iS
CHANNEL

Big Data and Its Implications

An Overview of Managerial Implications from Technical-Rational and Socially Embedded Perspectives

Maximilian Mende
MSc Management, Information Systems and Digital Innovation (2014/15)
Department of Management
London School of Economics and Political Science

KEYWORDS

Big Data
Bounded Rationality
Managerial Rationality
Social Embeddedness

ABSTRACT

Big Data has been shown to have strong influence on decision making processes, and organizational setups across all kinds of organizations. This critical literature review draws on peer-reviewed literature to identify scholars' perspectives on different implications, and questions the assumption that today's data offers the basis for completely rational decision making. Therefore process-inherent boundaries like irrational human interaction and data quality issues are identified. Apart from the technical-rational dimension, Big Data has an impact on the social environment. This includes questions about the ethical legitimacy of data accumulation and usage. The literature review concludes that there is further shaping of organizations necessary, in order to leverage the potential Big Data offers. Lastly it addresses a number of research questions, which bear the potential to further develop this field of study.

Introduction

During the 2002 Major League Baseball (MLB) season, the until then mediocre team of Oakland Athletics surprisingly won the national championship despite a below average budget. The secret to their success was the deviation from the traditional experience-based decisions in the transfer market, to an approach based on rigorous statistical analysis. Their success proved them right, so that today this methodology is referred to as sabermetrics. This is the combination of SABR, which is the acronym for the *Society of American Baseball Research* and metrics, which highlights the analytical component. Michael Lewis was following the team during its 2002 season and packaged the insights into his best-selling novel *Moneyball* (Lewis 2004).

This is only one very prominent example of Big Data significantly impacting management decision making. This paper aims to provide an overview of different perspectives on managerial rationality, and outline interesting avenues for future research in this regard. The term managerial rationality refers in this context to the underlying drivers that define the decision making process as well as the setup of an organization. Working definitions for technical-rationality, bounded-rationality and social embeddedness will be given throughout the paper.

Corresponding Author
Email Address: Maximilian.Mende@capgemini.com (M. Mende)

The literature review is structured as follows. Section one provides an overview about different perspectives on beneficial implications of Big Data on managerial rationality from a formal technical-rational perspective. In the following section, inherent boundaries to the rational implications as proposed by the literature are highlighted. Section three looks at the detrimental implications from a socially embedded perspective. Finally, future research questions are addressed, which have the potential to advance the scientific insights in this field.

Although the impact of Big Data is significantly broader than is examined in this critical literature review, in order to give a detailed insight, it will focus only on managerial rationality. Further research should discuss also the implications from an engineering- and economic-rationality point of view.

Implications of Big Data on Managerial Rationality from a Formal Technical-Rational Perspective

Formal technical-rationality refers to the perspective that corporate behavior is completely rational. It is derived from 'best practice', which can be proven with mathematical evidence. For example, in a rational company every investment decision is justified by sound mathematical models and is linked into the general business strategy.

Implications for Decision Making

Multiple authors who adopt this view implicitly assume that Big Data heavily influences the process of decision making. It is assumed to shift from what McAfee et al. call a hippo approach – highest paid person's opinion (McAffee et al. 2012) – towards one based largely on algorithms and mathematical patterns. These algorithms are used to predict the future and thus to define business strategy (LaValle et al. 2011; Andrejevic 2014). Chen, Andrejevic and LaValle build on this view stating that this shift is applicable in any context independent of function or industry (Chen, Chiang & Storey 2012; Andrejevic 2014; LaValle et al. 2011).

Anderson and Rosenzweig even extend this point to a more absolute level. In their opinion, the statistical examination of Big Data is superior to, and should replace, any other legacy decision making tool (Anderson 2008; Rosenzweig 2014). They point out that a decision making process based on Big Data is superior to traditional methods, as it prevents inherent human biases (Rosenzweig 2014). The argument of flaws emerging from human interaction in heuristics is supported by Goldstein and Gigerenzer: the authors state that at some point problems and solutions consist of too many variables to be grasped by human cognitive abilities, leading to mistaken decisions (Goldstein & Gigerenzer 2009).

The prevalent assumption behind the above outlined views is that underlying data sets are neither flawed nor biased. This is justified by Mayer-Schönberger and Cukier, as well as Booch's view that the mass of data can eliminate any flaws in its subsets, hence bringing us closer to objective truth (Mayer-Schönberger & Cukier 2013; Booch 2014). However the essay will further examine potential boundaries to this assumption at a later point.

Big Data supports decision making	
La Valle et al. (2011)	– In the future algorithms will be used, to help define business strategy – This shift is applicable independent of function or industry
Andrejevic et al. (2014)	
Chen et al. (2012)	
Big Data should replace legacy ways of decision making	
Rosenzweig (2014)	– Decisions based on Big Data are superior to and should consequently replace traditional decision making techniques
Anderson (2008)	
Goldsetein & Gigerenzer (2009)	– Limited human cognitive abilities promote use of Big Data
Big Data provides flawless insights	
Mayer-Schönberger & Cukier, 2013	– Combining data sets and leveraging the bulk of data can eliminate flaws in subsets
Booch, 2014	

Table 1. Classification of Scholars - Decision Making Benefits of Big Data

Implications for Organizational Design

In general, managerial rationality is not limited to decision making processes; furthermore it also takes into account effective organizing in order to achieve improved organizational performance. This includes, but is not limited to, setting rules and norms, obligations and roles, as well as hierarchical structures in an organization. LaValle et al. (2011) identified the impact of Big Data on this specific part of managerial rationality. They claim that insights gathered from internal data on process efficiency can be applied to streamline the internal flows of an organization. This is supported by McAffee et al.'s empirical research showing that companies applying Big Data are performing better in regard to productivity and profitability (McAffee et al. 2012). Pentland's case study shows the effect of Big Data in action. He outlines an analysis of the timing of coffee breaks and its subsequent optimization, which led to significant efficiency increases for a call center operator (Pentland 2013).

McAfee et al. and LaValle et al. examine the organizational implications in more detail (McAffee et al. 2012; LaValle et al. 2011). They highlight the need to set up suitable organizational structures and build strong analytical capabilities within the workforce, in order to support the data driven organization. This can be considered a technology deterministic point of view, where the social and organizational context follows technical advances.

Linked to those organizational changes McAfee et al. and the Journal of Strategic Direction see an additional implication, stating that not only will the processes and organization change, but also the distribution of power in the executive boards will shift towards the positions responsible for holding and analyzing data (McAffee et al. 2012; Strategic Direction 2012).

Pentland takes an even more radical view on Big Data's implications, and expands the technology deterministic view that Big Data will not only shape organizational processes, but it will go beyond and shape entire organizations, cities and governments, predicting that those will be significantly more efficient in the aftermath (Pentland 2013). This view seeing the impact of Big Data beyond the organization is shared by Brown et al., who assume that analytics will determine how companies and also nations compete and prosper (Brown, Chui & Manyika 2011).

The implications of Big Data on managerial rationality offered are twofold: first it will drive executives into a more rational decision making through mathematically supported insights and guidelines. The most absolute position in this context is taken by Rosenzweig, proposing that organizations should rely on decisions based solely on Big Data, as this would eliminate any human bias (Rosenzweig 2014). The second implication offered is that process redesign and organizational reorganization based on insights gained from Big Data will significantly increase efficiency and productivity in an organization. The more absolute perspective in this case is led by McAfee and Pentland, who claim that Big Data will not only affect organizations, but is basically applicable in any context to improve efficiency and shift power distributions (McAfee et al. 2012; Pentland 2013).

Big Data helps build stronger organizations	
La Valle et al. (2011)	– Analysis help restructure organizations
McAffee et al. (2012)	– Empirical studies show Big Data increases performance
Pentland (2013)	– Case study where Big Data improved performance – Big Data will not only improve corporate performance, but shape companies, cities and countries' competitive positions
Brown et al. (2013)	– Big Data goes beyond the organization and will define how corporations and nations prosper
Big Data requires new organizations and capabilities	
McAffee et al. (2012)	– To leverage the full potential of Big Data, companies need to build new analytics capabilities and reshape their organization
La Valle et al. (2011)	

Table 2. Classification of Scholars - Organizational Benefits of Big Data

Inherent Boundaries to Implications of Big Data on Managerial Rationality

The previous section outlined the implications of Big Data on an organization under the assumption that actors behave completely rationally. Still this view might not be universally applicable, as multiple scholars highlight constraints to managerial rationality based on Big Data. The views through this lens can be classified as bounded rationality, which depicts limited cognitive abilities of people and limited capabilities of technology.

Generally the identified boundaries can be grouped into two broad categories: the first is based on misapplication of tools or human error, whereas the second is emerging from the constraints inside the underlying data sets.

Boundary from Misapplication

In regard to misapplication of data, scholars identified two common patterns: first Bollier and Fioramonti independently from one another argue that any data is exposed to assumptions and human filtering. In their opinion, this is a necessary phenomenon, as the amounts of structured and unstructured data are just too vast to be fully analyzed. Hence human filtering is required to reduce complexity and make data usable in the context of Big Data (Bollier 2010; Fioramonti 2014).

The second limitation is identified by Boyd and Crawford, as well as Leinweber. Those scholars highlight the phenomenon of apophenia in the analysis of data (Boyd & Crawford 2012; Leinweber 2007). Apophenia refers to the idea that statistically significant patterns in data are identified, that are only coincidental. Leinweber (2007) refers in this case to the example of a statistically significant relationship between the development of the S&P500 index in the USA and butter production in Bangladesh, which is to our current knowledge not directly interlinked.

Considering the above boundaries of human filtering and over-interpretation, further research will be required in order to validate the degree to which Andrejevic (2014) and LaValle et al.'s (2011) claim that patterns identified in Big Data shape strategies and decision making is true. In addition the view of Rosenzweig (2014) that Big Data is eliminating human bias needs reconsideration, taking into account that human bias could affect the data already before its automated analysis.

Boundary from Data Quality

The second set of boundaries emerges from issues in the quality of underlying data sets. Boyd and Crawford (2012) state that often sources underlying the decision making process are flawed and do not provide statistically significant data sets. They provide the example of data gathered from social media sites. The authors claim, that this kind of data cannot be statistically accurate to the degree implied by technical-rational scholars, as for example social media is assumed not to be used by a statistically significant sample of people. This view is amplified by Wigan and Clarke (2013), who take into account the process of building data sets where different data from various sources and with different original purposes is combined. The authors claim that often data sets seem complementary, even though they are really not and by combination provide erroneous results. In addition, they highlight that the combination of different quality data sets reduces the overall quality to the level of the set with the lowest quality. This is a strong contrast to the objective truth implied by Mayer-Schöneberger & Cukier and Booch, who justified the rationality behind decisions based on Big Data with the elimination of data quality issues through the accumulation and combination of multiple data sets (Mayer-Schönberger & Cukier 2013; Booch 2014).

According to Wigan and Clarke (2013), one method to prevent wrong decisions based on error prone data is to verify the results through additional validation. However at the same time taking an economic perspective, the authors state that due to cost and time lags incurred by advanced validation those steps are most of the time not taken by organizations. This in turn leads in the worst case to wrong decisions based on low quality data. Still this solution requires further elaboration, as even the best additional analysis cannot be unbiased, when based on flawed data.

Taking the managerial rationality perspective again, Wigan and Clarke (2013) claim that although wrong, those decisions are most of the time authorative, due to the setup in organizations. This point is supported by Rosenzweig, who identified that the growing popularity of sophisticated statistical approaches promotes a mentality of blind-trust into numbers, rather than one that promotes critical thinking and a mentality of questioning their validity (Rosenzweig 2014). On the other hand Redman (2013) claims that this mentality of relying on decisions supported only by data is often not sustainable. He observed that once a significant decision is taken wrong, managers tend to completely abolish their positive feeling towards data based decisions and return to processes based radically on gut feeling and experience. This issue implies that with a growing influence of Big Data corporate mentality needs to be redefined (Redman 2013).

Drawing on the organizational setup, Redman (2013) and the Strategic Direction Journal (2013) identify an additional boundary to the rational implications of Big Data. They state that already today knowledge workers tend to work 50% of their time correcting errors and trying to validate given data sets. According to a survey based study of the Strategic Direction Journal (2013) most internal IT functions today are unable to cope with the vast amounts of data. This view however does not contrast the implications seen by the technical-rational scholars, furthermore it supports their perspective that one implication of Big Data will be the reshaping of the organizational setup. In addition it supports the point made earlier, that organizations today do not possess the necessary capabilities to make sense of the benefits Big Data could offer (McAfee et al. 2012; Pentland 2013).

Misapplication compromises influence of Big Data	
Bollier (2010)	– Any data is subject to assumptions and filtering, in order to reduce complexity – this creates space for human bias
Fioramonti (2014)	
Boyd & Crawford (2012)	– Risk of apophenia is deeply embedded in Big Data
Leinweber (2007)	
Data Quality compromises influence of Big Data	
Boyd & Crawford (2012)	– Data sources today (e.g. Social Media) are unreliable
Wigan & Clarke (2013)	– Often really incompatible data sets are combined, limiting significance of results – Combination of flawed data sets reduces overall quality
Organizational culture compromises influence of Big Data	
McAffee et al. (2012)	– To leverage the full potential of Big Data, companies need to build new analytics capabilities and reshape their organization
Redman et al. (2013)	– Mentality in most companies does not yet accept Big Data for decision making or quickly abolishes it

Table 3. Inherent Boundaries to Application of Big Data

A practical example where all aforementioned biases were present is given by Hoffmann and Podgurski from the public health sector. First they highlight existing errors in data capturing and processing, due to workload and misunderstanding, subsequently they identify filtering errors in the selection of relevant data sets for electronic health records (Hoffman & Podgurski 2013).

Finally, it is important to see that none of the authors supporting a bounded rationality in this case completely contradicts the beneficial implications promoted from a technical-rational point of view. Still they highlight significant drawbacks that limit the rationality of data based organizations and put forward questions that should be considered in future research.

Implications from Big Data originating from the social Environment

Big Data is not a stand-alone phenomenon, au contraire: it is deeply embedded in the context of today's social environment. Data privacy is the most prominent theme in this regard. It links to the question about ethical norms and guidelines for the accumulation and use of data.

Considering the existence of norms for data accumulation, one statement of Boyd and Crawford is central: "Just because it is accessible, does not make it ethical" (Boyd & Crawford 2012, p. 671). As Davenport (2011) outlines in his case study, the basic set of ethics in regard to data is most often company specific. According to Booch (2014) these guidelines are currently not clearly defined on an international level, as only in some places like the European Union strong data privacy regulations are applied. The paradox leading to the inherent complexity of this topic is described by Nissenbaum (2010): He states that today privacy is mostly contextual, whereas data, as seen in the previous section by Wigan and Clarke, Müller-Schönberger and Cukier, as well as Booch, is often de-contextualized, which in turn compromises its privacy pledges (Wigan & Clarke 2013; Mayer-Schönberger & Cukier 2013; Booch 2014). A very severe case of this paradox is outlined in Davenport's case study, where the decision is at hand to sell customer data thus compromise customer relationships and privacy for financial benefits (Davenport 2011). A question implied by this discussion about data usage is about its ethical boundaries. Considering again the example of public health, the sharing of electronic health records could have the potential to greatly advance human understanding of different diseases; still it would imply disclosure of the most private kinds of personal data. A potential solution to this problem could be the usage of anonymous data. However this view can also be considered flawed, considering Booch's assumption that in the future the vast amounts of data and their inter-linkages will give the possibility to retrace any kind of data based on only two or three variables (Booch 2014). In order to further evaluate this topic, future research should engage in the question which scope for national or even international data privacy laws would be desirable.

Big Data brings ethical / privacy issues	
Boyd & Crawford (2012)	– Availability and usage of data do not necessarily go hand in hand
Davenport (2011)	
Booch (2014)	– There is a lack of data privacy guidelines and international collaboration on this topic
Nissenbaum (2010)	– Problems emerge from the paradox of contextual nature of private data and its de-contextual use in companies

Table 4. Social Issues of Big Data

Conclusion and further Research Questions

The view on Big Data with a focus on managerial rationality promotes three implications: firstly, Big Data can be used to support organizational decision making. However, the claim of scholars to completely replace legacy decision making tools and rely solely on analytics should be refuted, as this essay identified persisting problems of data quality, which are linked to human filtering, as well as the combination of data sets. In addition, the current organizational capabilities and mentalities show a need for transformation for organizations to be able to fully make sense of available data. This is the second implication and it includes not only the potential for streamlining existing organizations, but also building new organizational capabilities in order to improve human filtering mechanisms and prevent apophenia. Lastly, the ethical implication of Big Data persists, which focuses on the misuse of private data. This can also be considered a managerial implication, as it has lead and will lead to norms and guidelines that will shape the new data driven organization. During the course of this critical literature review questions have been identified that could drive the future research agenda in this field:

- Does the combination of data sets, like Mayer-Schönberger and Cukier (2013), as well as Booch (2014) state, provide an objective truth, or does it compromise the overall data quality like Wigan and Clarke (2013) indicate? Again this question might be highly context specific, as the nature of examined data sets is determining if results are improved or compromised.
- To what extent does human filtering compromise the rationality of Big Data? This includes the question if despite improving processing capabilities this step of complexity reduction will still be required in the future, as well as the question whether human filtering compromises or complements the analysis conducted by machines? As this question will be highly context specific, it would be reasonable to assess it under variable circumstances.
- Which options exist for data scientists to increase prediction accuracy and prevent apophenia in Big Data analysis, building on the idea of additional validation runs, by Wigan and Clarke (2013)?
- Which impact does Big Data have on corporate mentality, and which steps need to be taken to promote sustainable decision processes based on Big Data? This question could be extended by the research of Goldstein and Gigerenzer (2009), who are contrasting heuristic and statistical approaches to decision making.

References

Anderson, C 2008, 'The end of theory', Wired Magazine, vol 16, no. 7, pp. 7 - 16.

Andrejevic, M 2014, 'Big Data, Big Questions | The Big Data Divide', International Journal of Communication, vol 8, p. 17.

Bollier, D 2010, The Promise and Peril of Big Data, viewed 01 December 2014, <http://www.aspeninstitute.org/publications/promise-peril-big-data>.

Booch, G 2014, 'The Human and Ethical Aspects of Big Data', IEEE Software, vol 31, no. 1, pp. 20 - 22.

Boyd, D & Crawford, K 2012, 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon', Information, Communication & Society, vol 15, no. 5, pp. 662 - 679.

Brown, B, Chui, M & Manyika, J 2011, 'Are you ready for the Era of 'Big Data'', McKinsey Quarterly, vol 4, pp. 24 - 35.

Chen, H, Chiang, HR & Storey, VC 2012, 'Business Intelligence and Analytics: From Big Data to Big Impact', MIS Quarterly, vol 36, no. 4, pp. 1165 - 1168.

Davenport, TH 2011, 'The Dark Side of Customer Analytics', Harvard Business Review, vol 85, no. 5, p. 37.

Fioramonti, L 2014, How numbers rule the world: The use and abuse of statistics in global politics, Zed Books, London.

Goldstein, DG & Gigerenzer, G 2009, 'Fast and Frugal Forecasting', International Journal of Forecasting, vol 25, no. 4, pp. 260 - 772.

Harding, S 1996, Feminism, science, and the anti-enlightenment critiques, Routledge, New York, NY.

Hoffman, S & Podgurski, A 2013, 'Big bad data: law, public health, and biomedical databases', The Journal of Law, Medicine and Ethics, vol 41, no. Supplement, pp. 56 - 60.

LaValle, S, Lesser, E, Shockley, R, Hopkins, MS & Kruschwitz, N 2011, 'Big data, analytics and the path from insights to value', MIT Sloan Management Review, vol 21.

Leinweber, DJ 2007, 'Stupid data miner tricks: overfitting the S&P 500', The Journal of Investing, vol 16, no. 1, pp. 15 - 22.

Lewis, M 2004, Moneyball: The art of winning an unfair game, W. W. Norton & Company Inc., New York, NY.

Manovich, L 2011, 'Trending: The Promises and the Challenges of Big Social Data', in Debates in the Digital Humanities, University of Minnesota Press, Minneapolis.

Mayer-Schönberger, V & Cukier, K 2013, Big data: A revolution that will transform how we live, work, and think, Houghton Mifflin Harcourt Publishing Company, New York, NY.

McAffee, A, Brynjolfsson, E, Davenport, TH, Patil, DJ & Barton, D 2012, 'Big Data - the Management Revolution', Harvard Business Review, vol 90, no. 10, pp. 61 - 67.

Nissenbaum, HF 2010, Privacy in context: Technology, policy, and the integrity of social life, Stanford University Press, Palo Alto.

Pentland, A 2013, 'The Data-Driven Society', Scientific American, vol 309.

Redman, TC 2013, 'Data's Credibility Problem', Harvard Business Review, vol 91, no. 12, pp. 84 - 86.

Rosenzweig, P 2014, Left Brain, Right Stuff: How Leaders Make Winning Decisions, Public Affairs, Philadelphia, PA.

Strategic Direction 2012, 'Big data needn't be a big headache: How to tackle mind-blowing amounts of information', Strategic Direction, vol 28, no. 8, pp. 22 - 24.

Strategic Direction 2013, 'So, are the geeks inheriting the earth?: Big Data, big decisions and best practice for information management', Strategic Direction, vol 29, no. 9, pp. 12 - 15.

Wigan, MR & Clarke, R 2013, 'Big data's big unintended consequences', Computer, vol 46, no. 6, pp. 46 - 53.