2*(1), 3.

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Books.

Rothwell, P. M. (2005). External validity of randomised controlled trials:"to whom do the results of this trial apply?". *The Lancet*, *365*(9453), 82–93.

Saleh, T., Lenhard, E., Goldon, N., & Opolon, D. (2015). *How to avoid the big bad data trap, BCG perspectives*. Retrieved from https://www.bcgperspectives.com/content/articles/big-data-digital-economy-how-to-avoid-big-data-trap/. Accessed on June 24, 2017.

Schield, M. (2011). Statistical literacy: A new mission for data producers. *Statistical Journal of the IAOS*, *27*(3, 4), 173–183.

Schmittlein, D. C., & Morrison, D. G. (2003). A live baby or your money back: The marketing of in vitro fertilization procedures. *Management Science*, *49*(12), 1617–1635.

Schrage, M. (2014). *The innovator's hypothesis: How cheap experiments are worth more than good ideas*. MIT Press.

Seggie, S. H., Soyer, E., & Pauwels, K. H. (2017). Combining big data and lean startup methods for business model evolution. *AMS Review*, *7*(3–4), 154–169.

Silva-Risso, J., & Ionova, I. (2008). A nested logit model of product and transaction-type choice for planning automakers' pricing and promotions. *Marketing Science*, *27*(4), 545–566.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Skiera, B., & Abou Nabout, N. (2013). Prosad: A bidding decision support system for profit optimizing search engine advertising. *Marketing Science*, *32*(2), 213–220.

Soyer, E., & Hogarth, R. M. (2012). The illusion of predictability: How regression statistics mislead experts. *International Journal of Forecasting*, *28*(3), 695–711.

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, *333*(6048), 1393–1400.

Sterman, J. D. (2011). Communicating climate change risks in a skeptical world. *Climatic Change*, *108*(4), 811.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (Vol. 2). Random house.

Taleb, N. N., & Goldstein, D. G. (2007). We don't quite know what we are talking about when we talk about volatility. *Journal of Portfolio Management*, *33*(4).

Taleb, N. N., & Goldstein, D. G. (2012). The problem is beyond psychology: The real world is more random than regression analyses. *International Journal of Forecasting*, *28*(3), 715–716.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Villano, M. (2013). Retrieved from https://www.entrepreneur.com/article/229461. Accessed on April 15, 2018.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88*(421), 1–8.

Ziliak, S. T. (2012). Visualizing uncertainty: On Soyer's and Hogarth's "The illusion of predictability: How regression statistics mislead experts". *International Journal of Forecasting*, *28*(3), 712–714.

AUTHOR QUERY FORM

Emerald	**Book: RMR-V016-3611843** **Chapter: CH004**	**Please e-mail or fax your responses and any corrections to:** **E-mail:** **Fax:**

Dear Author,

During the preparation of your manuscript for typesetting, some questions may have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

Disk use

Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

☐ Scanning (parts of) your article ☐ Rekeying (parts of) your article

☐ Scanning the artwork

Bibliography

If discrepancies were noted between the literature list and the text references, the following may apply:

☐ **The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.**

☐ ***UNCITED REFERENCES:*** **This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.**

Queries and/or remarks

Location in Article	Query / remark	Response
AU:1	Please check and confirm the shortened running head.	
AU:2	As per style, a minimum of 6 keywords are required. Please check and provide.	
AU:3	Reference "Davenport and Harris (2007); Simon (2015); O'Neil (2016); Wiesel et al. (2011)" has not been provided in the list. Please provide the reference details.	
AU:4	Please provide the expansion for the abbreviation CMO.	
AU:5	Please provide the page range for Ariely, D. (2010); Taleb, N. N., & Goldstein, D. G. (2007).	

AU:6	Please provide the publisher location for Baron, J. (2000); Kahneman, D. (2011); Makridakis, S., Hogarth, R., & Gaba, A. (2009); Pfeffer, J., & Sutton, R. I. (2006); Ries, E. (2011); Schrage, M. (2014); Taleb, N. N. (2007).	
AU:7	Please provide the page range and publisher location for Borek, C., & Oza, A. (2018).	
AU:8	Please provide the volume and page range for Catmull, E. (2008).	
AU:9	Please provide the volume for Chen, H., Chiang, R. H., & Storey, V. C. (2012); Christensen, C. M., Raynor, M. E., & McDonald, R. (2016); Hogarth, R. M., & Soyer, E. (June 2013); McAfee, A., & Brynjolfsson, E. (2012).	
AU:10	Please provide the editors name and publisher location for Evans, D. (2012); Klayman, J. (1995); Pauwels, K., & Reibstein, D. (2010).	
AU:11	Please provide the editors name and publisher details for Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January).	
AU:12	Please provide the complete author names for Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., et al. (2018).	