

Article

## Datafication and the Seductive Power of Uncertainty —A Critical Exploration of Big Data Enthusiasm

Stefan Strauß

Institute of Technology Assessment (ITA), Austrian Academy of Sciences, Strohgasse 45/5, Vienna A-1030, Austria; E-Mail: sstrauss@oeaw.ac.at; Tel.: +43-1-51581-6599

Academic Editors: Mark Burgin and Wolfgang Hofkirchner

Received: 15 October 2015 / Accepted: 26 November 2015 / Published: 9 December 2015

---

**Abstract:** This contribution explores the fine line between overestimated expectations and underrepresented momentums of uncertainty that correlate with the prevalence of big data. Big data promises a multitude of innovative options to enhance decision-making by employing algorithmic power to gather worthy information out of large unstructured data sets. Datafication—the exploitation of raw data in many different contexts—can be seen as an attempt to tackle complexity and reduce uncertainty. Accordingly promising are the prospects for innovative applications to gain new insights and valuable knowledge in a variety of domains ranging from business strategy, security to health and medical research, *etc.* However, big data also entails an increase in complexity that, together with growing automation, may trigger not merely uncertain but also unintended societal events. As a new source of networking power, big data has inherent risks to create new asymmetries and transform possibilities to probabilities that can *inter alia* affect the autonomy of the individual. To reduce these risks, challenges ahead include improving data quality and interpretation supported by new modalities to allow for scrutiny and verifiability of big data analytics.

**Keywords:** technology assessment; uncertainty; predictability; interpretation; autonomy; power asymmetries; data pragmatism

---

### 1. Introduction

Big data is among the greatest hypes related to information and communication technologies (ICT). It promises a multitude of innovative options to enhance decision-making by employing algorithmic power to gather worthy information out of unstructured data sets. The field of applications ranges from

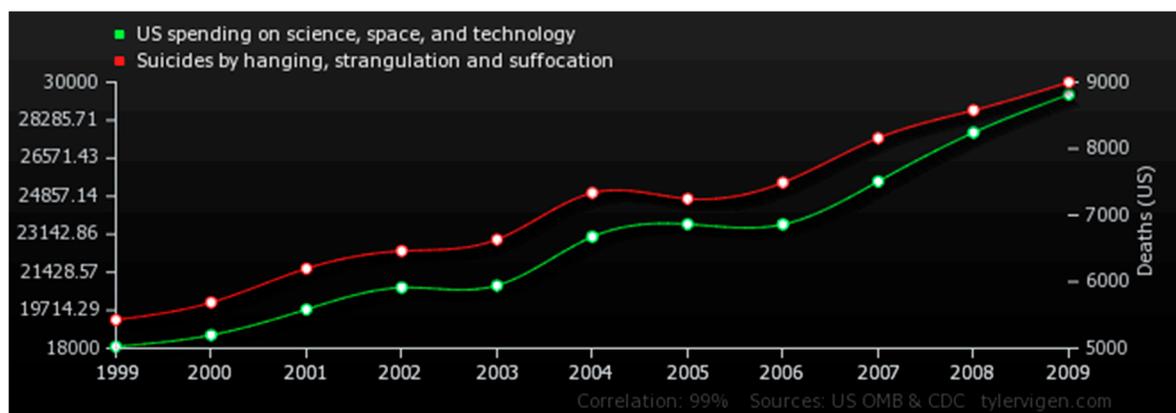
business process optimization, demand-oriented energy supply, market- and trend forecasting, uncovering illegal financial transactions, predictive policing, enhanced health research by analyzing population diseases, cancer research, software-supported medical diagnosis, *etc.* Hence, big data bears a lot of potential to support societal well-being [1,2]. Big data enthusiasts even tend to present it as a leap in evolution away from the stone age of unstructured data sets far ahead in the age of sophisticated algorithms and data visualization. Exploiting petabytes of data is framed as a remedy to deal with complexity and reduce uncertainty by paving the way for predictive analytics [3].

A number of trends and drivers promoted the emergence of big data: technological developments such as social networks, mobile devices, cloud computing, apps, machine-to-machine communication, smart technologies, *etc.* entail an increase in the processing and availability of digital information. In the economic and business sector, big data has been known for some years already under labels such as business intelligence or enterprise 2.0 aiming at taking strategic advantage of digital information for novel business models. In a societal context, the increase of social media and informational self-exposure and trends such as the “quantified self”, *etc.* also contribute to a further growth in digitally available personal data across many different domains. Last not least, also political trends such as securitization, increasing emphasis on pre-emptive and preventive security measures and developments towards so-called predictive policing foster big data. Put together, big data is closely linked to so-called “datafication” [1,4] aiming at gathering large amounts of every-day-life information to transform it into computerized, machine-readable data. Once digitized, algorithms then can be fed with the data in order to unleash the assumed enormous assets hidden in the large amounts of information. Accordingly, big data is often defined as “high-volume, -velocity, -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making” [5]. This definition mirrors the strong role IT-marketing plays in the big data discourse as it puts emphasis on presenting big data as a novel form of information processing that efficiently enriches decision-making. Less mystifying, [6] define big data as “a cultural, technological, and scholarly phenomenon” that rests on the interplay of technology, analysis and mythology. The latter addresses the “widespread belief that large data sets offer a higher form of intelligence and knowledge to generate insights previously impossible with the aura of truth, objectivity and accuracy” (p. 663, [6]). This dimension of mythology is of particular interest in this contribution that critically explores some of the major claims of big data enthusiasm. The analysis presents several examples to elaborate on the power of big data analytics and its capability to challenge the interpretation of its results. The aim here is to point out important fallacies and societal challenges in order to contribute to a deeper understanding and reasonable use of big data technology. This paper is structured as follows: Section 2 provides an overview on the meaning of big data and its pragmatic approach in exploiting information; Section 3 deals with the role of uncertainty and discusses the predictive capacity of big data to enhance decision-making; Section 4 elaborates the power of big data and its relation to autonomy and the final Section 5 provides a summary and some concluding remarks.

## 2. Meaning and Mythology of Big Data

A major claim of big data is that the exploitation of large, messy data sets allows analysts to gain more insights in a natural/self-evident way as “[w]ith enough data, the numbers speak for themselves” [7]. This

claim follows a “the bigger the better” logic that metaphorically speaking suggests to consider the whole haystack as a gold mine instead of just searching for a needle. In line with this delusive view is the perception that data quality decreases in importance and finding a correlation is key to come to better decision making [1,2]. In this regard, big data seems to be embraced by a mystique of seemingly causal correlations neglecting the fact that correlation is not causation. Revealing patterns and exploring correlations can of course be very helpful in a number of contexts *inter alia* for medical research and in the health sector; e.g., by showing yet hidden interrelations between symptoms of different diseases, exploring side effects of drugs, *etc.* This can support medical treatment and benefit forecasting and early warning; analysis of anonymized health data about the population can contribute to explore how diseases develop over time such as spreading of cancer and so on. However, this potential is not *per se* given and a critical reflection on the claims of big data seems to be important to come to a clearer understanding of the prospects and limits of big data. The following example in Figure 1 shows a correlation of nearly 100% between the US spending on science, space and technology and suicides by hanging, strangulation and suffocation.



**Figure 1.** Spurious correlation.

This correlation (provided by tylerjvigen.com) is obviously complete nonsense and aims here to highlight the misleading power of spurious correlations that is inherent in big data technology. The requirements for the interpretational performance to understand what the big data results are actually providing as well as their limits are not in each case as simple as here (as will be exemplified in the following sections).

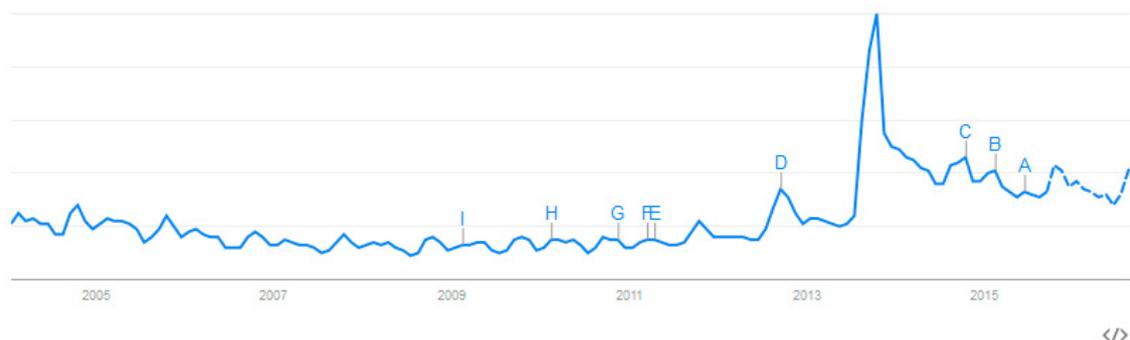
Behind the scenes of the big data mystique and related trends, there might be a new paradigm of data pragmatism on the rise: “Algorithmic living is displacing artificial intelligence as the modality by which computing is seen to shape society: a paradigm of semantics, of understanding, is becoming a paradigm of pragmatics, of search” [8]. This new data pragmatism gathers data from many different sources to look for hidden patterns and win new insights. Actually, big data is mainly based on algorithms that use the so-called mapreduce programming models to analyze the basic structure of data and then aggregate parts of the structure. This allows for fast analysis and parallel computing of large data sets. For the implementation of mapreduce, the calculation of probabilities plays a crucial role [9]. In short terms: big data technology mainly grounds on pattern-recognition and the calculation of probabilities. An important question related to that is: What if there is such a shift away from semantics as [8] assumes? Is syntax

then becoming more meaningful, especially in big data analysis? Referring to its aspirations, big data enriches a quantity of syntax with meaning. To illustrate what big data is and what it is not, language translation is a useful example: Online translation tools (e.g., babelfish, google translate) search for patterns in large data sets about terms, phrases, and syntax. They analyze the textual input information based on its structure and syntax to calculate the probabilities of the original text in a different language. Dependent on the complexity of the text, the results mostly do not provide an exact translation but it can still give helpful hints. In several cases, this is sufficient to get a basic idea of a text, but without a solid interpretation and basic knowledge about the other language, it is often only messy, complicated information. Thus, an essential challenge of big data is the correct interpretation of the information it provides. However, coping with this challenge can be very complicated as outlined in the following sections.

### 3. Uncertainty, Predictability and Interpretation

Big data can be seen as a noble attempt to reduce uncertainty by employing complex data analysis often propagated as a means to predict future developments with predictive analytics [3]. However, the illustrative capacity of big data to point out connections between items is mostly based on probabilities and not causalities. If, literally speaking, magic is traded for fact, respectively, correlation and probability is mixed up with causation, this can lead to increasing uncertainty. While another claim is that inaccurate analysis results could be compensated by “big enough data” [1], suggesting that facts come with quantity, big data here not merely risks some inaccuracy but also taking the wrong decisions and distorting reality. Despite of the enthusiasm it creates, big data is also not an oracle and thus not capable of pre-empting the future.

In Figure 2, the Google “trends” service shows a significantly increasing interest in the search term “Heisenberg” in autumn 2013.



**Figure 2.** Google trend curve for the search term “Heisenberg”.

What does this tell us? That people are astonished by the work of the great physicist? Or were people more interested in the last episode of the TV series “Breaking Bad” (The main protagonist of this TV series uses the pseudonym “Heisenberg”) aired during this time? In 1927, the Nobel Prize winner Werner von Heisenberg discovered the uncertainty principle (in German called “Heisenberg’sche Unschärferelation”), which is a cornerstone in quantum mechanics. In short terms, the principle asserts that two complementary properties (such as position and momentum) of a particle cannot be exactly

determined at the same time and thus not known simultaneously. The more accurate the position of a particle is determined, the less accurate its momentum can be ascertained and *vice versa* [10]. In a big data context, the uncertainty principle could look as follows: The position of a data set is the information it represents at a specific moment in time. With particular usage (*i.e.*, calculation), the data set gains momentum. If the principle is also valid for data, one could argue that the more data is gathered and aggregated, the less can be known about its original contexts. In other words: there are interpretational limits of big data that complicate its verifiability, and the dimension of time plays a crucial role in this regard. An in-depth analysis of the role of the uncertainty principle for big data is of course beyond the scope of this paper. But it is still used here to point out that the effectiveness of predictive big data analytics has natural limits.

Data is always a temporal construct, as [8] reminds us. Thus, no matter how big it is, data is always a construct emerging in the course of time. Existing data in the present can only provide information about the past until the present. The fact is thus important to consider that predictive analytics of whatever kind have history as backing but also as a natural limit to their effectiveness. In other words, the future remains unpredictable with or without big data. This rather simple fact seems to be neglected in the big data discourse. Similarly, data seems to be misleadingly treated as equivalent to a valid fact. Surely, data can be valid facts, but not *per se*. The term fact is to be understood as something that is actually the case or in the words of Ludwig Wittgenstein: “The world is the totality of facts, not of things” (p. 108, [11]). Above all, data are a set of numbers, characters, signs, *etc.* Without interpretation of the data, the only valid fact about data is its existence, but the data itself does not reveal whether it is valid or true in a certain context or not.

Of course, there is no doubt that already information about the existence of data, correlations, *etc.* can be very supportive in developing knowledge. Hence, the capacity of big data does not necessarily imply a predictive look in the future. It can help to improve the understanding of what has already happened and what is going on. In this regard, big data *inter alia* bears a lot of potential in the health sector, e.g., to explore yet unknown patterns that can be very supportive for diagnosis, tailored treatment modalities, preventive medicine and so on. However, in order to tap this potential, knowledge about its limits is important. Otherwise, eventual misunderstandings of the nature of “datafied” knowledge might become reinforced. The following examples reveal some of the limits and complications of big data results. One of the seemingly “big” success stories of big data, namely Google flu trends was celebrated for its highly accurate detection of the prevalence of flu. However, as [12] pointed out, in the end, the prevalence of flu was overestimated in the 2012/2013 and 2011/2012 seasons by more than 50%. The approach in total (using search terms for the analysis) is questionable, and this example already demonstrates how misleading big data can be. The following two cases in the health sector raise much more serious concerns. The analysis of medical data surely bears a lot of promising potential to support diagnosis and treatment. However, large data sets in general and medical data specifically are a complex thing and, with an increase in automated data analysis, it can become complicated to interpret the information provided by big data. A critical example is the faulty calculation of the US big data company 23andMe that analyses genetic data of their customers to check it against risks of diseases (trivia: The founder of 23andMe is the former wife of one of the Google founders). In one case, the company informed a customer via e-mail that the analysis revealed two mutations in his DNA which are typical for a genetic disease called limb-girdle muscular dystrophy. This disease leads to paralysis and is mostly lethal. After

the first shock, the customer decided to find out whether this can actually be true. He explored his genetic data set and learned that the analysis results provided to him were wrong. The big data software made a mistake by not considering the double helix of his DNA. While it seems to be true that his DNA has two mutations, these do not occur in the same gene. The big data analysis did not recognize this. The customer confronted the company with the error and all he received was a short confirmation and apology [13]. This is not a single case. In another case, 96 customers received incorrect DNA test results [14].

The following example reveals critical issues of big data as tool to support medical diagnosis. The diagnosis found unusual lupus symptoms and an association with a certain propensity for blood clots. The doctor decided to treat the patient with anticoagulant medication (which dilutes the blood and reduces clots), and the patient did not develop a blood clot. However, this does not prove that there even was an association as the big data analysis suggested. It is simply impossible to find out whether the big data-supported diagnosis was a great success and if the medication was correct or the analysis was simply completely bogus [15].

These cases teach some important lessons. First of all: big data is very seductive in supposing it as a novel tool to tackle complexity, predict future events and learn about yet unknown facts. Surely, it uses high-performance computing to exploit large amounts of data; and, in no time, analysis can provide information to interpret the data, which would not be possible by humans in an appropriate amount of time. Furthermore, big data can (as in the second case) reveal hidden information relevant for decision-making. However, the flip side of the coin is that complexity increases with the amount of data, and the increasing complexity of big data analysis fed with increasing automation may trigger not merely uncertain but also unintended societal events. A most prominent example is provided by the NSA (National Security Agency of the US) that collects about 20 billion communications events per day, far more than a typical analyst can reasonably make use of [16]. With complexity, also the proneness to errors and a certain risk of false positives increases. The outlined examples drastically highlight that big data has enormous inherent risks of errors. The knowledge required to correctly interpret what such analyses present and find out whether it is valid information or not can often be far beyond being trivial. Besides the high mental stress caused by e.g., wrong health information of the concerned individual the question of what happens if such errors remain hidden is pressing. As outlined above, the predictive capacity of big data is naturally limited. If the results of predictive analytics are blindly trusted, then their verification or falsification can become complicated. In particular, if a predicted event is taken for granted and actions are taken to prevent this event, then the prediction can hardly be verified or falsified—for instance, can a pre-crime be prevented? Or can a merely predicted disease be effectively treated?

Hence, big data entails high risks of failure and self-fulfilling prophecies, especially if correlation is mixed up with causation as the big data discourse suggests. In combination with an increasing trend towards automated pattern-recognition and decision-making, this can even lead to situations where the social and economic costs for correcting errors becomes higher than the costs for simply taking big data results for granted or, in other words, for accepting failure. In this regard, big data, to some extent, even seduces one to hazard errors. Finally, big data challenges the role and meaning of data quality and interpretation.

#### 4. Increasing Power Asymmetries and Technology-Dependencies

Lawrence Lessig once proclaimed that “code is law” [17] meaning that software algorithms increasingly affect society. Similarly, [18] used the term algorithmic authority in this regard. The emergence and employment of big data underlines this critical assessment. As mentioned above, big data algorithms are most likely applied for probability calculating pattern-recognition techniques. These can be supportive for decision-making, not least because making a decision also implies the assessment of options, excluding unrealistic or inappropriate ones and then selecting a suitable option. Thus, there are, without any doubt, several useful applications in many domains where big data can present valuable information for decision-support. Big data can also foster autonomy, for instance by improving the early recognition of health risks that may improve patients’ quality of life, or new options for health governance, early detection of risks, *etc.* can strengthen societal autonomy. However, big data also bears risks to overestimate the significance of quantities and probabilities in a sense that only the computable counts as a valid option while others become sorted out or even remain unrecognized. These risks can become reinforced by (semi-) automated decision-making, which is among the aims of big data. In this regard, the question then - what happens with rare, unlikely events? also gains high importance. So-called black swans are exceptionally and highly improbable events, but they can have particularly high impact [19]. From a meta-perspective, big data might thus be understood as a pave maker for a new techno-determinism that is capable of re-shaping the future by transforming possibilities into probabilities.

If big data is used for (semi-) automated decision making, new forms of technology dependencies occur that deeply affect the autonomy of the individual as well as of society. In this regard, there is also a certain threat that big data leads to an increase in power asymmetries and conflicts between human autonomy and software- or algorithmic autonomy. Examples such as correlations between beer and diaper seemingly observed by the US company Wal-Mart [20] are weird but harmless. In another strange less harmless case, a girl’s pregnancy was predicted before she knew it herself by analysing her consumer behaviour [21]. These examples show how algorithms may slightly strain human autonomy by secretly analysing behavioural patterns. The power of predictability is also exploited by some business models. For instance, by a platform called the-numbers.com that intends to predict the revenues of a movie before it even appears on screen, so it calculates blockbuster probability. This prediction is based on factors such as how many famous actors are involved in a film *etc.*; a company called Marketpsych concentrates on *inter alia* exploiting information gathered from social media to learn about the impact of conflicts, wars *etc.* as well as emotional states such as fears to use it for investment and trading opportunities [22]. Recent business models are also based on algorithms that automatically generate news articles [23,24].

The mass surveillance activities revealed by Edward Snowden in 2013 alarmingly highlight the close and supportive relationship of big data with surveillance and control [16,25]. As announced last year by the former NSA director Michael Hayden, the metadata gathered by this surveillance is also used for so-called targeted killings conducted by the CIA [26]. There is also evidence for false positives in drone wars where civilians were killed. Hence, big data can literally also kill people [27].

Against this background, big data might become a new source of political, economic and military power [28]. Implications range from sharpened views on realistic options for decision-making to constrained rooms of possibilities that impact privacy and autonomy of the individual. Particularly, if the human factor becomes a subject of statistics, this raises a number of serious concerns in this regard. As (p. 54,

[29]) states, “the most interesting thing about human nature is its indeterminacy and the vast possibilities this implies: our non-essentialist essence is that we are correlatable humans before being correlated data subjects. Whatever our profile predicts about our future, a radical unpredictability remains that constitutes the core of our identity”. With growing amounts of data, the concept of informational self-determination, which is a core principle of privacy, becomes further strained. Large data sets facilitate techniques for de-anonymization and re-identification. Thus, the boundaries between personal and non-personal information increasingly blur [30]. Together with its “supportive relationship with surveillance” [25], big data can reinforce a number of related threats, such as profiling, social sorting and digital discrimination. For instance, users of privacy tools such as Tor might become classified as terrorists by the NSA surveillance software called “xkeyscore” [31]. The use of big data technology is also spreading for law enforcement and police work. Together with developments towards predictive policing, aiming at identifying “likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions” [32], big data entails a number of serious challenges than can even strain cornerstones of democracy such as the presumption of innocence or the principle of proportionality. Predictive policing is already in use e.g., by the Memphis police which employs IBM’s software blue C.R.U.S.H. (Criminal Reduction Using Statistical History) [33]. Another example is “TrapWire”, a system aiming at predicting terrorist attacks by gathering data from a large network of surveillance cameras linked to databases, *inter alia* used by several police departments in the US but also in larger cities in the UK [34]. Threat scenarios referring to the movie “Minority Report” might still be overestimated. However, automated predictive analytics might increase the pressure to act and challenge to realize the red line between appropriate intervention and excessive pre-emption. Another scenario dealing with threats of social sorting and discrimination is what can be called “Gattaca scenario” where individuals become discriminated because of their DNA. Gattaca is a movie from 1997 that draws a dystopia where people have pre-selected careers based on their genetic code [35]. People with risks for genetic diseases have lower chances for societal development and well-being. This scenario presents another red line that might be important to consider not to be crossed.

## 5. Summary and Conclusions

The de-mystification of big data reveals it as a new source of networking power in many different domains, which (as every technology) can be boost or barrier to innovation in many respects. The claims of big data to reduce uncertainty by predicting future events are widely misleading. It is of course completely unsurprising that it is definitely not a future-predicting oracle. However, a problem of big data is its seductive power to perceive it as sort of crystal ball. Instead of seducing society to believe that the future would be less unpredictable, big data applications should be presented and implemented as what they can actually be: useful tools for model-based learning by revealing yet undiscovered interrelations and correlations. Big data technologies and applications present probabilities that can provide information useful for planning. This information is not predictive, but it contributes to designing a frame that can be supportive in organizing ways to deal with uncertainty, as it suggests possibilities by presenting probabilities. These possibilities are thus not factual, they do not present causal relations *per se*, and it should be kept in mind that the frame constructed based on this information can also have an impact in itself.

The “shady side” of winning new insights for decision-making is thus the emergence of new power asymmetries where a new data pragmatism celebrating quantity and probability curtails quality and innovation. In this regard, this data pragmatism risks becoming a new techno-determinism that increases the proneness to errors and false positives and at the same time disguises failure. In a way, big data entails a threat of what could be called “normalized uncertainty”. The algorithms are highly complex and, together with trends towards automated decision-making, it becomes increasingly tricky to scrutinize the results of big data analyses. In particular, if these tools are used in domains that lack the capability and resources to critically assess their outcome. To reduce the risks of big data, it is likely reasonable to reconsider the thin line between overestimated expectations and underrepresented momentums of uncertainty that correlate with the big data discourse.

Big data can foster datafication of the individual and surveillance tendencies. Hence, there are big new challenges for effective concepts to effectively protect privacy, security, and also, increasingly, autonomy. Questions that are likely to become more pressing *inter alia* are: how to use big data for enhanced anonymization techniques, how to deal with automated decisions, and how to improve transparency and accountability. Essential challenges ahead also include enhanced requirements for data quality and interpretation. In this regard, the human factor plays a crucial role as being the one who has to interpret correctly, uncover failure and putting results in applicable contexts. This also refers to a need for new analytical skills to handle predictive analytics with care—not least because of the increasing complexity big data entails. As regards technology development, there are new challenges in the field of human computer interaction in order to design interfaces that facilitate the handling of complex data sets and their correct interpretation without reducing too much information. This is crucial to allow for scrutiny of big data which is among its major weaknesses. Thus, there is high demand for transparency, replicability and verifiability of big data analytics, particularly as regards predicting events. This demand is somewhat pressing to reduce certain risks of autonomy loss.

To make use of big data and reduce its inherent risks of automated false positives, it might be essential to set the focus, not merely on the myths and the magic of big numbers but on the exploration/information process and its components, *i.e.*, what information is used for big data, from which sources, for what purpose was it collected, what is the intention of the analysis (what should be explored), is the result plausible or not, what are the reasons for (im-) plausibility, what additional information did the analysis reveal, what are the limits of the results, and similar questions. Such a process-oriented view could alleviate getting useful information out of big data without drastically increasing complexity. Of course, this is not a panacea for all of these risks, but it can at least contribute to re-vitalize Ernst Friedrich Schumacher’s “small is beautiful” [36] to allow for a sharpened focus on the essentials of large arrays of information.

## Acknowledgments

I would like to thank all my colleagues that encouraged me to publish this work as well as the reviewers that helped to further improve the quality of this paper.

## Conflicts of Interest

The author declares no conflict of interest.

## References

1. Mayer-Schönberger, V.; Cukier, K. *Big Data: A Revolution that Will Transform how We Live, Work and Think*; Houghton Mifflin Harcourt: New York, NY, USA, 2013.
2. Bollier, D. The Promise and Peril of Big Data, 2010. Available online: <http://23.66.85.199/collateral/analyst-reports/10334-ar-promise-peril-of-big-data.pdf> (accessed on 7 December 2015).
3. Intel. Predictive Analytics 101: Next Generation Big Data Intelligence, 2003. Available online: <http://www.intel.com/content/dam/www/public/us/en/documents/best-practices/big-data-predictive-analytics-overview.pdf> (accessed on 7 December 2015).
4. Lycett, M. Datafication: Making Sense of (Big) Data in a Complex World. *Eur. J. Inf. Syst.* **2013**, *22*, 381–386.
5. Gartner Group. 3D Data Management: Controlling Data Volume, Velocity and Variety, 2001. Available online: <http://www.gartner.com/it-glossary/big-data/> (accessed on 7 December 2015).
6. Boyd, D.; Crawford, K. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Inf. Commun. Soc.* **2012**, *15*, 662–679.
7. Anderson, C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired* 2008. Available online: [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory) (accessed on 7 December 2015).
8. Boellstorff, T. Making big data, in theory. *First Monday* **2013**, *18*, doi:10.5210/fm.v18i10.4869.
9. Dyer, C.; Aaron, C.; Mont, A.; Lin, J. Fast, Easy, and Cheap: Construction of Statistical Machine Translation Models with MapReduce. In Proceedings of the Third Workshop on Statistical Machine Translation, Ohio State University, Columbus, OH, USA, 19 June 2008; pp. 199–207.
10. Hilgevoord, J.; Uffink, J. The Uncertainty Principle. In *The Stanford Encyclopedia of Philosophy*, Spring 2014 ed.; Edward, N.Z., Ed.; Available online: <http://plato.stanford.edu/archives/spr2014/entries/qt-uncertainty/> (accessed on 7 December 2015).
11. Casey, G. Wittgenstein: World, Reality and States of Affairs. *Philos. Stud.* **1992**, *33*, 107–111.
12. Lazer, D.; Kennedy, R.; King, G.; Vespignani, A. The Parable of Google Flu: Traps in Big Data Analysis. *Sci. Mag.* **2014**, *343*, 1203–1205, doi:10.1126/science.1248506.
13. Bernau, V. Falsche Genanalyse von US-Firma - Sorry, Fehldiagnose. *Süddeutsche Zeitung* 2014. Available online: <http://www.sueddeutsche.de/digital/falsche-genanalyse-von-us-firma-sorry-fehldiagnose-1.2031046> (accessed on 7 December 2015). (In German)
14. Kincaid, J. 23andMe Sends Wrong DNA Test Results To 96 Customers. *TechCrunch* 2010. Available online: <http://techcrunch.com/2010/06/07/23andme-sends-wrong-dna-test-results-to-96-customers> (accessed on 7 December 2015).
15. Bruce, P. Big Data Accelerates Medical Research? Or Not? 2014. Available online: <http://www.kdnuggets.com/2014/10/big-data-accelerates-medical-research-or-not.html> (accessed on 7 December 2015).
16. Greenwald, G. *No Place to Hide: Edward Snowden, the NSA, and the US Surveillance State*; Henry Holt: New York, NY, USA, 2014.
17. Lessig, L. *Code Version 2.0*; Basic Books: New York, NY, USA, 2006.

18. Shirky, C. A Speculative Post on the Idea of Algorithmic Authority, 2009. Available online: <http://www.shirky.com/weblog/2009/11/a-speculative-post-on-the-idea-of-algorithmic-authority/> (accessed on 7 December 2015).
19. Taleb, N.N. *The Black Swan: The Impact of the Highly Improbable*; Random House: New York, NY, USA, 2007.
20. Whitehorn, M. The parable of the beer and diapers: Never let the facts get in the way of a good story. *The Register* 2006. Available online: [http://www.theregister.co.uk/2006/08/15/beer\\_diapers/](http://www.theregister.co.uk/2006/08/15/beer_diapers/) (accessed on 7 December 2015).
21. Hill, K. How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did. *Forbes* 2012. Available online: <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/> (accessed on 7 December 2015).
22. Peterson, R. The predictability of oil prices, fear and war. *MarketPsych* 2013. Available online: <https://www.marketpsych.com/article/22/the-predictability-of-oil-prices> (accessed on 7 December 2015).
23. Rutkin, A. Rise of robot reporters: When software writes the news. *New Scientist*. **2014**, 221, doi:10.1016/S0262-4079(14)60627-8.
24. Finley, K. In the Future, Robots Will Write News That's All About You. *Wired* 2015. Available online: <http://www.wired.com/2015/03/future-news-robots-writing-audiences-one/> (accessed on 7 December 2015).
25. Lyon, D. Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data Soc.* **2014**, doi:10.1177/2053951714541861.
26. Cole, D. We Kill People Based on Metadata. *NYR Daily* 2014. Available online: <http://www.nybooks.com/daily/2014/05/10/we-kill-people-based-metadata/> (accessed on 7 December 2015).
27. Ackerman, S. CIA Drones Kill Large Groups Without Knowing Who They Are. *Wired* 2011. Available online: <http://www.wired.com/2011/11/cia-drones-marked-for-death/> (accessed on 7 December 2015).
28. Zwitter, A.J.; Hadfield, A. Governing big data. *Politics Govern.* **2014**, 2, 1–3.
29. Hildebrandt, M. Privacy and Identity. In *Privacy and the Criminal Law*; Claes, E., Duff, A., Gutwirth, S., Eds.; Antwerpen, Intersentia: Oxford, UK, 2006; pp.43–57.
30. Strauß, S.; Nentwich, M. Social network sites, privacy and the blurring boundary between public and private spaces. *Science & Public Policy* **2013**, 40, 724–732.
31. Zetter, K. The NSA Is Targeting Users of Privacy Services, Leaked Code Shows. *Wired* 2014. Available online: <http://www.wired.com/2014/07/nsa-targets-users-of-privacy-services/> (accessed on 7 December 2015).
32. Perry, W.L.; McInnis, B.; Price, C.C.; Smith, S.; Hollywood, J.S. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*; RAND Cooperation: Santa Monica, CA, USA, 2013.
33. Hickman, L. How algorithms rule the world. *Guardian* 2013. Available online: <http://www.theguardian.com/science/2013/jul/01/how-algorithms-rule-world-nsa> (accessed on 1 July 2013).
34. Shachtman, N. Trapwire: It's Not the Surveillance, It's the Sleaze. *Wired* 2012. Available online: <http://www.wired.com/2012/08/trapwire-strafor-biz/> (accessed on 7 December 2015).

35. Wikipedia. Gattaca. Available online: <https://en.wikipedia.org/wiki/Gattaca> (accessed on 7 December 2015)
36. Schumacher, E.F. *Small Is Beautiful: A Study of Economics as if People Mattered*, 25th Anniversary eds.; HarperCollins: New York, NY, USA, 2000.

© 2015 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).