

Bruno Gransche:

The Oracle of Big Data – Prophecies without Prophets

Abstract:

The need for foreknowledge intensifies and a prophetic promise of today's palm readers causes us wet palms: *letting the world speak for itself*. Big Data comes with the promise of enabling people to listen to that speaking world and of gaining accurate foreknowledge by Big Data predictions. The uncertainty of our modern, complex world overstrains our present coping capabilities, causing a feeling of slipping off a slippery slope, which in turn causes a need for increasing our own foreknowledge. Part of the Big Data promise is to grant better foreknowledge by overcoming the wrongness of scientific theory or causation assumptions. But thus, people have no other option than to *believe* in these results and perform their actions in *good faith*. This makes Big Data based outcomes *a matter of faith*. This article argues that Big Data based outcomes can be seen as today's oracle, as prophecies without prophets and reflects on the consequences of that perspective.

Agenda:

Seeking foreknowledge – The perfect conjecture	. 56
The promise of Big Data – Listen to the world itself	. 57
The problem with the promise – A matter of faith	. 58
The Oracle of Big Data – Prophecies without Prophets	. 60
Conclusion	. 61
References	. 61

Author:

Dr. Bruno Gransche:

- Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Str. 48, 76139 Karlsruhe, Germany
- 27 + 49 721 68 09 424 , 🖂 bruno.gransche@isi.fraunhofer.de, 💻 www.brunogran.de
- Relevant publications:
 - Vorausschauendes Denken: Philosophie und Zukunftsforschung jenseits von Statistik und Kalkül. Bielefeld: transcript 2015, 453 p.
 - Philosophie und Zukunftsforschung. Aktuelle Beiträge zum Begriff der Zukunft. Philosophische Rundschau, 61/3, 2014: pp. 206–227.
 - Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen: Grundsatzfragen autonomieorientierter Mensch-Technik-Verhältnisse. With Shala, E.; Hubig, Ch. et al., Stuttgart: Fraunhofer Verlag 2014, 231 p.

Seeking foreknowledge – The perfect conjecture

The future is today's hot topic. Our world is apparently always ahead in time and widely focused on future events and developments. The new is at large considered better than the old, the time to come more important than the past, innovation beats tradition, trend researchers and prediction specialists earn much more money, attention, and appreciation than historians and archaeologists. When tradition and ancient custom were generally held in high regard, they provided a liable orientation and good foreseeability for everyday decision making. Metaphorically speaking: In hardly and slowly changing cities, a ten-year-old street map did just fine – in highly dynamic and rapidly changing environments, old maps turn useless ever faster. With the same pace, our world structures are liquefying, static orientation approaches get obsolete, the need for constantly updated information, predictive efforts and anticipation rises. In that situation of a "liquid modernity"⁹⁵, the son of a blacksmith is no longer automatically becoming a blacksmith and he is no longer sure to be needed as such his entire lifetime. Today's decisions – as the choice of occupation – need a great deal of anticipation. Will welders be needed in 2050 or welding robot operators instead? Will more red or yellow shirts be sold next summer? Will male insurance clients keep on causing more severe car accidents than female ones?

The future is considered more important than ever and knowledge about the future seems to be the oil of the 21st century. The problem is: There is no such thing as 'knowledge about the future' in a complex, dynamic and non-deterministic world. Prometheus, the Greek Titan and epitome of science, alone had a perfect foreknowledge of the one single and only possible future in the deterministic Greek mythological cosmos. But – because of that – he could not do a thing to make a difference, to change this future because that would have meant to instantiate a second future different from the foreseen one, which in turn would make that perfect foreknowledge impossible. Prometheus was famous and envied for that foreknowledge and even tortured by Zeus for it.⁹⁶ This ancient myth already understood that you cannot have both: Either you can, in principle, gain knowledge of the future (being a Titan, an Oracle or a prophet) and this at the cost of not being able to perform different actions than those that inevitably lead to that one and only deterministic future, or you are free to make a difference, to influence the future, to manipulate or create different alternative futures which comes with the impossibility of foreseeing them.⁹⁷

The future is not the realm of facts but of objectives and ambitions, there is nothing true or false about the sentence 'I will buy a Richter painting.' One can believe in the so communicated plan, maybe even based on whether it is considered probable or not. But believing in a stated ambition, judging future alternatives according to alleged intuitive or scientifically calculated probabilities is far from positive knowledge about the future. The best grasp we can get about the future for epistemological reasons are more or less educated guesses, better or worse underpinned assumptions – that is: conjectures.

All we can do to approximately satisfy our rising need for foreknowledge is to further educate our guessing capabilities and develop our *art of conjecture*.⁹⁸ Our ability to perfect this kind of artistry seems to lose – despite remarkable progress – the arms race with the world's increasing complexity. This feeling of slipping off a slippery slope causes a call for new arms in increasing our own foreknowledge, which, in principle, can never

98 Jouvenel 1964.

⁹⁵ See: Bauman 2000

⁹⁶ There is one other Greek mythological figure being famous for her accurate foreknowledge of some parts of the coming: Cassandra. So why did Zeus not interrogate this mortal woman instead of meddling with a Titan – who, of course, knew he will be tortured but could not help it? The fact that knowing the future and being free to change it exclude each other holds for Cassandra as well. Here coming with the curse that no one ever believed her prophecies. So Zeus could have extorted the foreknowledge from Cassandra, but in turn he would not have believed her anyway.

⁹⁷ According to Kant, foreseeing the future of freely acting people is impossible, or as he puts it: If actions could be foreseen, there would be no freedom. "[U]nd wenn wir alle Erscheinungen seiner [des Menschen, BG] Willkür bis auf den Grund erforschen könnten, so würde es keine einzige menschliche Handlung geben, die wir nicht mit Gewißheit vorhersagen und aus ihren vorhergehenden Bedingungen als nothwendig erkennen könnten. In Ansehung dieses empirischen Charakters giebt es also keine Freiheit" (Kant 1998, 634-635 [577-578]).

exceed its conjectural nature but extols itself as predictive knowledge. One approach to ease the disturbing uncertainty of an open future was stochastics and calculating probabilities. Nonetheless, this 'statistical foreknowledge' is still conjectures expressed in numbers.⁹⁹ So the need for foreknowledge remains and intensifies. It is causing us – as we are addicted to anticipation – wet palms when we encounter the prophetic promise of today's palm readers.

The promise of Big Data – Listen to the world itself

So far, all efforts to anticipate future developments have been somehow limited by the cognitive capabilities of the anticipator. The ever limited conjecturing ability, even at the level of artistry, falls behind the open future's uncertainty. Models and theories are at the very core of the efforts to deal with uncertainty and to anticipate possible futures. Causality, for instance, insight into causal connections, is probably the dominant way of anticipating future events. Causal connections allow to predict the effect of a certain cause given similar enough circumstances. Models and theories (e.g., probability theory) are what enables the above mentioned 'statistical foreknowledge' but also all sorts of explanation of what might happen based on what happened. A theory based guess – a hypothesis – is considered improved in contrast to a mere wild guess; often its conjectural character is hidden and then called prognosis, forecast, or prediction. On the slippery slope of today's dynamic world, even the most advanced anticipation efforts, even those based on highly elaborate scientific theories, are witnessed to fail epically as seen at the financial crisis 2008. Obviously, our best anticipation capabilities are not good enough for our immense need for foreknowledge and improved theories have not brought a breakthrough so far which leads some to the suspicion that the theory foundation itself might be a shortcoming.

Big Data is now claimed to lessen the need for theories and it comes with the promise of enabling people to listen to 'the world itself'.

"The promise is that, with high levels of data generation and developments in computational analysis, the world (coded through datafication) can begin to speak for itself without its (more than) fallible human interpreter."¹⁰⁰

Or as put in the much cited article "The End of Theory": "With enough data, the numbers speak for themselves. [...] Data without a model is just noise. But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete."¹⁰¹ Even the best models are flawed and "a caricature of a more complex underlying reality."¹⁰² This leads to the promise of a Big Data enabled 'better way':

"There is now a better way. Petabytes allow us to say: 'Correlation is enough.' We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can [...] let statistical algorithms find patterns where science cannot. [...] Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all."¹⁰³

Reality in its vibrant abundance – so the luring promise – could be accessed through their authentic data, thus circumventing the anemic and essentially curtailed scientific models and theories. Understanding the datafied

⁹⁹ All calculations, even the most sophisticated, that distribute probabilities to alternative developments or events still suffer from the flaw that it is a mere guessing how many percent were to be distributed. It has become a habit to distribute 100% making three equally probable events each 33% probable. But given a fourth unknown possible event, maybe only 75%, had to be distributed on the three known events in the first place. Stochastics can provide quite sophisticated information on the known futures, but the number of known futures taken into account is restricted by one's conjecturing abilities. And are not the unknown futures much more in number, thus in likeliness, and – being unknown – causing much more uncertainty?

¹⁰⁰ Chandler 2015, 837–838.

¹⁰¹ Anderson 2008.

¹⁰² ibid.

¹⁰³ ibid.



language of our IT system pervaded world in its alleged original richness with the help of nowadays computational 'superpowers' – such as Big Data algorithms – seems to let the proverbial dream of the emperor who wanted a map of his empire being as detailed as the reality come true.¹⁰⁴ A map provides more orientation than the actual reality because it omits all unimportant details. Concerning these omitted details, the map is wrong, but that is just how it can provide orientation. "Essentially, all models are wrong, but some are useful."¹⁰⁵ – This famous aphorism is about to be outdated as expressed by Google's research director Peter Norvig: "All models are wrong, and increasingly you can succeed without them."¹⁰⁶

According to that promise of 'Correlation supersedes causation', Big Data algorithms mapping the 'data empire' could lead to such a 'perfectly accurate' map of the reality because they would overcome the constitutive difference between map and empire, between model and world. Therefore, Big Data is becoming notorious for its "unreasonable effectiveness"¹⁰⁷, the 'end of theory', and thus being responsible for the "death of the theorist"¹⁰⁸. If theory is the base of our best conjecturing abilities and if theory itself is the shortcoming of our anticipatory efforts, does the alleged death of theory then imply the death of conjecture, thus giving room to flawless since theoryless predictions? Does this scientific deicide committed by Big Data finally offer us direct foreknowledge?

Data directly derived from our very movements, actions, communications, interactions, body functions, etc. would allegedly not be distorted by any theory of causation imposed by the people trying to make sense of it. Brave new world, where Big Data systems are used to find correlations that could not have been even searched for. The wrongness of the models does no longer matter if there are no model-based hypotheses guiding the questions and defining what counts as an answer. For these algorithms, there is no such thing as unimportant details because the purpose that it has to prove useful for (such as orientation for a map) is no longer predefined. Big Data is so delightfully longed for because it is expected to give us answers we did not even know the question for, which is to bring digital serendipity to a whole new level. This is just the kind of uncertainty about our futures we are confronted with in our complex world and that stochastics failed to tackle: We need answers even if the questions were already too complex to ask, we need to approach the 'unknown unknowns', the things we do not even know that we do not know them.¹⁰⁹

The problem with the promise – A matter of faith

Just as the scientific method, the use of theory and models was not just invented as some sort of elitist brain jogging, the 'death of theory' would come with some major problems.

The first problem refers to the misunderstanding that mistakenly identifies the 'datafied world' with the 'world itself', meaning that already the promise of listening to the world itself via Big Data technologies is a modern myth. To state the obvious: Any set of data – no matter how incomprehensibly gigantic – is selective. The promise clearly disregards the fact that data is no pre-social phenomena but always already socially constructed or socially determined in its condition of formation. Data is influenced by people with certain interests and mind-sets and the data producing, collecting, storing, and processing technologies are so as well, thus selecting only data within their sensing capabilities and their scopes, that people with certain objectives and with theories about the means by which these objectives are possibly obtainable designed. The datafied world is distinguished

¹⁰⁶ Anderson 2008.

¹⁰⁴ Lyotard 1984, 55.

¹⁰⁵ Box; Draper 1987, 424.

¹⁰⁷ Halevy et al. 2009.

¹⁰⁸ Steadman 2013.

¹⁰⁹ At this point, the said aspects mainly concerning science reach into political, governance, and resilience debates. See: Chandler 2014. For a quite famous use of the concept of `unknown unknowns' see: Rumsfeld 2011.

from the world itself, at least by its inscribed traces of theory and models¹¹⁰; so claiming the complete death of theory by Big Data analytics is techno-deterministically biased and myopically dealing with the illusion of presocial objective data.

Big Data systems do not 'kill' the theoretical inheritance of data itself. They do whatsoever circumvent much theoretical wrongness in data collection and pattern recognition, what might be enough for Norvig and others to hold on to that promise. But theory comes in not only in data formation but also at the point where information meets human actors. The problem in having an answer to an unknown question is that you never know how to make sense of it. If the answer is 42, for instance, you are in trouble figuring whether kilo, percent, or years, etc. If this information should make sense and be used to motivate actions, then theory and causality have to be invested by human actors inevitably all along with the allegedly overcome wrongness again. So, at the very moment the algorithmic findings are perceived by human actors, they get subjected to some sort of causal or theoretic assumption – consciously or not, be it in a careful methodologically structured scientific or an intuitive emotional prejudicial superstitious way. For example: If Big Data systems would find a strong correlation between being depressed and being a teacher, and given a will to change that situation, people have to come up with some cause-effect assumption whether the job might depress people working in it or people predisposed for depression choose to be a teacher. In short: Do teachers get depressed or do depressed get teachers? The mere correlation cannot guide any action to solve this problem – theory can. Correlation does not supersede causation if you wish to change something and you need to know how.

In our complex world, human actors are no longer the only entities performing actions or action-like processes. Artificial agents sell stocks, filter and channel information flows, and perform all sorts of actions human actors come to deal with as mere results or as participants in all forms of human-technology interactions or coactions.¹¹¹ If assistive systems give recommendations on how to act according to found correlations (or nudges or forces people in a certain direction by modifying interfaces, contents, systems behavior, etc.), it is crucial to be able to deduce the system's behavior and its underlying processes in order to understand and evaluate the recommendations. If this theory-based validation by people is still possible, then the whole human-technology interaction is still as 'defective' as the theories are. In order to unleash its promised potential to deal with a data quantity and heterogeneity being impossible for humans to grasp even with much time and effort – which is one definiens of Big Data. Delivering insights – or predictions based on them – that people without algorithmic help could never have found is the alleged potential of Big Data systems and it is at the same time the exclusion of scientific validation because accountability, verification (for the time being), and falsification are essential for science.

When people get confronted with information, processes, part-actions and actions, or results based on Big Data algorithms they have no chance to retrace how these outcomes were generated, what they were based on, and if they are 'true, right, or correct' (if any of these concept applies at all). Thus, in a datafied world widely pervaded with Big Data technology and artificial agents acting on this basis, people have no other option than to *believe* in these results and perform their actions in *good faith*. Within these systems that are claimed to render scientific method obsolete, there is no space for scientific falsification. This makes Big Data based outcomes *a matter of faith*. It is information (or hybrid actions based on this information) coming from a source that is principally obscure to human actors. And at the very moment it enters the human sphere, it becomes an orienting force, guiding people's and agents' actions no matter of their original correctness. As for the claim of 'death of theory', this is where its potential validity ends: Algorithms, systems, artificial agents may be able to perform beyond theory only on the ground of abundant data¹¹², but human beings are not. When

¹¹⁰ GPS data, for instance, with which movement profiles can be created inherit assumptions of both the special and the general theory of relativity and, thus, of course, their theoretical correctness and wrongness.

¹¹¹ Gransche et al. 2014

¹¹² NB: This 'beyond theory' refers only to their performance. As well as data, IT systems and artificial agents are no pre-social phenomena but underlie a theory-compromised formation process.

encountering human actors, the (if so ever at all) flawless since theoryless information is 'corrupted' by more or less theory-based interpretations and validations of people before being transformed into actions.

Beyond scientific validation possibilities, Big Data findings are indistinguishable from Big Data creations or data noise artifacts. Given a large enough search room, there are always correlations.¹¹³ Big Data findings and creations have the same potential impact on human behavior: How could people distinguish them in the first place? Those who do not know their difference are forced to believe in both or none equally. If only enough people believe in this guiding character of Big Data based outcomes – and the current hype is strongly suggesting that this is absolutely the case –, then these outcomes develop a self-fulfilling and self-defeating power as known from the respective prophecies.¹¹⁴ So, does the end of theory correlate with a renaissance of prophecy?

The Oracle of Big Data – Prophecies without Prophets

Wherever people are ignorant, there will be prophets.¹¹⁵ Scientific prognoses – in terms of probabilities including their range of uncertainty, their limitations of validity, and their condition of formation like transparency about the set of information and hypotheses they are based on - can be used to inform decisions; they lessen people's ignorance if not mistaken in its conjectural nature. Post-theory Big Data predictions, on the other hand, lack this self-referring information. People, nevertheless, using them to base their actions on are ignorant about their range of uncertainty and validity, their formation circumstances, etc. Big Data based outcomes, being a matter of faith, can be seen as today's prophecies. As they are not claimed by deficient mortal beings but by some sort of pseudo omniscient algorithmic deity, they are the paradox of prophecies without prophets. Thus, Big Data becomes some sort of today's oracle, a voice revealing insights and predictions from an abundant yet obscure source that is claimed to be the world itself - or at least as close to it as we can hope to get. And just like the ancient oracles, its power does not derive from any correctness of the content of any single prophecy but from the people believing in it. In contrast to scientific prognosis, which is a matter of doubt, those Big Data prophecies being a matter of faith are immune to critique or falsification. Both the oracles of ancient times and those of Big Data have this immunity in common; the former because they were seen as an authentic direct message from the Gods in a deterministic cosmos, the latter because it is broadly believed to be the world speaking for itself. The actual events either prove the correctness of their prediction or the wrongness of the fallible interpreter.

Prognosis and prophecy are two ways of dealing with future unknowns.¹¹⁶ The modern approach of prognosis accepts the existence of indispensable unknowns along with the notion of an open future. The ancient prophecies placed all the uncertainty in the impartial human knowledge and misunderstandings of a principally knowable future. If Big Data prophecies take the fallible interpreter out of the equation providing prophecies without prophets, this would not only mean that positive foreknowledge would after all be possible but also even directly accessible. Prognosis and prophecies have similar power as socially effective speech-acts. Prophecies, in addition, have two advantages as powerful speech-acts of which the first is the said bonus of infallibility. The second one is a strong awareness of its circularity, which primarily holds for ancient prophecies. While prognoses inherit the scientific tendency to see themselves as uninvolved observers, as mere describing objective entities, prophecies always included their effect in the prophesied future (Oedipus for instance). That is why we know self-fulfilling and suicidal or self-defeating prophecies but not such prognoses. This valuable

¹¹³ On a global scale, there is a good chance that, every time I breathe in and out, one human being dies and another one is born at the same time what obviously does not make my breath lethal or life giving; nonetheless this correlation could get 'recognised' by Big Data pattern recognition. It is human causal common sense that instantly classifies this correlation as absurd. Who knows how many artificial agents already sold, filtered, channelled masses of stocks, information, services, and wares on that kind of correlation? All we might see is a changed price in the end with no chance to check which correlations lead to it.

¹¹⁴ Merton 1948.

¹¹⁵ "Partout où les hommes seront ignorants, il y aura des prophètes", d'Holbach, Paul Henri Thiry, 123.

¹¹⁶ Esposito 2013.



awareness of circularity is one lesson to learn from the ancient prophecies and should be transferred to scientific prognoses¹¹⁷ and to today's Big Data predictions.

Conclusion

Big Data systems do not bring the end of theory, but – apart from the theory inheritance within data itself – they postpone theoretical interpretation within the information-action chain to a point where it might cause less wrongness on the one hand but also less possibility to evaluate and correct previous parts in that chain on the other hand. This might lead to problems concerning accountability of co-actions to which a hybrid variety of human an artificial actors contribute. Model wrongness is not overcome but relocated and in disguise, thus withdrawn from scientific critique and improvement processes. Predictions – shifting from prognoses to modern prophecies – change their nature from being a matter of doubt to a matter of faith. As decreased fallibility of prophetic foreknowledge comes with decreased freedom of action (Prometheus) and as the appearance of prophets is connected with increased ignorance (d'Holbach), the renaissance of prophecies should alert a progressive democratic society but yet not lead to defensive overreaction as there are insights to be learned from prophetic future anticipation such as a strong awareness of prediction circularity.

Big Data services are indubitably playing an increasing role not only in science but also in politics and economy as well and, therefore, many questions are to be dealt with. How should a society reintroducing the concepts of oracles and prophecies (even if not under these names) at the expense of scientific methods deal with that kind of strategy shift in approaching complex and open futures? What do powerful oracles and prophecies mean in terms of responsibility, accountability, democracy, resilience, governmental influence, and (self-)governance capabilities? Who and where are the new prophets staging themselves as 'out of the equation' and staging the objectivity of 'the world speaking for itself' while strategically acting from behind the curtain? What do they win with this disguise? Are Google's and other Big Data Titans' imperatives actually a surprisingly honest totalitarian rule – "So, follow the data." – and are they a revealing witness of their potentially hazardous approach on (not) shaping the future – "Now go out and gather some data, and see what it can do."¹¹⁸?

References

- Anderson, Chris: The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, 2008. Available from <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory>. Accessed 27 September 2015.
- Bauman, Zygmunt: Liquid modernity. Cambridge, UK, Malden, MA: Polity Press; Blackwell, 2000.
- *Box, George E. P.; Draper, Norman Richard: Empirical model-building and response surfaces. New York: Wiley 1987.*
- Chandler, David: Beyond neoliberalism: resilience, the new art of governing complexity. Resilience 2 (1), 2014, pp. 47–63.
- Chandler, David: A World without Causation: Big Data and the Coming of Age of Posthumanism. Millennium -Journal of International Studies 43 (3), 2015, pp. 833–851.
- d'Holbach, Paul Henri Thiry: Le Christianisme dévoilé: In Premières Oeuvres, pp. 94–137. Paris: 1797.

¹¹⁷ This is widely the case in stock market prognosis, bets, and futures but still rare in scientific prognosis.

¹¹⁸ Halevy et al. 2009, 12.

- *Esposito, Elena: Formen der Zirkularität in der Konstruktion der Zukunft. In Prophetie und Prognostik: Verfügungen über Zukunft in Wissenschaften, Religionen und Künsten, edited by Daniel Weidner and Stefan Willer, München: Wilhelm Fink 2013, pp. 325–340.*
- Gransche, Bruno; Shala, Erduana; Hubig, Christoph et al.: Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen: Grundsatzfragen autonomieorientierter Mensch-Technik-Verhältnisse. Stuttgart: Fraunhofer Verlag 2014.
- Halevy, Alon; Norvig, Peter; Pereira, Fernando: The Unreasonable Effectiveness of Data. IEEE Computer Society, 2009, pp. 8–12.
- Jouvenel, Bertrand de: L'Art de la conjecture. Monaco: Éditions du Rocher 1964.
- Kant, Immanuel: Kritik der reinen Vernunft. Hamburg: Felix Meiner 1998.
- Lyotard, Jean-François: The postmodern condition: A report on knowledge. Manchester: Manchester University Press 1984.
- Merton, Robert K: The Self-Fulfilling Prophecy. The Antioch Review 8 (2), 1948, pp. 193–210.
- Rumsfeld, Donald: Known and unknown: A memoir. New York: Sentinel 2011.
- Steadman, Ian: Big data and the death of the theorist, 2013. Available from http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory. Accessed 27. Sep. 2015.