Annals of Tourism Research 83 (2020) 102931

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

Research article

Hidden theorizing in big data analytics: With a reference to tourism design research

Josef A. Mazanec

Modul University, Vienna

ARTICLE INFO

Associate editor: Daniel Fesenmaier

Keywords:
Tourism design
Big data analytics
Inductive reasoning
Theory building

ABSTRACT

As demonstrated in Xiang and Fesenmaier's (2017) collection of articles Tourism Design involves the application of Big Data Analytics. This raises the question whether the tourism researchers' struggle for theory building may (have) come to an end since Big Data is doing it for them. Addressing several misunderstandings, the author discusses most recent studies, mainly from the field of Tourism Design, and identifies the masked elements of theory hidden in heavily data-driven big data analytical approaches.

Introduction

The case for inductive reasoning in the social sciences has been advanced whenever researchers struggled to escape the Procrustes scheme of the hypothetico-deductive method. A prototypical example is Edwin Locke's plead for inductive theory building. This author claims that '… science has not, and could not have, progressed by the process of falsification' but 'progressed only by the process of making positive discoveries' (Locke, 2007, p. 869). Besides being deceived by popular misinterpretations of critical rationalism he ignores the commonplace experience that one learns by trial and error. He also misses the old and highly instructive distinction between the 'context of discovery' and the 'context of justification' (Reichenbach, 1938, p. 6f.). Indeed, inductive reasoning has always been an accepted part of the process of discovery; however, despite enormous efforts in inductive logic (Carnap & Stegmüller, 1958) and progress in Bayesian approaches (Godfrey-Smith, 2003, Ch. 14), it could not properly solve the confirmation problem. Nevertheless, inductive reasoning is an essential element of science, but does it operate unbiased by theory?

Locke (2007) puts forward three examples of 'successful inductive theory building': Beck's cognitive theory of depression, Bandura's social-cognitive theory, and his own Locke and Latham goal setting theory. The latter yields puzzling findings like '…people with easy goals are easier to satisfy; those with hard goals are harder to satisfy' (p. 879). But what is really important to diagnose by critically studying the three examples is the fact that none of the underlying discoveries happened via observation free of preconceived opinions. Beck made his discovery while practicing psychoanalysis, perhaps not an undisputed theory, but certainly a source of guidance for collecting data. Bandura conducted experiments and, unless one flips a coin, the various decisions connected with experimental design are based on hypotheses. Reading Locke's description of how they developed the Locke and Latham management theory strengthens the impression that the authors may have abstained from formal hypothesizing, but the whole process is full of unspoken expectations.

So, what do we encounter in Big Data research? As Big Data missionaries announce the end of theory (Anderson, 2008) the old debate resurfaces and raises the question whether researchers in the empirical sciences have to cope with unprecedented phenomena. Is Big Data research indicative of scientific revolution, '… a transformation of the world within which scientific work was done' (Kuhn, 1962, 1970, p. 6) or extrapolated normal science?

E-mail addresses: josef@mazanec.com, josef.mazanec@modul.ac.at.

https://doi.org/10.1016/j.annals.2020.102931
Received 21 November 2019; Received in revised form 18 March 2020; Accepted 24 March 2020
0160-7383/

Kitchin (2014) describes the Big Data 'challenge' as: 'The challenge of analysing Big Data is coping with abundance, exhaustivity and variety, timeliness and dynamism, messiness and uncertainty, high relationality, and the fact that much of what is generated has no specific question in mind or is a by-product of another activity' (p. 2). The verdict of 'no specific question in mind' is the characteristic that clashes with scientific principles, unless purposeless activity counts as scientific research.

Typically, Big Data applications, and especially within tourism design related research, arrive as huge case studies. Think, e.g., of analyzing 17,066 geotagged photo records for detecting associative point-of-interest patterns (Lee, Cai, & Lee, 2014) which are time and place specific. The case-study character of Big Data Analytics becomes apparent in the critical literature survey of Günther, Rezazade, Huysman, and Feldberg (2017) where the authors focus on 'how organizations realize social and economic value from big data' (p. 205). Although they mention the deductive and hypothesis-driven approach to Big Data Analytics – the only place where the word 'theory' appears once – their concerns are predominantly practice-oriented and business-driven. Organizations keen on cumulative knowledge building may have to fulfill slightly different requirements than those trying to leverage Big Data insights case by case. Does this actually indicate absence of theory?

'Let the data speak for themselves'

Several authors favor correlation and prediction over causal explanation. 'Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all' (Anderson, 2008). When Gandomi and Haider (2015) elaborate 'distinctive features inherent in big data' they only think about the consequences for 'predictive analytics'; the adjective 'explanatory' appears in sentences such as 'capture the interdependencies between outcome variable (s) and explanatory variables, and exploit them to make predictions' (p. 143). There is no mentioning of explanation as an objective or of terms like 'causal' or 'causality'.

In their systematic literature survey 1996–2015 Sivarajah, Kamal, Irani, and Weerakkody (2017) describe data, process, and management challenges of Big Data Analytics and propose methods for mastering them. They discuss descriptive, predictive, and prescriptive analytics but do not mention explanatory purposes. In other words, 'predictive' analytics comprise pattern recognition and relationship detection without claiming explanation. How does 'prescriptive' analytics then 'determine the cause-effect relationship among analytic results and business process optimization policies'? It is no surprise that 'There are very limited examples of good prescriptive analytics in the real world' (p. 276).

Big Data Analytics have been discovered by researchers in most of the empirical sciences and certainly the disciplinary background moderates the scientists' attitude toward data-driven research. Coveney, Dougherty, and Highfield (2016) argue from the biology and medicine point of view and violently ask for 'big theory' in Big Data research. They hold that '…we need models and theoretical insights to help guide the collection, curation and interpretation of data.' But there is little doubt that the 'complex stochastic systems' encountered in the life sciences are no less complex in the social sciences. Perhaps it need not be 'big' theory. Elragal and Klischewski (2017) outline the epistemological pitfalls in all stages of Big Data Analytics. They propose 'a theory-driven guidance for the BDA process including acquisition, pre-processing, analytics and interpretation' and recommend what they call a 'lightweight theory-driven approach' (p. 5). Theory may enter through the backdoor if Big Data Analytics are combined with accompanying design principles. Miah, Vub, Gammack, and McGrath (2017) apply design science principles for developing a procedure for analyzing social media Big Data. The design guidelines are made explicit (in Table 2, p. 775) and they would be futile unless resting on hypotheses relating obedience to these guidelines to the quality of the design outcome.

Sometimes, the advent of Big Data Analytics is considered a transition into a new paradigm in the Kuhnian sense. As Kitchin (2014) puts it: "rather than testing a theory by analyzing relevant data, new data analytics seek to gain insights 'born from the data' " (p. 2). However, after enthusiastically welcoming Big Data Analytics the same author later concludes that 'just as data are not generated free from theory, neither can they simply speak for themselves free of human bias or framing' (p. 5). It is typical for this and similar discussions pondering the pros and cons that finally Big Data arrive where all inductive theory-building ends up, in the context of discovery. Accordingly, the Big Data movement does not fundamentally alter *the* scientific method. But it lets even the most unimaginative empirical researcher become an empowered and resourceful generator of bold (systems of) hypotheses.

Leo Breiman (2001) explicates the difference between the two 'cultures' of 'data modeling' and 'algorithmic modeling' in a convincing manner. By data models he means assumptions about the data-generating process while algorithmic models are black-box models. He demonstrates how algorithmic methods like random forests, support vector machines or neural nets – themselves representing black boxes 'slightly less inscrutable than nature's' – excel in problem-solving by sacrificing 'interpretability' for 'information accuracy' (p. 209f.). One must not overlook, however, that Breiman's critique addresses the overwhelming interest of statisticians in data modeling, a reproach that does not hit social science research with equal force. For example, statisticians are still sceptic of Partial Least Squares and PLS Path Modeling while social scientists have embraced this method which its inventor recommended for 'data-rich and theory-primitive' problem settings (Wold, 1982, p. 4f). Wedel and Kannan's (2016) historical account of data and analysis in marketing science proves marketing researchers' attention for unstructured data. Nevertheless, understanding marketing intelligence as a decision-making aid, the authors speak of Big Data Analytics serving 'estimation of causal effects' (p. 114). Thus, if social scientists claim to provide decision support for policy-makers they must acknowledge that decision-making means intervention reliant on causal relationships. '…the explanatory account of causation is merely a variant of the manipulative account, albeit one where interventions are dormant' (Pearl, 2001, p. 26). At least for marketing science Big Data and marketing theory are not antagonists as Lukosius and Hyman (2019) provide nine examples of academic papers testing marketing theories with Big Data. And in information systems research Agarwal and Dhar (2014) propagate a balanced view. The authors state that the '… tension between correlational analysis and causal testing of hypotheses represents a fundamental dilemma in the use of big data for explanation versus prediction' (p. 446). However, they insist on Big Data being useful for hypothesis generation as well as testing. Predictive success may occur without explicit hypothesizing, but certainly opens a path to subsequent developing of explanatory models.

'Let the data speak for themselves' is an attractive slogan and a tempting strategy. **But we must give them a language first.** Whether we choose this language after thorough reasoning or casually or by tacit agreement does not matter. It always implies theory. Semantics provides the terms and concepts to signify the observations. Syntactic rules govern the admissible data manipulations. Even the simplest statement like 'At this moment a book is displayed on my desk' requires a theory of time, space, and desk. Actually, a 'linguistic framework' that exceeds the 'world of things' (Carnap, 1950) is needed as some of the observables in Big Data collection most likely serve as indicators for abstract entities. For instance, a star rating in a hotel review does not merely represent a tick mark in an online questionnaire, but quantifies the strength of an evaluative belief. This means a correspondence relation to a latent variable such as attitude. Thus, a number of background hypotheses get involved including assumptions regarding measurement level (ordinal, interval) and attitude dimension (cognitive, affective).

Lessons from big data analytics in tourism design

How may we characterize the situation in tourism research? The bulk of tourism Big Data consists of the electronic vertices of the tourists' travel planning and *en route* behavior where the trend favors use of multiple channels (Xiang, Magnini, & Fesenmaier, 2015). In their comprehensive literature survey covering 165 publications, Li, Xu, Tang, Wang, and Li (2018) distinguish user-generated, device, and transaction data. For each category the authors examine research focus, data characteristics and analytic techniques in great detail. Given the overwhelming wealth of data and the meticulous description of how they were collected, purified and analyzed in all these empirical studies one baffling observation remains: The expression 'valuable information' pops up five times, but there is no mentioning of generalizations or cause-effect findings that have been derived from the massive input of data and analytics. Further, it seems that this study misses the point: Where did Big Data study actually lead to an inductively generated theory? No theory in, no theory out … is it as simple as this?

Consider a typical Big Data tourism study by Chareyron, Da-Rugna, and Raimbault (2014) who analyze traveler route maps constructed from geotagged and chronologically ordered photos and suggest relating the itineraries and activities to social networks of community websites. The research question '…is there a correlation between user information (age, sex, origin, camera model, etc.) and reviews or paths?' is equivalent to hypothesizing such relationships. However, it would gain much in theoretical relevance, if it were backed up by pondering potential reasons of such correlations. Further highlighting of the situation in tourism studies focuses on the particular research avenue of Tourism Design. The discussion proceeds with the publication by Xiang and Fesenmaier (2017) who collected 15 up-to-date contributions on Big Data Analytics in Smart Tourism Design. These articles represent an instructive overview of the tourism research issues currently investigated with Big Data methodology. What kind of position do the authors of the contributed articles take along the continuum between data-driven and theory-guided work?

When discussing the characteristics of tourism Big Data Song and Liu (2017) regard veracity as the biggest challenge. They raise the question 'Is the data being stored and mined meaningful to the problem being analyzed?' (p. 18). What they do not add is that answering this question necessitates prior hypothesizing, at least implicitly. The empiricists' criterion for meaningfulness has always been testability. And one cannot empirically test what is not phrased as a reasoned expectation of some observations.

Sensory data is becoming an important resource for Big Data Analytics (Kim & Fesenmaier, 2017). In principle, the use of psychophysical measures dates back to the 1960ies when advertising research applied the polygraph for inferring psychic arousal levels from galvanic skin response and accompanying physiological indicators. But despite the technological development into telemetric and brain wave measures one aspect has not changed since the early days. Interpretation of these data and the very idea of relying on physiological indicators heavily depend on theory (Nelson, 2016). By the same token, the method of sentiment analysis lacks orientation without taking recourse to some theory of emotions. E.g., Scharl, Lalicic, and Önder (2017) choose the five affective dimensions propagated in marketing research by Aaker (1997). In the future, more elaborate theoretical systems like Plutchik's Wheel of Emotions (Plutchik, 1980) may attract the Big Data research community.

In analyzing online travel reviews, Marine-Roig (2017) explicitly speaks of a 'theoretical framework' when emphasizing the importance of considering the paratextual elements of postings. There may not (yet) exist a full-fledged theory of the interrelationships and effects of the various types of textual elements in online reviews. However, as the exposition by Marine-Roig (2017) nicely demonstrates, the discussion is full of implicit assumptions about the review writers' and readers' perceptions and behaviors. When Wan and Law (2017) evaluate the role of online reviews in hospitality management it becomes apparent that the leveraging of this resource would not work out without being rooted in explanatory models of service quality, brand attitude, satisfaction, word-of-mouth, or consumer complaining behavior.

Other Big Data Analytics articles explicitly pursue a model-based approach. Marchiori and Cantoni (2017) propagate 'user experience risk assessment' and a 'destination online reputation' model. Xiang, Schwartz, and Uysal (2017) place online reviews into a five-step sequence leading to competitive strategy formulation. Li and Yang (2017) analyze traveler flows with a negative binomial model, as does Park (2017) in his analysis of star ratings in online reviews. What goes by the name of 'model' in these articles does not always imply clearly stated if-then or cause-effect relationships. However, an explanatory purpose always lurks behind the various flow charts and pictorial representations. Proposed metrics for assessing the economic value of utilizing social media networks (Pan & You, 2017) clearly assume causal relationships between these measures and business performance.

Quite frequently it turns out that 'research questions' are nothing other than hypotheses in disguise. By example, Kirilenko and Stepchenkova (2017), in their Sochi Olympics case study, ask about the geographic location of Twitter messages, their temporal distribution and contents. But why analyze millions of tweets? There must be a reason which readers learn in the concluding lines of the article: '… investigation of differences between the countries and regions in relation to a specific sporting event' and '… mapping the connections inside the collected data has the potential to provide insights into information flows between people and groups discussing events' (p. 234). Apparently, the analytical exercise is based on the premise that it captures explanatory variables of users' information behavior.

One may argue that Big Data Analytics in the Xiang and Fesenmaier (2017) book on tourism design does not fully reflect the situation of data-driven tourism research in general. Considering more examples from the typical areas of application, viz. forecasting, text mining, and exploiting geo-tagged photo collections, broadens the perspective. For example, Gunter and Önder (2016) use Google Analytics indicators for predicting visitor arrivals at the city of Vienna. Already in their introduction the authors state that the '…theory explaining the search behavior of the visitors to the DMOs' websites is called information foraging theory, which is derived from behavioral ecology and which is similar to food foraging theories in anthropology (p. 199). This uses Big Data, but there is not blind faith in data-drivenness. A little later the authors write: 'There exist more than 20 Google Analytics indicators, 10 of which were used in the present study based on data availability and potential predictive power' (p. 200). 'Potential predictive power' depends on expectations aka hypotheses. In another study, Pan and Yang (2016) develop a forecasting hotel occupancy model using multiple Big Data sources such as local tourism bureau's website traffic and detailed weather information. Their statement 'A consumer will mostly likely visit a business's website prior to making a purchase' (p. 959) today appears like a trivial background hypothesis. Certainly, it was not when the Web and eCommerce were in their infancy. Further, the sentiment analysis published by Scharl et al. (2017) was mentioned above and exploited results from consumer behavior research. Finally, the review article by Alaei, Becken, and Stantic (2019) sets out to 'examine the state-of-the-art sentiment analysis methods' (p. 176). In their examination it becomes obvious that the necessary polarity classification implies a theory of meaning and language structure irrespective of whether machine-learning classifiers are trained with examples or the assignment of terms follows a dictionary. It would be easy to demonstrate that any text-mining approach based on analyzing term-document matrices rests on numerous background assumptions from the selection of text corpora to text preprocessing, classification, or latent topic extraction.

What about analyzing the user-generated content of photo collection platforms? These initiatives may appear as candidates for purely data-driven work - inductive behavior at its finest. Kim, Kim, Lee, Lee, and Andrada (2019) raise the research questions 'Which locations do people prefer?', and 'What are the characteristics of visitation patterns regarding features inside tourist destinations?'; the authors seek answers by analyzing 'spatial patterns of visitation using 10 years of Flickr geo-tagged photographs' (p. 249). They implicitly recognize the hypothetical character of the detected visitation patterns as they report on a 'validation' step involving focus group interviews with regional experts and management officers. The whole endeavor clearly rests on the a-priori assumptions that geo-tagged photo sequences reflect meaningful visitation patterns, and that they are influenced by location attributes and capable of predicting visitation revenue. The hypothesizing is not just correlational but directional as required by the chosen method of geographically weighted regression. In their study on special interest tourism (SIT) Ma, Kirilenko, and Stepchenkova (2020) deal with social media big data (41,747 geo-tagged Instagram photos taken by 37,652 users). They speak about testing: 'Thus, the main purpose of this study was to put tourism typologies associated with SIT to the empirical test in order to get further insight into the debated relationship between SIT and mass tourism' (p. 2). And accordingly, they propose three explicit hypotheses stating that the dependence of visitation rate on travel distance, education, and income differs between two tourist types. The study setup resembles the marketing strategy originally named 'a posteriori segmentation' (Mazanec, 1992, p. 40). It is regarded to be data-driven, but always demands the fundamental hypothesis that the consumer or tourist population can be partitioned into subgroups that are homogeneous in terms of psychographic or behavioral characteristics.

The final study in our chasing of hypotheses in Big Data Analytics operates on meta-level. Mandal (2018) does not apply BDA itself, but investigates how tourism businesses cope with this challenge; as 'tourism enterprises operate in a complex and competitive environment, they need to rely on cross-domain analytics to analyse the trends in environmental, economic, and social sustainability' (p. 1107). With four explicit hypotheses the author develops an explanatory model of how sustainable tourism supply chain performance is determined by the organization's BDA planning, investment, coordination, and control capabilities. In his review of Marquez and Lev's (2017) book on 'Big Data Management' Cheng (2017) finds that 'big data has started to shift us to work backward with the start of the data collection followed by analysis and finally insights' (p. 455). This is a common picture: In the beginning there was the data. However, this is no novelty as so-called secondary data have been material for analysis in the empirical sciences ever since. The examples discussed so far demonstrate that Big Data researchers choose their material as purposeful as their small data colleagues. The difference is that they may feel exempt from sharing their study-guiding expectations explicitly.

Reconciling big data and theory

Information systems research is certainly a forerunner in Big Data Analytics. Hence, it is worthwhile asking how IS researchers are coping with the data-and/or-theory problem. Berente, Seidel, and Safadic (2018) describe two routes to theory development by processing trace data via manual (grounded theory) and/or automated (computational theory) identification of concepts and associations. The key element is the 'lexicon' as 'the pre-existing set of concepts, constructs, and their implied relationships' (p. 3). In traditional philosophy of science parlance, it represents nothing else than the well-known tenet that observation and the ensuing scientific processes are theory-laden (Brewer & Lambert, 2001; Kordig, 1971). Even the most enthusiastic adept of inductive theory building chooses a language 'as the pre-theoretical lexicons in the process of novel theorizing' (Berente et al., 2018, p. 12).

Empirical data, big or small, populate the world of experience. In his famous paper on 'Two Dogmas of Empiricism' (1953) Willard Quine (1951) depicts its relation to knowledge in a metaphor: 'The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience' (p. 35). It seems that analysts are doomed to haul along the invisible heritage of a body of knowledge. This bag is never empty. It contains all the unvoiced beliefs, expectations, tacit assumptions and background hypotheses, and 'auxiliary statements' (Putnam, 1991, Chapter 6) needed for purposive data gathering and developing plus testing theories as well.

'A science needs points of view, and theoretical problems' (Popper, 1959, p. 88; *'Wir brauchen Gesichtspunkte, theoretische Fragestellungen'*, Popper, 1934, p. 71). Big Data Analytics transcend naïve data collection insofar as one or more methods of data manipulation and processing get always involved. Each of these methods carries its own backpack of required data properties and application assumptions. But of even more importance, the analysts, unless acting randomly, are posing a question (*Fragestellung*), perhaps not yet regarded as a problem, rather out of curiosity. Even this inquisitiveness is based on rudimentary hypothesizing. Choosing some Internet platforms, repositories of unstructured data, text corpora, remote sensing data, a combination of several sources … rests on the expectation of detecting associations and recognizing patterns. Inductive theory builders have not changed strategy; they only have discovered a new toolbox way more powerful than old craft accessories. Theory still means preserved experiential knowledge and its noble function is preventing scientists from reinventing the wheel over and over again.

The examples from tourism Big Data Analytics exhibit theoretical building blocks to a varying degree. However, none of them lacks theory totally. Stripped of all Big Data glamour they still harbor theoretical remnants. As a consequence, they pave the way for entering the context of justification and of rigorous empirical testing with a chance of improving in generalizability. Theory-guided research work is not necessarily linked to the customary way of null hypothesis significance (NHST) and variable directional relationship (VDR) testing which has to face increasing criticism. Hubbard (2004) denounced the indiscriminate blending of the Fisherian and Neyman-Pearson approaches to statistical testing. Ziliak and McCloskey (2008) crusaded against 'the cult of statistical significance' though their arguments did not remain unwithspoken (Spanos, 2008). In business research the criticism has led to fervent pleading for 'Good Science' that dismisses NHST and turns to asymmetric modeling (Woodside, 2016). Qualitative Comparison Analysis (QCA) in its crisp or fuzzy form has been the workhorse of asymmetric configural modeling. In a tourism-related study Ferguson, Megehee, and Woodside (2017) provide a best-practice example of how to apply QCA setting out with carefully justified hypotheses rather than data-driven exploration. The authors demonstrate how fuzzy QCA assists in examining determinants of nation-specific tipping behavior. Hofstede's cultural values, religiosity, income inequality, and purchasing power parity serve as determinants combined in various configurations whose composition the authors meticulously substantiate and justify with previous literature findings. However, it remains an open issue whether and how QCA will be able to cope with the enormous sample sizes in Big Data Analytics. The founder of QCA, Charles Ragin, had 5–50 case examples in mind (Ragin, n.d.) and later extensions speak of moderate sample sizes of 50+ (Greckhamer, Misanyi, & Fiss, 2013).

Conclusion

Application of Big Data Analytics was the starting point for surveying Tourism Design studies. There may be divergent opinions as to what extent the examples from tourism research actually implement Design Science principles. However, they share one commonality. The 'artifacts' under study are various sorts of social systems defined by behaviors of tourism stakeholders (mostly tourists or their electronic traces). 'IT & Tourism' has become a flourishing field of research. The discussion about Design Science in Information Systems Research seems to shift from IT artifacts to social systems, a development where the 'input for the scientific design process is provided by theories' (Drechsler, 2012, p. 192). Tourism is no different. Both, studying tourism artifacts and constructing design rules rest on behavioral assumptions.

The Big Data *tour d'horizon* shows compelling evidence that '…entering the empirical domain without being prejudiced by prior hypotheses' once unmasked as 'Myth 1' in tourism research (Mazanec, 2009, p. 319) rightly deserves this label. There is a striking analogy to Paul Watzlawick's first axiom of communication which states the 'impossibility of not communicating' (Watzlawick, Beavin, & Jackson, 1967, p. 275). In the same vein one may say that humans cannot *not* hypothesize. This applies to everyday behavior and the more to scientific activity. John Little's concept of the 'intuitive model' in managerial problem-solving comes to mind (Little, 1970). Implicit assumptions and expectations always guide our behavior. We are bound to hypothesize if we like it or not. If making research processes vulnerable to criticism is taken seriously there mustn't remain any hidden background hypotheses. Explicitness renders them scientific.

References

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research, 34*, 347–356.
Agarwal, R., & Dhar, V. (2014). Editorial-big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research, 25*(3), 443–448.
Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research, 58*(2), 175–191.
Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. http://www.wired.com/2008/06/pb-theory/, Accessed date: 29 February 2020.
Berente, N., Seidel, S., & Safadic, H. (2018). Data-driven computationally intensive theory development. *Information Systems Research,* 1–15. https://doi.org/10.1287/isre.2018.0774 Articles in Advance. (last accessed 02-29-2020).
Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science, 16*(3), 199–231.
Brewer, W. F., & Lambert, B. L. (2001). The theory-ladenness of observation and the theory-ladenness of the rest of the scientific process. *Philosophy of Science, 3*(68), S176–S186 Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part I: Contributed Papers.
Carnap, R. (1950). Empiricsm, semantics, and ontology. *Revue Internationale de Philosophie, 4*, 20–40.

Carnap, R., & Stegmüller, W. (1958). *Induktive Logik und Wahrscheinlichkeit.* Vienna: Springer.

Chareyron, G., Da-Rugna, J., & Raimbault, T. (2014). Big data: A new challenge for tourism. *IEEE international conference on big data* (pp. 5–7). . https://ieeexplore.ieee.org/document/7004475, Accessed date: 29 February 2020.

Cheng, M. (2017). Big data management. In F. P. G. Marquez, & B. Lev (Vol. Eds.), *Information technology & tourism. Vol. 17. Information technology tourism* (pp. 455–456). Switzerland: Springer International Publishing 2017, 267 pp.

Coveney, P. V., Dougherty, E. R., & Highfield, R. R. (2016). Big data need big theory too. *Philosophical Transactions of the Royal Society A, 374*, 20160153. https://doi.org/10.1098/rsta.2016.0153 (last accessed 02-29-2020).

Drechsler, A. (2012). Design science as design of social systems – Implications for information systems research. In K. Peffers, M. Rothenberger, & B. Kuechler (Eds.). *Design science research in information systems. Advances in theory and practice* (pp. 191–205). Berlin-Heidelberg: Springer.

Elragal, A., & Klischewski, R. (2017). Theory-driven or process-driven prediction? Epistemological challenges of big data analytics. *Journal of Big Data, 4*, 19. https://doi.org/10.1186/s40537-017-0079-2 last accessed 02-29-2020.

Ferguson, G., Megehee, C. M., & Woodside, A. G. (2017). Culture, religiosity, and economic configural models explaining tipping-behavior prevalence across nations. *Tourism Management, 62*, 218–233.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137–144.

Godfrey-Smith, P. (2003). *Theory and reality, an introduction to the philosophy of science.* Chicago: The Univ. of Chicago Press.

Greckhamer, T., Misanyi, V. F., & Fiss, P. C. (2013). The two QCAs: From a small-N to a Large_N set theoretic approach. *Configurational Theory and Methods in Organizational Research Research in the Sociology of Organizations, 38*, 49–75. https://www.emerald.com/insight/content/doi/10.1108/S0733-558X(2013)0000038007/full/html, Accessed date: 29 February 2020.

Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google analytics. *Annals of Tourism Research, 61*, 199–212.

Günther, W. A., Rezazade, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems, 26*, 191–209.

Hubbard, R. (2004). Alphabet soup, blurring the distinction between *p*'s and *α's* in psychological research. *Theory & Psychology, 14*(3), 295–327.

Kim, J., & Fesenmaier, D. R. (2017). Measuring human senses and the touristic experience: Methods and applications. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 47–63). Springer International Publishing Switzerland.

Kim, Y., Kim, C., Lee, D. K., Lee, H., & Andrada, R. I. T. (2019). Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management, 72*, 249–256.

Kirilenko, A. P., & Stepchenkova, S. O. (2017). Sochi Olympics on twitter: Topics, geographical landscape, and temporal dynamics. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 215–234). Springer International Publishing Switzerland.

Kitchin, R. (2014). *Big data, new epistemologies and paradigm shifts.* Big Data & Society1–12. April–June 2014 https://journals.sagepub.com/doi/full/10.1177/2053951714528481, Accessed date: 29 February 2020.

Kordig, C. R. (1971). The theory-ladenness of observation. *The Review of Metaphysics, 24*(3), 448–484.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* 2nd ed. 1970Chicago: Univ. of Chicago Press.

Lee, I., Cai, G., & Lee, K. (2014). Exploration of geo-tagged photos through data mining approaches. *Expert Systems with Applications, 41*(2), 397–405.

Li, D., & Yang, Y. (2017). GIS monitoring of traveler flows based on big data. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 111–126). Springer International Publishing Switzerland.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301–323.

Little, J. D. C. (1970). Models and managers: The concept of a decision calculus. *Management Science, 16*(8), B–466-B-485.

Locke, E. A. (2007). The case for inductive theory building. *Journal of Management, 33*(6), 867–890.

Lukosius, V., & Hyman, M. R. (2019). Marketing theory and big data. *Journal of Developing Areas, 53*(4), https://doi.org/10.1353/jda.2018.0082.

Ma, S., Kirilenko, A. P., & Stepchenkova, S. (2020). Special interest tourism is not so special after all: Big data evidence from the 2017 great American solar eclipse. *Tourism Management, 77*, 1–13 https://www.sciencedirect.com/science/article/abs/pii/S0261517719302195?via%3Dihub, last accessed 02-29-2020.

Mandal, S. (2018). Exploring the influence of big data analytics management capabilities on sustainable tourism supply chain performance: The moderating role of technology orientation. *Journal of Travel & Tourism Marketing, 35*(8), 1104–1118.

Marchiori, E., & Cantoni, L. (2017). Evaluating destination communications on the internet. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 253–279). Springer International Publishing Switzerland.

Marine-Roig, E. (2017). Online travel reviews: A massive paratextual analysis. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 179–202). Springer International Publishing Switzerland.

Marquez, F. P. G., & Lev, B. (2017). *Big Data Management.* Switzerland: Springer.

Mazanec, J. A. (1992). Classifying tourists into market segments: A neural networks approach. *Journal of Travel & Tourism Marketing, 1*(1), 39–59.

Mazanec, J. A. (2009). Unravelling myths in tourism research. *Tourism Recreation Research, 34*(3), 319–323.

Miah, S. J., Vub, H. Q., Gammack, J., & McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. *Information & Management, 54*, 771–785.

Nelson, R. (2016). Scientific (analytic) theory of polygraph testing. *APA Magazine, 49*(5), 69–82.

Pan, B., & Yang, Y. (2016). Forecasting Destination Weekly Hotel Occupancy with Big Data. *Journal of Travel Research, 56*(7), 957–970.

Pan, B., & You, Y. (2017). Conceptualizing and measuring online behavior through social media metrics. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 203–211). Springer International Publishing Switzerland.

Park, S. (2017). Estimating the effect of online consumer reviews: An application of count data models. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 147–163). Springer International Publishing Switzerland.

Pearl, J. (2001). *Causality: Models, reasoning, and inference, 2nd printing.* Cambridge: Cambridge University Press.

Plutchik, R. (1980). A general pschoevolutionary theory of emotion. In R. Plutchik, & H. Kellerman (Eds.). *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 3–34). New York: Academic.

Popper, K. (1959). *The logic of scientific discovery.* London and New York: Routledge.

Popper, K. R. (1934). *Logik der Forschung* (3rd ed.). Tübingen: J. C. B. Mohr1969.

Putnam, H. (1991 [1974]). The corroboration of theories. In R. Boyd, P. Gasper, & J. D. Trout (Eds.). *The philosophy of science* (pp. 121–137). MIT Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60*, 20–43.

Ragin, C. C. (n.d.). What is qualitative comparative analysis (QCA)?, http://eprints.ncrm.ac.uk/250/1/What_is_QCA.pdf, Feb 29, 2020.

Reichenbach, H. (1938). *Experience and prediction, an analysis of the foundation and the structure of knowledge.* Chicago: The Univ. of Chicago Press.

Scharl, A., Lalicic, L., & Önder, I. (2017). Tourism intelligence and visual media analytics for destination management organizations. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 165–178). Springer International Publishing Switzerland.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research, 70*, 263–286.

Song, H., & Liu, H. (2017). Predicting tourist demand using big data. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 13–29). Springer International Publishing Switzerland.

Spanos, A. (2008). Review of Stephen T. Ziliak & Deirdre N. McCloskey's the cult of statistical significance: How the standard error costs us jobs, justice, and lives. Ann Arbor (MI): The University of Michigan Press, 2008. *Erasmus Journal for Philosophy and Economics, 1*(1), 154–164.

Wan, S., & Law, R. (2017). Leveraging online reviews in the hotel industry. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 235–252). Springer International Publishing Switzerland.

Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes.* New York: W. W. Norton & Company48–71.

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing. 80*(Nov), 97–121.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog, & H. Wold (Vol. Eds.), *Systems under indirect observation: Causality, structure, prediction*. *Vol. 2*. *Systems under indirect observation: Causality, structure, prediction* (pp. 1–54). Amsterdam: North-Holland.

Woodside, A. G. (2016). Diffusion and adoption of good science: Overcoming the dominant logic of NHST and the reporting of rubbish. *Journal of Business-to-Business Marketing, 23*, 327–333.

Xiang, Z., & Fesenmaier, D. R. (Eds.). (2017). *Analytics in smart tourism design, concepts and methods*. Springer International Publishing Switzerland.

Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services, 22*, 244–249.

Xiang, Z., Schwartz, Z., & Uysal, M. (2017). Market intelligence: Social media analytics and hotel online reviews. In Z. Xiang, & D. R. Fesenmaier (Eds.). *Analytics in smart tourism design, concepts and methods* (pp. 281–295). Springer International Publishing Switzerland.

Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives.* Ann Arbor: The University of Michigan Press.

Josef A. Mazanec is professor emeritus of the Vienna University of Economics and Business (WU) and full professor at the Department of Tourism and Service Management at MODUL University Vienna. He functioned as head of the Institute for Tourism and Leisure Studies of WU (1981–2010). He was a visiting scholar at the Alfred P. Sloan School of Management, MIT (1992). During 1997–2002 he served as the Vice-Rector for Research of WU and in 1997–2000 as the Speaker of the Joint Research Program on “Adaptive Models and Systems in Economics and Management Science” (comprising researchers from WU, Vienna University of Technology, and University of Vienna). He is a founding member of the International Academy for the Study of Tourism and an editorial board member for numerous tourism and marketing journals. His research interests are in consumer behavior, multivariate methods, decision-support systems, and marketing and management science applications in hospitality and tourism.