© Academy of Management Journal
2019, Vol. 62, No. 4, 971–978.
https://doi.org/10.5465/amj.2019.4004

FROM THE EDITORS

NEW WAYS OF SEEING BIG DATA

Few topics have received as much recent attention from researchers across disciplines, practitioners, policymakers, and popular media as "big data." Yet, from our experiences on the *Academy of Management Journal* editorial team, we believe a great deal of ambiguity and even confusion still prevails around key questions such as: What does big data encompass? Does big data mean the end of theory? In what ways does big data research differ from conventional scientific methods of inquiry in management research? What does it take to publish a big data study in management journals?

Therefore, our aim in this editorial is to offer insights into big data aligned with our editorial team's focus on "new ways of seeing." We readily acknowledge that big data can stretch our theoretical reach and expand the repertoire of methodological approaches for studying management phenomena in new ways. As a pervasive but emergent business phenomenon, big data lends fruitful opportunities for management scholars to not only challenge, change, and extend existing theories, but also to inform the practice of big data through systematic investigation. Big data also provides valuable analytical and visualization tools to supplement, turbocharge, and even transform some areas of management research—such as the use of unstructured data, real-time data processing, and pattern recognition. At the same time, big data also makes it necessary to revisit some research assumptions, practices, processes, and tools developed against the backdrop of constrained data considerations.

We contend that the field will be in a stronger position to take advantage of big data opportunities—and to avoid the pitfalls—when we not only transfer knowledge from other disciplines, but also engage in the coproduction of knowledge on big data. To that end, we start by clarifying the logic of big data from a research perspective, arguing that management researchers may enrich the perspective in some important ways. We next outline research opportunities that can leverage the strengths of big data and management scholarship to mutual advantage. We then conclude the editorial with a host of suggestions for overcoming the basic barriers to publishing these research opportunities in the field's journals. Together, the three themes—(1) enriching the perspective, (2) leveraging strengths to mutual advantage, and (3) overcoming barriers to publishing—enable us to offer new insights about how and in what ways management scholarship might shape the content and evolutionary trajectory of knowledge on big data. They also enable an integrated discussion on the core issues of big data—the paradigmatic and methodological, as well as the conceptual and phenomenological.

Our message is one of optimism tempered with realism. We believe that innovative research approaches *concurrently* leveraging the power of big data and the plurality of theoretical and empirical approaches can complement to advance both management research and big data practices. But, even as the need and opportunities for such innovations are manifold, they remain complex, challenging, and perhaps risky pursuits for individual researchers. We hope that this editorial, then, will serve as a springboard for those wishing to move the conversation from a one-way emphasis on the implications of big data to a two-way dialogue for advancing both big data and management scholarship.

ENRICHING THE PERSPECTIVE

A bigger-picture reflection on big data as a research approach should, we suggest, be a part of any dialogue on big data in the field because of the implications it holds for research questions, model construction, designing research, data collection, and analyzing and visualizing data. The perspective has been characterized in several ways, as follows: (a) from theory or small-sample data to be interpreted by humans to processing huge amounts of data to reach data-driven discoveries (Elragal & Klischewski, 2017); (b) from causality to patterns and correlations in the data (Mayer-Schönberger & Cukier, 2013); (c) from testing a theory to insights born from the data (Kitchin, 2014); and (d) the prominence and status acquired by data as commodity and recognized output (Leonelli, 2014). These all seem reasonable descriptions, but, in our judgment, what truly anchors the approach is the "law of large numbers"—the notion that, with enough data and samples, errors (uncertainty) are bound to surrender to certainty (Succi & Coveney, 2018). As Cohen (2013: 1921) stated, "big data's claims to epistemological privilege stem from its asserted fidelity to reality at a very high level of detail."

There has been considerable concern with the perspective as a sort of "empiricism on steroids" that involves gathering and going through data to find patterns and making predictions about dependencies and causation (Frické, 2015; Sætra, 2018). We also observe that big data applications so far appear to have predominantly tackled *the question of what is happening now and likely happens next*. For example, a common focus is not on why a single variable might explain an outcome variable, but how the outcome varies with many potential predictors—with or without theory as to which predictors are relevant (Einav & Levin, 2014).

We would argue against the assertion that the perspective diminishes the importance of causal adequacy and depth in research. Because data are a means to an end, big data's informativeness to reach justifiable conclusions matters more than its volume, velocity, or variety (Bowman, 2018). A correlational finding, for example, may not morph into a causal one by simply increasing the volume, variety, and velocity of the underlying data. The real issue, however, is not data per se, but the perspective that undergirds the manner in which data are considered, collected, curated, and investigated (Coveney, Dougherty, & Highfield, 2016). Specifically, we concur with others that the claim that researchers need not start with theory but could rather acquire more objective insights and explanation from big data models and analyses is tenuous and unconvincing (Chan & Moses, 2016; Sætra, 2018).

To the contrary, given the complexity and resource requirements of accessing and processing big data sets, it seems to us asking the right questions is crucial. With no theory guiding the questions, an explanation of *what is going on, and why*, may not be adequately addressed. Moreover, an enhanced ability to detect correlations and clusters in the data can hardly substitute for theory to provide a stronger foundation with which to avoid errors and derive appropriate inferences from these correlations. Without theory, thus, pure big data approaches in the management field could routinely fail to provide conceptual accounts for the managerial phenomena and processes to which they are applied—as has been observed with some other disciplines as well, such as biology and medicine (Coveney et al., 2016).

Indeed, the ideal of pure empiricism or pure induction seldom works per se, and no theory can be so good as to supplant the need for data and testing (Calude & Longo, 2017). Thus, maybe the interpretation and use of big data as a perspective should resemble what is generally seen as "abduction" (the combination of deductive and inductive logics to derive causal inferences). If so, what are the implications for our predominant model of knowledge production and use? Abductive research involves a logic of discovery and doubt (Locke, Golden-Biddle, & Feldman, 2008), and such dispositions and capabilities warrant further attention with big data use. For example, how might big data patterns serve as a source for the development of a new theory, which is then further elaborated and tested deductively? And, more broadly, how might a big data perspective be made more theory driven for investigating managerial phenomena?

The question of when and under what conditions a big data approach could produce managerially actionable insights better than "smaller" high-quality data, and vice versa, is also intriguing to consider. We suspect that, as the situational complexity and ambiguity increases in an organizational decision, process, or system, the comparative advantage of big data may decrease, especially when data quality is mixed and systematic biases (unknown) exist. Collecting data from millions of individuals may provide little benefit in improving predictive accuracy, for example, if only a subset causes the most variance in the data. More broadly, big data might not perform well if data quality does not permit true replicability of the models and a rich understanding of the specific sources of instability in the models (Oswald & Putka, 2016). As Succi and Coveney (2018: 11) observed:

> In the end, most of [big data] comes down to more or less sophisticated forms of curve fitting based on error minimization. Such minimization procedures fare well if the error landscape is smooth, but they exhibit fragility towards corrugated ones in other situations, which are the rule in complex systems.

Finally, the perspective inherently demands that the process of data exploration be contextually informed, but the wider context is often entirely sidestepped. What might management research—in particular, qualitative researchers—say about the how and why of the context in the perspective's enrichment? Without such enrichments, big data models in management research could experience slow progress and a higher failure rate, as well as hindering the researcher's ability to understand the failures' root causes.

LEVERAGING STRENGTHS TO MUTUAL ADVANTAGE

Beyond a richer perspective, we also encourage attention to big data research aligned with the field's scholarly strengths and priorities. This is where the two-way dialogue could lead to specific advances and enable management researchers to envisage the diverse routes for a sustainable synthesis of management and big data scholarship. To facilitate an organized approach, we next discuss a framework of big data as a concept, methodology, and phenomenon.

Concept

Given big data's diverse uses across settings, disciplines, and applications, the concept is in danger of becoming "everything and nothing." The popular definition in terms of data properties such as volume and variety has created ambiguity about what might count as big data. For example, it is not entirely clear what determines the threshold to qualify data as "big" across different settings and applications. Management researchers with a strong emphasis on clearer definitions and constructs could help advance the current definitional ambiguity by moving the conversation toward more encompassing understanding on the domain, boundaries, and precision of big data concepts and constructs. Our own working definition is to view big data as a label that refers to the generation, organization, storage, retrieval, analysis, and visualization of data sets involving large volumes and a variety of data, involving new kinds of methodological, epistemological, and politico-ethical issues and questions.

Relatedly, even as researchers have devoted attention to the dimensions of big data, a consensus is yet to emerge. Three are prevalent: volume (the magnitude of data), variety (structural heterogeneity in a data set), and velocity (the rate at which data are generated and speed at which they are analyzed and used) (Tonidandel, King, & Cortina, 2018). But, researchers have also advanced other dimensions (curiously, many start with the letter "v"), such as veracity, vision, visibility, and value, among others. Each dimension poses distinct challenges and ways to overcome them for researchers and managers in accessing, storing, and utilizing big data. For example, the velocity dimension is associated with issues such as transfer speed, storage scalability, and timing, while veracity comes with issues such as uncertainty, authenticity, trustworthiness, and accountability. An examination of substantive research questions, the level of analysis, and the theoretical lenses used to construct hypotheses and propositions call for clarity regarding these dimensional manifestations. Such lower-level ordering, classification, or other aggregation of issues and characteristics across big data dimensions could also serve as a foundation for the development of clearer big data constructs for testing. A conceptual understanding of big data characteristics that could help with the generation of big data sets with a story about managers and organizations represents a promising direction for qualitative researchers in the field.

Methodology

Big data studies commonly begin with a researcher having access to a data source or a data set on a phenomenon, rather than with theory (Johnson, Gray, & Sarker, 2019). Thereafter, the analysis process involves specific issues in data access and clean up, search, and processing that are different from conventional approaches. Executing the phases might call for distinct computational and programming skills (e.g., R and Python). Data for "smaller" research are normally produced in structured ways and captured at certain point(s). A key challenge for the big data methodology is how to integrate and store structured and unstructured data in a way that would make the later analyses and visualization efficient and secure. Another challenge is that big data sets are often not created to examine specific questions and constructs. Thus, the researcher must deal with various issues pertaining to data construction and quality.

It is across these challenging methodological phases of big data where we would encourage management researchers to attain a greater understanding of the advantages and disadvantages of "starting with theory" in big data studies. Precisely in what ways (and when) might theory help to guide the various decisions pertaining to the cleaning, construction, aggregation, and storage of big data sets? For example, a multi- or meso-level theory could inform the decision of whether a big data set should be constructed as "horizontally deep" (many variables but fewer observations) rather than "vertically deep" (fewer variables but many observations). We would also encourage researchers to develop a deeper understanding of each facet of the methodological process. It could, for example, be a productive practice for the field if big data studies were to routinely contain a summary of methodological steps undertaken, including the challenges encountered and solutions implemented.

Even as it might be possible to examine some large data set using traditional statistical and computational techniques, many do not scale to diverse and unstructured data sets. Statistics focuses overwhelmingly on inferences from data, while computational architectures and algorithms that can extract and discern valuable knowledge from complex data sets are among the key considerations in big data approaches. These architectures and algorithms are used to analyze big data sets for specific purposes, such clustering, pattern identification, and prediction. Some of the techniques include data mining, machine learning, neural networks, and deep learning (convolutional, deep belief, and recurrent nets). We also observe that a straightforward application of some of these techniques, especially unsupervised machine learning, in management research could result in several challenges. For example, one strength of deep learning techniques is to search for and then extract patterns from unstructured data sets. By this means, questions might be raised concerning, for instance, how to build an explanatory model around a pattern, and how to communicate the boundaries and constraints of the final model.

What is also obvious to us, after reviewing the relevant research, is that these techniques tend to be highly specialized across different research tasks and evolve dynamically, which makes it difficult for individual researchers to make use of the potential of these techniques. For example, big data visualization techniques demand computational, statistical, and informational knowledge. Another set of challenges could come from the required computing power and infrastructure, which might be not easily accessible to individual researchers. Together, we thus suggest that management researchers need to develop a more systematic understanding of the advantages and disadvantages of the available big data analytical techniques in the context of management studies and phenomenon—for example, how might the field's empirical approaches be combined with big data techniques such as experimental data or findings and subsequent applications of machine learning techniques to attain more generalizable insights? Or, how might the field take advantages of the techniques to calibrate covariates, or address multicollinearity for "smaller" data studies in which the outcome variable is complex and distal, such as organizational performance? Relatedly, the field could benefit from a richer understanding of the challenges and opportunities of investigating the findings from the machine learning and other predictive techniques. For example, in what ways might unsupervised machine learning techniques be combined with qualitative research, such as using the clusters and patterns to inform concepts selection, aggregations, and initial themes?

It is also noteworthy that different technologies and platforms might be appropriate for a variety of purposes across the methodological tasks, such as storage, access, and processing of data. While some platforms exist and provide impressive capabilities for processing big data sets, management researchers will need to obtain a more complete understanding of how to choose among the ever-expanding menu of big data technologies. Thus, we encourage management scholars to devote more attention to new research designs and analytical complementarities at the nexus of conventional and big data methodologies. We also encourage more attention toward understanding the advantages and disadvantages of big data techniques and technologies in the context of managerial phenomena, individually and comparatively vis-à-vis conventional techniques such as multivariate statistics.

Phenomenon

Big data practices and applications have occurred in diverse settings, industries, and economies. We thus suggest that management scholars can and should focus special attention on big data as a phenomenon in organizations, institutions, and societies. There is a great need for understanding where and in what organizational and industrial contexts big data applications might be more consequential for organizations and managers. More broadly, we believe that the field should lead in the development of new theories, approaches, and frameworks that could help managers and their firms to better use and extract value from big data. For example, how might big data technologies and tools be used in support of corporate strategy, such as by integrating diverse data technologies across business units? A related question concerns the firm's strategic choice regarding "where and how to play" in the big data space. From a decision-making perspective, machine-learning approaches that automatically identify actionable patterns could help to alleviate some of the cognitive burden on managers. This potential raises several intriguing questions. What is the nature and consequence of the trade-off between bounded executive cognition and cognitive requirements of big data? How might a capability to quickly analyze and visualize patterns hidden in big data shape the quality and speed of decision-making? What types of managers are more likely to embrace (or avoid) big data in making decisions? Addressing these questions could lead to new theory or refinements to existing theories such as the resource-based view, organizational learning, upper-echelons, among others, as well as help managers improve use of big data in their own decision-making.

Some researchers have argued that the notion of big data as objective and fact based is a myth (Gitelman, 2013). Given the possibility for a subjective interpretation, individual micro-foundations might be crucial in understanding big data processes and uses in organizations. Several broad questions beg attention: How do individuals and groups choose and interpret big data? What are some of the psychological barriers to individuals' adoption of big data? These questions could be investigated by drawing on a range of distinct theoretical lenses. Attention-based perspectives, judgment and heuristic theories, and counterfactual thinking could be especially pertinent in understanding how individuals might utilize and interpret big data.

Big data might also create some research opportunities around its own ecosystems. For example, whereas much has been said about how big data is revolutionizing management processes and how decision-making teams can benefit from using it, little has been said about the challenges and processes of the big data teams that generate and manage big data in organizations (Saltz, 2015). Understanding the novel interpersonal challenges that big data teams face is an important direction that also could lend considerable prescriptive value, such as in the context of new product development. More broadly, creating big data infrastructure requires senior executives to put in place appropriate structures and capabilities that support integration and unification of the many islands of data and analytical capabilities that could exist throughout the organization. At the same time, this integration creates several relational and cultural challenges, such as resistance to sharing and combining data because of organizational silos and disputes over the implications of the associated analytical insights. Galbraith (2014: 3) observed that, as organizations embrace big data, there is "a shift in power from experienced and judgmental decision-makers to digital decision-makers." How do organizations structure this shift in power? Does the typical top management team include a separate chief digital officer or does the chief information officer wear two hats: IT and big data? Another question concerns how organizations might create norms and values concerning information sharing, transparency, and trust.

We would be remiss not to touch upon the ethical and privacy issues surrounding big data. It has by now become clear that the generation and storage of big data sets involve more challenges than usually anticipated, as shown for instance with the recent scandals such as Cambridge Analytica. To begin with, the availability of data is not a guarantee that their use would be ethical or even legal. In addition, there are issues and contradictory demands of transparency and protection of individuals' identities and personal knowledge (e.g., Acquisti, Brandimarte, & Loewenstein, 2015). Although legal regulations and organizations codes can serve as helpful markers, individual differences are critical in understanding the propensity of individuals in going over and beyond the minimum compliance, or, alternatively, the tendency of individuals to engage in ethical wrongdoing with regard to big data acquisition and utilization. The standards and practices regarding individual data rights, ethics, and privacy are in a state of development and debate globally. These ethical complexities in big data provide opportunities to enrich theories in the areas of ethics and values, such as ethical leadership, moral values, and identity.

Moreover, although the ideals about big data speak of openness and access to all, this is not entirely the case. Big data is becoming an increasingly important business in which various actors not only control the databases but also regulate the marketing, sales, and use of such data and analytical capabilities (Cohen, 2013). Is this going to lead to asymmetric access and a new big data divide among researchers and practitioners, and within and across societies and nations more broadly? Several indications suggest that big data can lead to a "Matthew effect," by which we simply mean, to paraphrase Merton (1973), that the data-and-analytical-capability-rich might get richer, and the data-and-analytical-capability-poor might get poorer. Relatedly, research transparency and replication issues could become problematic if big data sets and the analytics that underpin them were to be kept secret for a variety of reasons, such as competitive advantage (Cohen, 2013).

OVERCOMING BASIC BARRIERS TO PUBLISHING

Our discussion on big data as a research perspective and its associated research priorities in the preceding sections make it clear that big data gives rise to some distinct issues at each stage of the research process and design—from starting and/or building theory, and accessing and integrating the data, to the analysis and reporting and visualization. It also seems to us that big data research is developing in a way that might be beyond a single researcher's capabilities and resources, due to data access and management, the required computational power, and the necessary knowledge of the analytical tools and techniques. We believe that researchers will also need to consider and overcome some rather basic barriers to publishing big data studies in the field's journals.

First, big data cannot substitute for careful and credible research designs and the appropriate consideration of research questions. With no clear and theoretically pertinent question guiding their creation and preparation, big data sets might come across as a large convenience sample or a "fad." A key question for researchers therefore is this: Why is big data most appropriate in studying the research question of interest? Researchers may thus have to provide additional justification for the way and the types of data and variables collected, constructed, and aggregated. We would particularly encourage that researchers incorporate (explicitly or implicitly) the logic of data access and collection, integration and aggregation, analysis, and reporting and visualization to craft and communicate the research design of their big data studies.

Second, it may be difficult if not impossible for reviewers and other authors to replicate and extend studies if there is little transparency about how the data are created, manipulated, and/or analyzed. Private companies often own and store big data sets. Without some built-in quality checks and controls, reliabilities and validities of variables might be systematically compromised. Systematic errors cannot be resolved by collecting more of the same data. Reviewers and readers are used to seeing empirical studies that typically use small samples wherein the variables are operationalized in a specific fashion. While some variables may have face validity and require less justification, researchers might have to find solutions about the operationalization of latent and profile constructs embedded in big data sets. One solution is to use the small sample contexts to establish the validity of those measures before employing them in the big data contexts. Another solution might be to combine big data analysis with other methods—either quantitative or qualitative—to establish validity and/or illuminate the key processes or mechanisms at play. Yet another option is to work closely with practitioner experts to ensure strong face validity of the assumptions and approaches used by the researchers.

Third, the selection of constructs or variables in the current empirical approaches is typically done with guidance from the underlying theory. With big data, the process of converting data into constructs of interests can lack clarity and transparency because some associated techniques, such as machine learning, might be barely guided by an explicit theory. Here, one could distinguish between supervised and unsupervised learning techniques. In supervised techniques, the researcher could specify the variables to be incorporated into the model, and, so, the approach is like the conventional small-sample research. But, in unsupervised algorithms, the algorithm will select the variables from the available variables to be included in the model. Reviewers and readers are not used to seeing papers that select variables in such "random" (from the perspective of the small-sample research paradigm) fashion. Relatedly, given that the existing paradigm uses control variables in regressions to control for alternative influences correlated with the explanatory variables, researchers must pay attention to how they can convince reviewers that the patterns of associations found from the data are reasonable and are not just associations "by chance."

Fourth, whatever analytical technique(s) is utilized, but especially for machine and deep learning, we encourage researchers to describe the content and process of specific variables and associations examined, rather than having them obscured within a "computational black box." The unsupervised and deep machine-learning techniques, by automating multiple hypothesis testing with opaque modifiers and biases, could in fact convolute the meaning of constructs and predictions—with the added risk of spitting out spurious correlations at an unprecedented scale. A related concern is that, because technologies and techniques of big data are rapidly changing, researchers might rely on outdated techniques and modeling. The selected tools and technique will need to be justified vis-a-vis the study's question and testing needed for a credible answer.

Fifth, most big data approaches employ predictive techniques rather than statistical inference approaches. So, scholars employing big data approaches must convince reviewers and readers that the approaches are equally good, if not better, for testing the theories, relative to the statistical inference approaches. Moreover, when simultaneous associations among multiple variables need to be presented in a single model, researchers must present them in a format understandable to reviewers trained in different paradigms. Finally, because statistical significance is irrelevant with massive sample sizes, researchers should work to justify and demonstrate the importance of the findings—for example, with effect sizes and contextualized significance. Visual representation of the results would also likely be a necessary approach.

Despite these pitfalls, hurdles, and challenges, we argue that management scholarship will be in stronger position to the extent that it not only transfers relevant knowledge on big data into the field, but also actively shapes the content and evolutionary trajectory of that knowledge. We have discussed herein several directions for synthesizing the strengths of big data and management scholarship to mutual advantage. The various paradigmatic, conceptual, methodological, and phenomenological issues surrounding big data also signify to us that individual researchers will need to weigh the benefits and risks and proceed cautiously when pursuing big data research.

Zeki Simsek
Clemson University

Eero Vaara
Aalto University School of Business

Srikanth Paruchuri
Pennsylvania State University

Sucheta Nadkarni
University of Cambridge

Jason D. Shaw
Nanyang Technological University

REFERENCES

Acquisti, A., Brandimarte, L., & Loewenstein, G. 2015. Privacy and human behavior in the age of information. ***Science***, 347: 509–514.

Bowman, A. W. 2018. Big questions, informative data, excellent science. ***Statistics & Probability Letters***, 136: 34–36.

Calude, C. S., & Longo, G. 2017. The deluge of spurious correlations in big data. ***Foundations of Science***, 22: 595–612.

Chan, J., & Moses, B. L. 2016. Is big data challenging criminology? ***Theoretical Criminology***, 20: 21–39.

Cohen, J. 2013. What privacy is for. ***Harvard Law Review***, 126: 1904–1933.

Coveney, P. V., Dougherty, E. R., & Highfield, R. R. 2016. Big data need big theory too. ***Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences***, 374: 20160153. Retrieved from https://royalsocietypublishing.org/doi/full/10.1098/rsta.2016.0153.

Einav, L., & Levin, J. 2014. The data revolution and economic analysis. ***Innovation Policy and the Economy***, 14: 1–24.

Elragal, A., & Klischewski, R. 2017. Theory-driven or process-driven prediction? Epistemological challenges of big data analytics. ***Journal of Big Data***, 4: 19. Retrieved from https://link.springer.com/article/10.1186/s40537-017-0079-2.

Frické, M. 2015. Big data and its epistemology. ***Journal of the Association for Information Science and Technology***, 66: 651–661.

Galbraith, J. R. 2014. Organization design challenges resulting from big data. ***Journal of Organization Design***, 3: 2–13.

Gitelman, L. (Ed.) 2013. ***"Raw data" is an oxymoron*** Cambridge, MA: MIT Press.

Johnson, P., Gray, P., & Sarker, S. 2019. Revisiting IS research practice in the era of big data. ***Information and Organization***, 29: 41–56.

Kitchin, R. 2014. Big Data, new epistemologies and paradigm shifts. ***Big Data & Society***, 1: 1–12.

Leonelli, S. 2014. What difference does quantity make? On the epistemology of Big Data in biology. ***Big Data & Society***, 1: 1–11.

Locke, K., Golden-Biddle, K., & Feldman, M. 2008. Making doubt generative: Rethinking the role of doubt in research process. ***Organization Science***, 19: 907–918.

Mayer-Schönberger, V., & Cukier, K. 2013. ***Big data: A revolution that will transform how we live, work, and think***. Boston, MA: Houghton Mifflin Harcourt.

Merton, R. K. 1973. ***The sociology of science: Theoretical and empirical investigations***. Chicago, IL: University of Chicago Press.

Oswald, F. L., & Putka, D. J. 2016. Statistical methods for big data: A scenic tour. In S. Tonidandel, E. B. King, &

J. M. Cortina (Eds.), ***Big data at work: The data science revolution and organizational psychology***: 43–63. New York, NY: Routledge.

Sætra, H. K. 2018. Science as a vocation in the era of big data: The philosophy of science behind big data and humanity's continued part in science. ***Integrative Psychological & Behavioral Science***, 4: 508–522.

Saltz, J. S. 2015. The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. In IEEE Computer Society (Ed.), ***2015 IEEE international conference on big data:*** 2066–2071. Los Alamitos, CA: IEEE Computer Society.

Succi, S., & Coveney, P. V. 2018. Big data: The end of the scientific method? ***Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences***, 377: 20180145. Retrieved from https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0145.

Tonidandel, S., King, E. B., & Cortina, J. M. 2018. Big data methods: Leveraging modern data analytic techniques to build organizational science. ***Organizational Research Methods***, 21: 525–547.