Journal of the Association for Information Systems (2020) **21**(2), 268-291
doi: 10.17705/1jais.00601

EDITORIAL

ISSN 1536-9323

The Perils and Promises of Big Data Research in Information Systems

Varun Grover[1], Aron Lindberg[2], Izak Benbasat[3], Kalle Lyytinen[4]
[1]University of Arkansas, USA, vgrover@uark.edu
[2]Stevens Institute of Technology, USA, aron.lindberg@stevens.edu
[3]University of British Columbia, Canada, izak.benbasat@sauder.ubc.ca
[4]Case Western Reserve University, USA, kalle@case.edu

Abstract

With the proliferation of "big data" and powerful analytical techniques, information systems (IS) researchers are increasingly engaged in what we label as big data research (BDR)—research based on large digital trace datasets and computationally intensive methods. The number of such research papers has been growing rapidly in the top IS journals during the last decade, with roughly 16% of papers in 2018 employing this approach. In this editorial, we propose five conjectures that articulate the potential consequences of increasing BDR prevalence for the IS field's research goals and outputs. We discuss ways in which IS researchers may be able to better leverage big data and new analysis techniques to conduct more impactful research. Our intent with these conjectures and analyses is to stimulate debate in the IS community. Indeed, we need a productive discussion about how emerging new research methods, digital trace data, and the development of indigenous theory relate to and can support one another.

Keywords: Big Data Research, IS Research, Theory, Empiricism, Digital Traces, IT Artifact

Dorothy E. Leidner was the accepting senior editor. This article was submitted on December 17, 2019 and underwent two revisions.

1 Introduction

Contemporary information systems (IS) research seems to be increasingly turning toward big data and many groups within the IS community have welcomed this change. The pervasiveness of digital phenomena—including social media, mobile commerce, analytics, machine learning, cloud, and the internet of things—has led to the emergence of massive datasets and access to increasingly sophisticated analytical techniques. This is creating a different texture for some IS research (Abbasi, Sarker, & Chiang, 2016), one that capitalizes on big data and explores unprecedentedly large datasets and/or uses computationally intensive analysis techniques (e.g., neural nets, machine learning, dynamic algorithms, etc.). We refer to this research orientation as big data research (BDR).

Overall, we welcome BDR as a positive development in that it opens new vistas for the study of IS. It has enabled novel forms of evidence provision and attendant theory development in both the natural and the social sciences (Lazer et al., 2009). Expectations are high for the potential of big data to advance our understanding of science and business (Bell, Hey, & Szalay, 2009; Dhar, 2013; Maass et al., 2018). Our sister business disciplines, e.g., marketing (Sudhir, 2016), organization science (Davis, 2015), and management (Simsek et al., 2019), have debated the benefits of BDR for several years. Similarly, some IS scholars have raised issues around the costs and benefits of BDR (Abbasi et al., 2016; Johnson, Gray,

& Sarker, 2019) while others examine how big data traces can be used effectively to study specific research questions (e.g., Lindberg, 2020; Østerlund, Crowton, & Jackson, 2020). Albeit highly valuable, especially in illuminating how BDR can advance the knowledge frontiers of our field by tackling new kinds of research problems, or through approaching old problems with a new type of vigor, the debate around BDR so far has not offered an evidence-based analysis of how BDR may influence the varied forms of knowledge production within our field. We argue that this question is important for the IS research community and that it calls for an examination of how the community can learn to balance the benefits and potential limitations of BDR as it continues to gather momentum.

As a preface, we would like to state the boundary conditions of our arguments. First, we define BDR broadly as research that involves large and often heterogeneous datasets represented in multiple formats (qualitative, quantitative, video, image, audio, etc.). These datasets are obtained mainly from the digital traces left behind by various groups of users (including bots and other relatively autonomous software or hardware agents) interacting with online platforms. These data traces are mainly noninvoked, capture naturally occurring interactions, and are therefore collected for purposes unrelated to academic research (Howison, Wiggins, & Crowston, 2011). Importantly, such data corpora are reflective of the contemporary explosion in the volume, variety, and velocity of the data available (George et al., 2016). This can be contrasted with "small" data studies that use either structured (e.g., survey instruments) or nonstructured data (e.g., interviews), mostly collected by the researchers themselves from primary or secondary sources. Such small data studies have largely characterized traditional IS research over its first 50 years.

Second, consistent with the editorial missions of our major journals, we assume that the IS field is fundamentally geared toward providing "theoretical insights that advance our understanding of information systems and information technology in organizations and society."[1] Hence, any assessment of BDR needs to be made based on the contribution to generalized knowledge within this domain space.

Third, while we are optimistic about the potential of big data in IS research, we are nevertheless concerned about certain BDR practices. Unprecedented access to large datasets, which can be retrieved at relatively low cost from repositories, archives, or websites, present significant and novel opportunities for IS researchers. Large sample sizes will have a positive and direct relationship with the desire for higher power in statistical tests; that power can be further catalyzed by sophisticated, well-matched analysis techniques. Importantly, well-conceived BDR practices may not only be instrumental in various stages of scholarship involving various forms of reasoning (i.e., inductive, deductive, or abductive) but, as we will discuss later, they may also contribute directly to theory development through the discovery of anomalies, alternative conceptualizations of constructs, and new kinds of powerful field experiments. While we continue to have legitimate debates about the relative value of significance tests vs. effect sizes in this new environment, there are other potentially troubling signs associated with the recent rise of data munificence. The ready access to big data and powerful analysis methods may also encourage researchers to rely on size as a justification for the novelty of datasets and to draw on the power of computationally intensive methods to impress the reader. If this is, indeed, a strong tendency in BDR, it may come at the cost of neglecting deeper intellectual engagement with substantive research questions that provide disciplinary value through their long-term relevance to the corpus of cumulative IS knowledge.

While such patterns may or may not be at play in the IS field at large, there is anecdotal evidence that this is indeed happening. When three of the authors of this paper participated in the 2017 ICIS doctoral consortium in Seoul, they heard impressive presentations and discussions of multiple novel techniques. However, many doctoral students seemed somewhat oblivious to the importance of theory and whether or not the questions they sought to answer had a pertinent audience in the IS community, whether academic or practice oriented. Some studies deployed massive digital traces and/or archival data, but the research problems addressed appeared more appropriate for urban planning, consumer marketing, or health sciences. As similar fruits of BDR gain more prominence in our journals and body of knowledge, future generations of IS scholars may unintentionally inherit a brave new world of research where big data, computationally intensive analysis techniques, and evidence triangulation will reign over theory, disciplinary relevance, and the importance of having a cumulative tradition.

These observations and the potential concerns related to their long-term effects on the future of our discipline motivate this editorial. We hope to dig deeper into the conduct of BDR in the field and its impact on the field's research mission and output and we aspire to start a conversation on how to chart the future directions of our research that leverages the rise of large, novel data pools and powerful analysis

[1] https://aisel.aisnet.org/jais/authorinfo.html

techniques. To this end, *we are deliberately taking a critical stance in our framing of BDR and its outcomes*, which we then subject to examination using emerging evidence from some of our major outlets. This allows us to provide constructive, evidence-based guidelines for the field on how to respond to the challenge of BDR in a productive manner. Specifically, we address three key research questions:

1. How is BDR currently being practiced as manifested in the output published in our major outlets?
2. What are the potential implications of such practices for the IS field and its future?

And, if we observe some potentially harmful outcomes and tendencies that violate the goals of cumulative, theory-based (impactful) research:

3. How can IS researchers better leverage the recently emergent, extraordinary access to data and analysis techniques while conducting impactful IS research?

To answer these research questions, we begin by explaining why BDR has been readily embraced by the field. Next, we formulate five conjectures that project the possible consequences of BDR as currently practiced on the type of knowledge produced in our field. We then analyze and compare 41 BDR studies and 41 non-BDR studies to provide evidence in support of our conjectures. We conclude by broadening the discussion about what we can do, both as individual IS scholars and as a collective, to leverage the potential of BDR more effectively.

2 Why Has BDR Been Embraced by the IS Field

In a relatively short period of time, the IS field seems to have readily embraced BDR. This is reflected in the multiple special journal issues that have focused on big data and related topics (e.g., *MISQ* 2016, *JMIS* 2018)[2] as well as broader publication trends. The number and proportion of BDR articles in our major journals has been steadily rising: we found that 41 of the 392 papers (ca. 10%) in our sample were BDR, whereas ten years ago (ca. 2008-2009), such articles were rare. There are also indications that with the current proliferation of big data and extraction techniques, this trend is accelerating.[3] These observations prompt questions regarding the reasons for the burgeoning quantity and significance attributed to such research. To explain this, we offer a number of arguments for why BDR is attractive to the IS field.

Impressive datasets: We are moving from a period of data deficiency to one of data abundance.[4] Large sample sizes that are readily available appeal to both authors and reviewers. The availability of such data provides increasing opportunities for researchers to interact dynamically with a dataset. Rather than having to do a "one-shot" data collection, as is the case in traditional survey or experimental research, in some cases researchers can now continuously interact with a digital dataset to further refine their analysis based on comments from reviewers. The size and uniqueness of such datasets also make studies seem more comprehensive and "scientific," thus impressing reviewers and readers. Sophisticated, computationally intensive methods also contribute to this impression.

Increased ease of demonstrating statistical significance: As a research field we have an institutionalized belief that supported hypotheses with significance—notwithstanding low-effect sizes and practical significance—is preferable to non-statistically significant ones with higher effect sizes. Using massive datasets, statistical significance can almost always be demonstrated, even if the effect sizes are negligible.

The lure of objective data: As digital trace datasets are not collected for the purpose of research, researcher-induced bias is eliminated (Lindberg, 2020). Such data also represent behavior captured "in the wild," and therefore serve to increase the real-world realism of the studies conducted using such datasets. Hence, such data are believed to increase the overall validity of the findings.

The availability of powerful tools for analyzing large datasets: Large datasets demand large computational power, which now can be supplied by either GPU-driven desktop work stations or by using computing clusters. The latter are hosted by many universities or rented from platforms such as AWS or Google Cloud. Furthermore, new computationally intensive techniques can analyze not only patterns of correlation between predetermined variables, but also construct models of social networks, processes, text, and a myriad of other features that can be inferred from digital trace datasets.

[2] *MIS Quarterly* (2016): Special Issue: Transformational Issues of Big Data and Analytics in Networked Business; *Journal of Management Information Systems* (2018): Special Issue: Strategic Value of Big Data and Business Analytics.

[3] Top journal receptivity to BDR suggests that academics focused on such research are better positioned to be tenured in their institutions, as publishing in journals ranked on the FT50 list has become a necessity in many academic institutions.

[4] There is a caveat to this. Corporate data, being the "new oil" is increasingly being monetized and may therefore become less free and accessible in the future. However, there will likely be a growth of digital traces that are readily accessible, at least in some form.

Increased synergy between teaching, research, and consulting: In the burgeoning area of business analytics, research focused on analyzing data, as opposed to developing theory, can more readily cross over to teaching and help researchers develop consulting assignments. This advantage has traditionally not been associated with many IS research topics. A possible exception in the past could be found in systems analysis and design or security, neither of which has received the level of attention in the research community currently garnered by BDR.

IRB approval issues: Since BDR does not involve data collection from human subjects, most of these studies fall out of IRB jurisdiction unless there are privacy concerns. While the ethical frameworks for big data are being challenged, researchers leveraging BDR can avoid long IRB delays.

In many ways, BDR marks a departure from the way IS research has been conducted in the past. We also have little doubt that datasets will continue to increase in size and richness and that computational analysis methods will continue to increase their capabilities, thus making this trend and the underlying reasons to use BDR even more powerful in the future. These developments beg the question: What are the potential implications of embracing BDR for the mode of knowledge production and related outputs of our field?

3 Our Conjectures: What Are the Consequences of the Proliferation of BDR for the IS Field?

In assessing the implications of BDR for the IS field, especially its mode of knowledge production and outputs, we argue that the field ultimately needs to deal with some core matter against which the value of any knowledge claim or related contribution can be assessed by the IS community. In our view, this core matter can be loosely articulated as knowledge concerning the "development of IT-based services, the management of IT resources, and the use, impact, and economics of IT with managerial, organizational, and societal implications."[5] There are indeed disagreements on how restrictive this domain of coverage should be and whether or not there are core theories within our field. However, we all agree that to make unique and impactful contributions and to survive in the long run, preserving and possibly expanding its core matter and related topics is a critical success factor for the IS field. Uniqueness implies that our intellectual contributions differ and are original when compared to contributions offered by other fields. By impactful, we mean research output that has salience and offers utility to key stakeholders by describing, explaining, and accounting for the focal phenomenon of interest. There are other requirements: The IS field needs to address important and salient problems within its selected core domain, and such research streams should be synergistic and offer cross-pollination of results and findings. IS scholars also need to strive for generalizability and cumulative and theoretically valid knowledge while acknowledging the need for accuracy related to theories that cover our core matter.

Based on these requirements, we propose five conjectures regarding BDR's potential implications for the IS field, assuming that the conduct of BDR continues along its current trajectory. We use the term "conjectures" to denote theory-free suppositions formed on the basis of the currently incomplete information that we currently hold about BDR and its potential impact. Our conjectures represent a "prescientific" understanding along with related explanations and predictions of the impacts of BDR in our field—essentially presenting our best assessment of the likely consequences of BDR given our understanding of how research knowledge is currently produced in our field. In our findings section, we provide evidence based on our comparison of published BDR and non-BDR papers. While the conjectures we put forth below may appear controversial, we envision them as *bold* suppositions, i.e., both thought-provoking and salient with regard to our modes of knowledge production. Because of their nature, they should also be capable of inducing reflection and discussion within our community. By adopting a bold approach that seeks controversy, we hope to stimulate a healthy debate on this vitally important topic.

3.1 Example Five Controversial Conjectures Concerning the Impact of BDR on IS Research Outcomes

The five conjectures that we propose are largely driven by how we understand and characterize the nature of BDR, in terms of its leveraging of big data as well as the novel techniques used for analysis. We acknowledge that we are deliberately provocative in adopting a narrow focus on *data-driven* practices that relate to BDR and that highly nuanced ways of leveraging big data are being developed every day. Thus, our analysis does not apply to all instances of BDR. We are, however, more interested in broadly manifested tendencies that characterize such data-driven practices of BDR. This approach serves our

[5] Stated in the editorial objective of *MIS Quarterly*, one of the leading AIS journals, see https://misq.org/about/

intention to characterize critical contours of BDR and contrast them with those of non-BDR, thereby identifying potential consequences of the proliferation of BDR as currently practiced for the IS field and its research.

As noted, the types of datasets available for scholars are increasingly large and heterogeneous, and many such datasets are automatically captured through technology. For example, users often leave traces on digital platforms and such traces make it possible to conduct "digital experiments." Companies such as Uber, Google, and Amazon conduct hundreds of such experiments every day under the label of A/B testing. Uber, for instance, tests the effects of surge pricing on demand in a city and the effects of gamification on driver behaviors.[6] Data used for such analyses may, for example, include the traces left behind by a completed service incidence, such as a car repair, as captured by both the actions and interactions of the customer and the service provider, each of which are recorded by the digital platform that mediates these interactions.

Such datasets are strictly behavioral:[7] They indicate what people have done in the context of a digital or physical process and where such actions leave traces on a digital platform of some kind to be stored and used later. These may, for example, be traces of posting, payments, bidding, social connections, viewing, editing, downloading, or linking to various types of user-generated content. While such data often include text, there is usually little in the data as such that would indicate what higher level constructs the data may manifest, such as, for example, specific emotional or cognitive states, structured social practices (e.g., norms, rules) or cultural values (e.g., reactions to gender or race).

The research that emanates from big datasets tends to focus on relationships between variables encoded in digital traces. Not surprisingly, many (although not all) such datasets are collected by private firms and made accessible to researchers. Such datasets are especially amenable to solving a firm's *tactical* problems. Tactical means that the research problems being studied are confined to largely local issues (i.e., issues that are narrow and contextually specific) and are mostly concerned with the immediate, empirical connections between the variables included in the dataset (as measured either directly or through computational transformation).[8] Consequently, the research emanating from using such datasets tends to have lower degrees of abstraction and tends to not reach beyond the immediate meaning of variables captured as digital traces. Therefore, we propose the following conjecture:

Conjecture #1: BDR will exhibit a tendency to address tactical problems

In BDR the data are not researcher invoked, and therefore the specific research problem addressed is ultimately dictated by the characteristics of the dataset, which is used either inductively or abductively. Hence, while there may be a broad practical need motivating the study and setting the parameters of the dataset, the specificity of the research questions mainly emerges from the data. Rather than utilizing a deductive approach where the research question informs the theoretical framing, which then guides data collection and analyses, BDR may invert this order. We note that inductive research, including interpretive approaches, has a long tradition in IS research. While BDR often starts with discovering empirical regularities in preexisting datasets, it generally does not present itself as inductive, exploratory, theory-building research.

Thus, as the availability of varied datasets increases, the research problems within the field are likely to grow increasingly diverse and concomitantly incommensurate with other studies and their findings. Different researchers may concentrate on different, highly localized datasets and may not strive (or will not be incentivized to strive) for the requisite amounts of abstraction necessary to connect deeply to the literature; therefore, the value for an individual researcher (as expressed within an individual journal article) to contribute back to a particular theoretical discourse becomes more limited (Lindberg, 2020). Rather, the value and raison d'être of such research lies in explicating relationships between patterns of data so that insights with regard to empirical relationships can be directly applied to fine tune a particular platform or application, which, in turn, helps to maximize sales, customer retention, user influence across a social network, or some other desired performance metric.

It is therefore unlikely that, following such an approach, we will reach significant levels of consistency in constructs, operational measures, or in the way problems are formulated in the long run, for the simple reason that it is not necessary to do so when trying to fine tune an idiosyncratic sociotechnical system from which the data are collected. If the goal is not to contribute to an overall theoretical and abstract

[6] Hal Varian, chief economist at Google, indicated in his talk, "Beyond Big Data," at the NABE meeting in September 2013, that Google runs about 10,000 experiments a year and that each time you access google.com you are participating in dozens of such experiments.

[7] Textual, image, and video data reflecting human behavior can sometimes be used to infer internal cognitive and affective states (e.g., positive and negative emotional valence) that go beyond behavior. This, however, is not as common within BDR as it is, for example, in survey-based SEM research.

[8] For example, Cavusoglu et al. (2016) analyzed over two million Facebook messages to study changes in user communications after the 2009 privacy policy change.

discourse regarding an important research topic and a set of research questions shared by a community, then it is clear that BDR may weaken the communal goal of building a theoretically and empirically *cumulative tradition*. Therefore, we propose the following conjecture:

Conjecture #2: BDR will result in widespread local diversity in research to the detriment of a cumulative tradition.

The relationship between IS research and the IT artifact has been debated for almost twenty years (Orlikowski & Iacono, 2001). While some have argued that the artifact should have an explicit presence as a construct in our research models (Benbasat, & Zmud, 2003), others have been comfortable with greater latitude regarding how the IT artifact is treated in our research discourse (Lyytinen & King, 2004; Robey, 2003). There is also strong evidence that many IS studies lack an explicit focus on the IT artifact, and that this "IT artifact deficit" has not improved over the years (see, e.g., Grover & Lyytinen, 2015; Orlikowski & Iacono, 2001). Despite this diversity of views regarding the core matter of the field, we anticipate that an increasing number of BDR papers will deal mostly with digitally collected data on phenomena increasingly distant from the present core. Indeed, some research practices indicate that the lure of big, *digital* data trumps relevance to IS. As a result, the IT artifact is increasingly unlikely to be included in the focal phenomenon being studied. The only connection of such studies to the IT artifact may be the fact that the data were collected using digital means or generated as part of behaviors unfolding on a digital platform. We do not, however, think that this focus is necessarily a distinguishing characteristic of present IS research and do not believe that it will be in the future, because scholars in marketing, human resources, and operations management increasingly rely on similar types of datasets (e.g., Chae, Bruno, & Feinberg, 2019).

Furthermore, some views of the IT artifact, such as the "ensemble view" (Orlikowski & Iacono, 2001), rely on inferring constructs and aspects of the IT artifact that may not be directly observable in terms of digital traces. The ensemble view focuses on understanding how people frame, appropriate, and enact technologies and related activities, i.e., how an IT artifact is constituted by an ensemble of social and technical forces. Such socially constructed aspects, however, are difficult to directly observe in people's behaviors. To capture IT artifacts as ensembles, researchers need to talk to users of said IT artifacts and ask them to account for particular courses of action to understand how they see the world, themselves, and the technologies with which they work in concert. If such data are not available or require tedious *primary* data collection, then the access to presumably more objective forms of behavioral data, such as digital traces, may take precedence in IS research.

We expect that the usage of digital trace data to drive research will encourage researchers to adopt a "nominal" view of the IT artifact (Orlikowski & Iacono, 2001). Such a view of the IT artifact considers IT "in name only." It is likely that such research will focus on the raw action of individuals using technologies, as represented in digital traces of use, rather than identifying *why* they act as they do, thus limiting our understanding of how IT and its development and/or usage is embedded in sociotechnical contexts. To accomplish the latter, researchers have to create and engage with unobservable constructs that need to be inferred interpretively using interviews or observational data including user accounts, or at least through access to texts in which people talk about the use of IT (e.g., Lindberg et al., 2016). This would help researchers assess what is going on "inside people's minds" as they enact technologies. Therefore, we propose the following conjecture:

Conjecture #3: BDR will exhibit a bias toward a nominal treatment of the IT artifact

BDR tends to investigate problems focused on local concerns and how to validly establish related claims of cause and effect. Such problems do not need broad theoretical support (although theory is often sprinkled throughout the manuscript in a cursory manner), because the cause-effect chains examined are often motivated by common sense rooted in the local setting of a particular study. The lack of more generalized causal explanations is not paramount since the primary driver of the research is the practical motivation of addressing a tactical problem. It remains to be seen whether the results of such explorations will later be abstracted using meta-analysis or other inductive theory development approaches to arrive at higher echelons of theoretical knowledge if scholars put theory at the backend of the research process, as advocated by inductive theory building. Or, as we conjecture next, without strong expectations of theoretical generalization, as long as data and analytics are deemed sufficient to justify a contribution (Leidner, 2020), it is quite likely that theory will receive little attention. Hence, we propose the following conjecture:

Conjecture #4: BDR will exhibit a bias toward cursory treatment of theory.

We have argued (through Conjectures #1 and #4) that BDR papers prioritize the practical value of the predictive accuracy of a model over generalizable theory. We suspect, moreover, that many BDR studies emphasize contributions that emanate from using large and unique datasets and new extraction techniques, as well as the associated use of advanced algorithms and sophisticated statistical tools. This is naturally positive as it advances the types of evidence and methods available to IS scholars. In many cases, however, the lure of BDR will also have the consequence that the IT artifact will be either absent or nominally represented. Instead, a loose, indirect association with the IT artifact will still be present to make the study qualify as IS research because sophisticated software and hardware were used to produce the data and/or analyze it. As a result, many BDR papers seek to make strong claims with regard to their contribution based on the novelty of data (referring to the size or uniqueness of the dataset) or analytic technique used (e.g., application of a novel, rarely used before machine learning technique). Therefore, we propose the following conjecture:

Conjecture #5: BDR will have a tendency to focus on data and methods, as opposed to theoretical knowledge of the IT artifact associated with non-BDR papers.

4 Our Conjectures: The Current State of BDR in IS Journals

To determine whether our conjectures have merit we contrasted BDR studies identified in a representative sample of IS papers with a randomized sample of non-BDR studies published in the same outlets. Toward this end, we chose a subset of three journals from the AIS Senior Scholars' Basket of Eight—*MIS Quarterly*, *Information Systems Research*, and *Journal of Management Information Systems*—which are popular outlets for quantitative research. From all papers published in these journals during 2016, 2017, and 2018 (through the third out of four issues for each journal in 2018), we excluded pure theory and method papers, editorials, and research commentaries to arrive at an initial set of 392 papers that we coded for measures helpful for assessing the veracity of our conjectures (see Table 1 below). Appendix A explains our data collection and coding process in greater detail. Appendix B presents a complete list of the BDR papers and Appendix C offers a complete list of the non-BDR papers. The results of the analyses are shown in Table 2. Based on this evidence, we return to our conjectures and assess their veracity.

Conjecture #1 is already happening: Tactical research is prevalent in BDR studies. This conjecture asserts that BDR will lead to more "tactical" as opposed to abstract research and is supported by the analysis. Among the 41 identified BDR articles, 31 (76%) were tactical, while in the matched non-BDR sample, only 13 (32%) articles were tactical. The difference is statistically significant ($\chi^2 = 14.17$, $df = 1$, $p < 0.01$).

Conjectures #2 and #5 are also happening: Less theory and more data feature prominently in BDR studies. We observe that the "mid-pages" in BDR papers are 2.53 pages *longer* than those in non-BDR papers, indicating a stronger focus on describing the methodology and techniques used in BDR papers. The difference is statistically significant ($t = 3.19$, $df = 79.98$, $p < 0.01$). In fact, this 2.53-page difference accounts for the shorter total length of non-BDR papers versus BDR papers (19.10 pages on average for non-BDR papers vs. 21.71 pages on average for BDR papers); this difference is also statistically significant ($t = 2.22$, $df = 76.86$, $p = 0.03$). The finding that the method/findings pages are growing longer in relation to the theoretical setup and discussion of the study results in the backend of BDR papers provides support for Conjecture #2 (see Table 2, "Back pages" row). BDR papers increasingly focus on a limited and local study topic and less on tying it to a chosen theoretical framing, thereby diminishing contributions to a cumulative research tradition. This also lends support to Conjecture #5, which suggests that the data/technique is emphasized in BDR papers, and that generalizability using ties with theoretical claims in the literature is less of a concern.

Conjecture #3 is somewhat less prevalent: The treatment of the IT artifact shows a less pronounced pattern. The nominal treatment of the IT artifact is more common in BDR. In the BDR sample, 23 of the 41 (56%) articles share a nominal view, whereas in the non-BDR sample the incidence of nominal views of the IT artifact are lower (16/41 = 39%). The direction of the difference suggests that understanding and explaining the role of the IT artifact is less of a focus in BDR. This is corroborated by the assessment that the majority of papers (56%) in the non-BDR sample deal with design, management, and impact of the IT artifact, while the minority of the papers in the BDR sample (39%) do so. This difference, however, is not statistically significant ($\chi^2 = 1.76$, $df = 1$, $p = 0.18$).

5 Our Conjectures: Why We Should Be Concerned?

Based on the conjectures we have proposed and tested, we now extrapolate several potentially negative consequences of current BDR practices on the IS field, assuming a broad agreement on the goals of knowledge production in the IS field pertaining to the IT artifact and a cumulative tradition. We may expect a significant change to unfold in what type of knowledge and in what ways knowledge is produced and deemed valid if such research starts to dominate our journals. These consequences assume the practicing of a pure and simplistic form of BDR, as outlined above, that eschews the need for abstraction and instead uses data and technique as the primary drivers of research justification and contribution.

Table 1. Coding of BDR and Non-BDR Studies

Aspect	Measures	Provides Evidence for:
Phenomenon	What is the phenomenon being studied? Is it tactical (narrow), i.e., a practical problem? Or is it an abstract (theoretical) issue that transcends the specific context?	Conjecture #1
Link to literature	What is the relative investment in terms of pages in describing method/findings vs. prestudy setup and discussion of findings?	Conjecture #2
IT Artifact	How is the IT artifact treated in the study? Is it treated nominally (as opposed to a proxy, computational, tool, or ensemble view)? Or, does the study deal with the design, management and/or impact of the IT artifact?	Conjecture #3
Theory	Is there testing of hypotheses? Is there merely passive application of received theory (instantiation)? Or is there theoretical development (modification or extension of theory)?	Conjecture #4
Technique	What is the proportion of the number of pages allocated to method and results?	Conjecture #5

Table 2. Results Contrasting BDR with Non-BDR Studies

	No. of papers or pages		**% of papers**		**Statistical test**
Variable	**BDR**	**Non-BDR**	**BDR**	**Non-BDR**	
Tactical / Abstract	31/10	13/28	76/24	32/68	$\chi^2 = 14.17$, $df = 1$, $p < 0.01$
Hypothesis testing / No hypothesis testing	17/24	22/19	41/59	54/46	$\chi^2 = 0.78$, $df = 1$, $p = 0.38$
Front pages (introduction and theory sections) (average)	6.56	5.98	-	-	$t = 0.84$, $df = 78.44$, $p = 0.41$
Mid-pages (method & findings sections) (average)	12.85	10.32	-	-	$t = 3.19$, $df = 79.98$, $p < 0.01$
Back pages (discussion and conclusions) (average)	2.29	2.80	-	-	$t = -1.70$, $df = 77.98$, $p = 0.09$
Total pages (average)	21.71	19.10	-	-	$t = 2.22$, $df = 76.86$, $p = 0.03$
No theory development / Theory development	34/7	23/18	83/17	56/44	$\chi^2 = 5.75$, $df = 1$, $p = 0.02$
Nominal IT artifact / Other representation of artifact	23/18	16/25	56/44	39/61	$\chi^2 = 1.76$, $df = 1$, $p = 0.18$
IT artifact (design, management, impact) yes/no	16/25	23/18	39/61	56/44	$\chi^2 = 1.76$, $df = 1$, $p = 0.18$

Three interrelated consequences that may take hold within IS research and spawn a downward spiral are the dilution of the IS field's identity, greater fragmentation of the field, and greater corporate governance of research output. After discussing these three possible consequences, we consider how to best leverage BDR for the purposes of IS research and how BDR can positively complement more traditional IS research to benefit the field.

5.1 Dilution of the IS Field's Identity

Debates about the identity of the IS field have flourished since the publication of the first articles that emphasized the primacy of the IT artifact (Benbasat et al., 2003; Orlikowski & Iacono, 2001). At the time of the publication of these articles, e-commerce and the pervasiveness of the internet were still in their infancy. These new phenomena, however, posed the question of how much the field cares about theorizing and explaining the effects of the changing nature of IT. The rise of new IT artifacts prompted critical reflections on the possible changing identity of the IS field. Today we are living in a world in which almost every commercial and organizational transaction/process/communication is mediated by some type of information system. This is also reflected in the near universal presence of IT artifacts that are embedded in processes involving all subfields in a typical business school. Similarly, sociology, architecture, engineering disciplines, medical and health sciences, education, and law are all increasingly engaging with technology and IS in multifaceted ways.

The question then becomes, whether other business disciplines will potentially take a piece of "our cake," our unique contribution. A recent article by Sarker et al. (2019) argues that what holds the IS field together is a "sociotechnical axis of cohesion" centered on the interplay between the social world and the technical (i.e., information-related) world. This suggests that there is an opportunity to establish ourselves as a reference field in relation to other social science fields, specifically with regard to the intersection between the social and the technical (i.e., through crafting sophisticated, evidence-based concepts and theories of IT artifacts as ensembles). If we can develop sophisticated theories that deal with this intersection at an appropriate level of abstraction, we can generate a smorgasbord of theories that other fields, such as marketing, finance, economics, accounting, and management (as well sociology and education), can draw upon when they need to address specific issues related to this intersection.

BDR in IS, if it continues along the trajectory it has followed to date, may dilute the distinctive nature of our research output. If the primacy of data and technique becomes increasingly dominant (Conjecture #5), and because most BDR draws from digital repositories, almost *any* digital data can be claimed to bear relevance to IS research based on its loose relationship with the IT artifact (Conjecture #3). The increasing prominence of various apps within a growing variety of areas (e.g., entertainment, health, finance, news) as well as digital repositories collated by for-profit corporations or various governmental institutions, leads to a plethora of novel data that can be examined. These data can often be connected, at least rhetorically speaking, to IS research, because of their digital nature and origin. However, such studies are more likely to have stronger implications for and be more relevant to other fields in which related problems originate.[9] The dilution will be compounded if the knowledge outputs are not complementary to each other (Conjecture #2) and remain locally idiosyncratic to the data and/or technique (Conjecture #1).

5.2 Increased Fragmentation of the IS Field

A natural consequence of diluting the IS field's identity is a greater fragmentation of the field. While the term "fragmented adhocracy" has long been used to depict the sociological structure of knowledge and community in the IS field (Grover, London, & Craig, 2016; Hirschheim, Klein, & Lyytinen, 1996; Kling, Banville, & Landry, 1989), BDR holds the potential to accentuate this characteristic. If we accept the sort of data that commercial platforms provide to us today, we will become increasingly limited in terms of what claims we can make and what types of IT artifact conceptualizations will be possible. Often, data used in BDR studies are strictly positivistic in that they focus on traces of user or machine behaviors. In fact, the differences between machines and humans are likely to be diluted (see, e.g., Zuboff, 2015).

At times, however, such data also include text, thus enabling analysis of human expression and communication as it occurs in linguistic forms. Digital trace data as text may therefore invite a more wide-ranging set of analyses in terms of the cognitive, intentional, and emotional states of actors. Such data may even be combined with hermeneutic or idiographic analysis methods conducted manually by human analysts (Lindberg, 2020). This would help us to grapple with changes in the social structures, practices, cultural values, norms, and mores that underlie IT-

[9] For example, we observed, in the IS dissertations presented at the doctoral consortium at ICIS 2017 in Seoul, topics such as measuring sleep patterns or mining "smart city" data on traffic light synchronization and traffic patterns. Such studies made claims to belong to the IS field because they used digital trace data but are, substantively speaking, only peripherally relevant to the design, management, and/or implications of digital artifacts for human enterprise.

based interactions. However, much of the BDR that we examined did not include an analysis of such a text component though they were available, but rather focused on tallying and analyzing digitally recorded "actions," such as buying, selling, logging on, and participation. For example, Bapna et al. (2018) found that customers who convert from the free to the paid version of the music service Last.fm listen to more songs, create more playlists, make more forum posts, and gain more "friends" on the platform, as compared to those who continue to use the free version. These findings do not require abstracting from digital trace data but rather represent patterns directly observed within such data.

One of the crucial differences between BDR and non-BDR papers that emerged in our analysis is that BDR exhibits substantively higher degrees of tactical, and therefore locally focused, research (Conjecture #1). BDR does not start with higher-level theoretical constructs and then proceed to identify possible indicators of such constructs, neither does it attempt to abstract to such constructs from "found" trace data (Webb & Weick, 1979). Rather, BDR tends to identify a set of "raw" variables as being important in themselves and therefore rarely discusses their validity or potential for bias in variable selection or measurement. Because the raw variables are taken as constructs in themselves, their measures are assumed to involve a minimum of bias. For example, a common case is to look at the number of logins that a user has made on a platform and its connection to the contribution volume in the community hosted on the platform. The number of "logins" that a particular user has made is not treated as an indicator of another construct such as "use," rather it *is* the construct "logins." This is what we would call a tactical set of variables in the sense that the degree of abstraction from measure to construct is low.

On the upside, such research may be conducted with substantial rigor as indicated by the increased size of the method sections. However, such research also tends to be incremental and narrowly empirical in its contribution. The number, diversity, and sophistication of analytical tools available provide researchers with increasing degrees of freedom in the ways they structure and conduct their analysis (Conjecture #5). R-hacking (i.e., fishing for interesting relationships) and HARKing (i.e., hypothesizing after results are known) may now be easily embraced (sometimes unintentionally) by researchers. When analyzing rich datasets, this becomes an alluring option and many times some interesting results can be guaranteed given the size of the dataset. Such practices, however, raise serious questions regarding the overall reliability and stability of findings as the datasets are still local and the sampling methods used may be suspect. If BDR continues along its current trajectory, it may not only increase the granularity of fragmentation within the IS field but may also reduce the generalizability and value of findings overall.

These characteristics of focusing on method, i.e., diminished engagement with construct-level theorizing and a focus on tactical problems, make it more difficult for BDR to contribute to a cumulative tradition (Conjecture #2) that stretches across multiple empirical contexts, types of technologies, and varied forms of information processing. Generalizability of findings will become harder unless we establish our field's research criteria to include the expectation that produced knowledge needs to contribute to a cumulative tradition in which the aim is to understand behaviors and features of sociotechnical systems at a theoretical level. BDR, however, will likely underemphasize the importance of theory (Conjecture #4).

5.3 BDR and Corporate Control of Research Output

It seems as if BDR shares similarities with the sort of research that is now regularly conducted by corporations in their continued effort to better "tune" their own products and services for profit and market share (Conjecture #1). However, the types of datasets needed to accomplish this, and the research that is produced, mostly pertain to firms' current tactical problems (which often do not relate directly to corporate strategy or related, deeper issues with regard to a firm's structural arrangements or environmental pressures). Solving such tactical problems will have strong, immediate, and direct implications for target companies and their operations.

If academic research becomes barely distinguishable from the research conducted by corporations to improve their operations, how do we justify its value for other stakeholders including academia at large, the public, policy makers, and so on? In our minds, we cannot. Tactical research fails to draw upon or contribute to the theories developed by other scholars and can therefore be done in relative isolation (Conjecture #2). We also expect that corporations are able to conduct this type of research more effectively than the academic community given corporations' greater access to data and computing resources, as well as the technical talent possessed by companies such as Google, Facebook, Microsoft, and Amazon. What is lost is the type of research that seeks to contribute widely to a higher-level theory of how aspects of technology and information interact with individuals, groups, organizations, communities, and societies (Conjecture #4). Such research also locks us into the types of platforms (and their goals) that have already been developed by corporations. In effect, researchers are turned into "tuners" who work to maximize the efficiency of platforms that are developed by corporations. This makes it more difficult for the academic community to ask more fundamental questions, even critical ones, as regards the relationships between IT, organizing, and human enterprise. These questions indeed require higher levels of abstraction and alternative forms of theorizing.[10]

Furthermore, BDR requires the availability of multiple heterogeneous datasets and ample computing resources for continued success. Consequently, continued and expanded access to such resources can become a critical success factor and differentiator for BDR scholarship. It is likely to create a divide between data haves and have-nots, where access to data largely drives research success.

Today, we are fortunately still in a situation where most digital trace data provided by companies are publicly available. However, this situation could change quickly and the situation we anticipate above may become more likely. For example, Twitter currently provides access to its archive of tweets via multiple APIs with different degrees of access and associated costs. As data are becoming the defining resource of the twenty-first century, the fact that researchers increasingly rely on data owned by corporations represents a fundamental shift in power, holding the potential to redefine what is researchable and by whom. This trend may be associated with multiple, deleterious effects. First, corporations will have greater control over what type of research is possible. Through restricting what kinds of data are made available, corporations can actively influence and control the types of research that are being done, the types of questions being asked, and ensure that such research primarily benefits the interests of the corporation. Second, while access currently is quite open, some corporations, such as Twitter, have actively provided access only to select groups of scientists. Through actively restricting access for some scholars, while giving access to other scholars, corporations wield control over the kinds of research that are being done and who gets preference in the community.[11] Points 1 and 2 also grow more important when combined with the fact that BDR has a tendency to focus on tactical questions (Conjecture #1), which rely heavily on the formulation of variables and traces in the data themselves. This means that the type of data that are made available will have a stronger influence on the kind of research that will be conducted in the future. More theoretically oriented forms of research would be less vulnerable to who controls data resources, as such research seeks to abstract beyond the specifics of the data to a cohesive theoretical discourse and its attendant conceptualizations.

There is some evidence that the indirect hand of corporate interests is already playing a role in limiting what data are "allowable" in research. For example, the revised data provenance policy of *Management Science* states that they will not allow papers that scrape data from websites that "explicitly ban such a practice" and papers could be withdrawn if the "entity complains to INFORMS … and demonstrates … material harm" or if the data were "stolen or hacked," regardless of their availability in public spaces, unless they get permission from the company.[12] Would this not limit researchers from examining important questions regarding possible anticompetitive practices or biases of IT-based platform companies? Additionally, the access could also have implications for the replicability of results[13] which is more of a concern in BDR, as most datasets are local and particular. This will be more pronounced in situations where datasets become proprietary, such as increasingly is the case with, for example, social media.

6 Our Conjectures: Leveraging BDR for Academic Research

There has always been a tension between rich, localized qualitative research that uses small samples and quantitative research that uses large samples and relies on cross-sectional surveys or panel data. However, BDR introduces a tension of a different sort. Because of the volume, variety, and velocity of big data as well as the fact that it is collected primarily for commercial purposes, BDR will surely assist the corporate world (George et al., 2016). In other words, within corporate research, digital traces are becoming the primary data source. Within academic research, such data will also become an important data source, but it should not be applied uncritically for reasons discussed in concert with our conjectures.

[10] It seems as if theory has become a convenient scapegoat for what ails the field. For instance, recent articles by Hirschheim (2019) and Dennis (2019) take a negative stance on "conventional" theory. We argue that as a field we should not be too quick in dismissing the importance of theory. Arguably, theory is at least partially responsible for bringing the field to the place of respectability it occupies today. However, courteous debate on productive and unproductive forms of theorizing, particularly in an environment of increasing digitalization and big data, is very important and is the hallmark of a healthy field.

[11] See https://medium.com/on-archivy/twitters-developer-policies-for-researchers-archivists-and-librarians-63e9ba0433b2

[12] See https://www.informs.org/Blogs/ManSci-Blogs/From-the-Editor/From-the-Editor-January-2019

[13] We also note a countercase being made suggesting that BDR facilitates better replicability due to open access to common datasets. However, this position is contingent on accessibility. We suspect that corporations will be prepared to grant access to data to the extent that it helps their cause while meeting necessary privacy regulations, which obviously may be breached or abrogated in various ways.

We recognize that the magnitude and diversity of big data clearly present a tremendous opportunity for IS researchers. Big data can more accurately represent behaviors not biased by direct interactions with researchers that undeniably exist when primary, “small” datasets are collected. BDR can also help amass and operationalize a broader repertoire of variables and thereby facilitate examination of novel questions regarding emerging digital phenomena, or already established questions of importance, which can now be inquired into in ways that were not possible before.

We also recognize, however, that IS researchers should aim to create broad, generalizable knowledge that can be built on by others. The evidence provided in this editorial indicates that BDR may do that to a lesser degree than non-BDR. There are also constituents in the field that may suggest that abstract knowledge is not as important as increased accuracy of local predictions. We can subscribe to that view from a practice-based perspective. The goal of BDR for practice could, for example, be to accurately read a magnetic resonance imaging (MRI) output and analyze it to make predictions in a specific setting.[14] In contrast, as IS researchers, we should use such data to better understand the broader issues related to use of technology and information in a social context. To understand and address the latter, we need to ask “why” questions. Asking why predictions work creates a basis for improving predictive models as the context changes. We believe that the IS field has the talent to address both interesting “what” questions as well as the more abstract “why” questions, even if such questions are not asked or answered by the same set of researchers, or even within the same study.

Below, we offer some suggestions that may enable BDR to contribute toward more generalizable knowledge. If followed, we would not only leverage the vast power of big data, but we could also improve the field’s long-term welfare. We recognize that some in the field believe that big data, powerful algorithms, and vast computing capacity can “automate” number crunching and reveal novel empirical patterns that do not require formulation of a priori hypotheses (Agarwal & Dhar, 2014). While we do not subscribe to such brute empiricism as a preferred approach to BDR for reasons described earlier, we do appreciate that there is a strong constituency in the field that has invested and continues to invest in such BDR techniques, which help to address specific research questions, and do not see deductive theory-based approaches as highly relevant.

For those *not* theoretically inclined, we would suggest that some care can and should be taken in linking the results of BDR studies to a broader knowledge goal. We posit that this is vitally important for the future of the field. If the only output of a BDR project is a new, valuable prediction that is idiosyncratic to a specific dataset or a corporation (e.g., traffic light synchronization leads to less traffic congestion, high net worth customers use banking apps more frequently, open vs. sealed auctions have different effects on bid sizes, etc.), then the field will be stunted in its ability to produce novel academic knowledge. As noted previously, we will be competing with actors and powers with bigger guns and greater praxis smarts. Therefore, we encourage BDR scholars in our field to increasingly make connections to broader fields of generalized knowledge (Johnson et al., 2019) through raising the heights of concepts being used, and therefore also enabling them to connect to higher-level theoretical discourses. This will also enable BDR to link the topics and problems that are being studied to broader literatures. Below, we discuss how to best leverage BDR for the benefit of *academic IS research* by keeping in mind the importance of the problems investigated, the measurements used during analysis, and theory development.

First, we would be concerned if we heard a doctoral student linking his or her research questions to an available corporate dataset while suggesting that the research questions will stem solely from the data. It would be equally concerning for us to hear a student lock onto a theory that the literature has already examined exhaustively and argue that his or her research problem is based on how well the theory can be applied. Both data and theory are intellectual tools used to gain a deeper understanding of the mysteries and puzzles of the blooming, buzzing confusion that is our world. The research problem must therefore stem from such unresolved phenomena and have salience to the field. Hence, researchers need to constantly iterate between the data, the theoretical literature, and the phenomenon at hand. The phenomenon may yield insights into important new questions that can then be tempered and framed by the extant literature (which can help us see what we already know about this phenomenon or the general problem it is related to) and informed or validated by the data (which can inform us about what aspects of our questions we can truly examine with confidence). Such an iterative process can better alleviate concerns that the access to new digital trace data is the only issue of relevance in the IS field, while IT is treated as a nominal phenomenon. If we can distinguish between localized tactical problems and broader knowledge problems that have salience with the field, then we can better formulate research agendas that are grounded in the former yet let us contribute to the latter.

Second, big data is mostly collected in the form of low-level behavioral trace data (Hedman, Srinivasan, & Lindgren, 2013; Venturini & Latour, 2010). Analysis of such large datasets will typically identify relationships

[14] https://healthcareinamerica.us/how-machine-learning-and-ai-could-improve-mris-1a0f4d50816c

that have relatively low-level correlations. Furthermore, in many cases, while the data as such are likely to be devoid of contextual cues, interpreting such data could allow for a deeper and more refined understanding of the phenomenon at hand. Such data can yield insights into appropriate ways of combining low-level data to create valid proxy representations of higher-level constructs, which can then invite broader inferences from the lower-level relationships detected. Classical research models involve the use of theoretical models that are expected to be tested by using operationalizations of constructs (i.e., indicator variables) that must meet established validity standards. In the BDR case, the reverse becomes important. For datasets relevant to the research problem, scholars need to identify constructs and their proxy representations (Howison et al., 2011). The key idea is to connect low-level variables already existing within given datasets, and their combinations, to broader constructs to demonstrate fair levels of construct validity. Therefore, while BDR may carefully examine available data and practical questions, we would argue that thought should be given to the corresponding knowledge questions, and some investment made in iterating between such questions and the variables available in the dataset to foster greater knowledge impact.

Third, during the BDR process it is useful to ask, "How can I abstract my tactical, analytical problem to a more general or archetypal research problem (Rai, 2017) at either the frontend of the analysis, the backend, or both?" This calls for linking the problem at hand to the pertinent literature. If a theoretical base is found to be relevant (and there is nearly always relevant theoretical literature to draw upon because theory, by definition, is abstract, and therefore applicable to a wide range of specific contexts, at least to some degree), it will help frame the problem and hypotheses as novel. Or, it might allow the building of a rationale or theoretical logic for anticipated relationships. Navigating between the specific and the general problem helps to keep the research grounded and allows us to arrive at a knowledge product valuable for both academia and the corporate world.

Even if the findings of BDR remain specific to the dataset at hand, greater investment should be made in abstracting to general problems at the backend so that other researchers can build theories based on the results or seek to further examine the identified relationships in a broader context. When this course of action is followed, BDR empiricists can search for collaboration with other fellow researchers that are more theoretically inclined. In this way, by linking tactical problems to more general, theoretical problems, big data researchers can act as effective drivers of theoretical development through providing a steady supply of rigorously derived evidence. Such a symbiotic relationship between the discovery mode of BDR and more classic modes of theory development and testing will be beneficial for the IS field and will leverage specialized and valuable BDR skills more effectively.

6.1 Complementarity

There are complementary approaches that can take advantage of the strengths of both BDR and traditional theory-based approaches (Berente et al. 2019; Lindberg 2020; Østerlund et al. 2020). A key aspect of such approaches is the importance of injecting human creativity into the BDR process so that theory development is increasingly informed by novel big data patterns. This begs the question: How do we create synergies between large-scale number crunching that helps reveal local and rich patterns and classic theorizing at higher levels of abstraction that is informed by the versatility and rich tradition of sociotechnical theory?

At a basic level, traditional hypothetic-deductive research can benefit from BDR in terms of testing theory-driven hypotheses in novel ways. However, the ability to match a big dataset with theory-driven constructs remains challenging. Therefore, there may exist an opportunity for big data to help *triangulate* existing measures. While perceptual measures can be developed and validated through psychometrically sound techniques, they can and need to be complemented by using proxy variables offered by big datasets. This effectively amounts to a new form of mixed methods research utilizing multiple and different quantitative techniques.

It is also possible to use sophisticated algorithms to identify novel patterns in data that can then offer *initial structural frames* for deeper theory-building research that uses qualitative data or small data techniques. Such an approach can suggest patterns (e.g., clusters) of objective behaviors represented in big data. Such frames would be impossible to generate using traditional inductive approaches (such as grounded theory building) but can be deployed effectively as the basis for identifying constructs and examining emergent relationships in depth using theory-based approaches. For example, Vaast et al. (2017) utilized cluster analysis to identify groups of Twitter users, whose tweets could then be examined qualitatively to understand how various groups differed in their social media communications. Similar to structural framing, the researcher could also start with large-scale patterns and then use *abduction* to draw out reasonable inferences for the presence of such relationships. These inferences could then form the basis for new hypothesis development to be conducted in concert with some appropriate, underlying logic or theory.

Table 3. Summary of Approaches Using BDR

Approach	**Technique-driven**	**Technique-dominated**	**Symbiotic**
Focus of Research Team	Data extraction, mining, machine learning, and analytics	Mainly data and analytical skills but also some theoretical consideration	Balance of theoretical skills and data/analytical skills
Abstraction	Little abstraction	Some consideration of abstraction	Substantial consideration of abstraction
Key Idea	• Identify a topic relevant to practice and compile accessible big data on that phenomenon • Use machine learning and other big data techniques on accessible big data to identify novel patterns	• *Problems Matter:* Consideration of the practical knowledge problem • *Measurement Matters:* Consideration of proxy representations from the dataset that connect to higher-level constructs • *Theory Matters:* Abstract from the local to the general problem to foster links with the literature	• *Triangulate:* Use BDR to triangulate variables and results from traditional deductive research (mixed methods) • *Structural Frame:* Use BDR to identify patterns that can frame theory building and testing using small samples • *Abduction:* Derive reasonable inferences from BDR results that can generate new hypotheses for testing
Consequences	High probability of Conjectures #1-5 being true with focus on deriving accurate predictions that can be refined and applied in a practical context.	Moderate probability of Conjectures #1-5 being true, but offers both practical value and knowledge contributions at a moderate level of abstraction	Low probability of Conjectures #1-5 being true, and therefore offering potential to leverage both BDR (big data) and traditional theory-based (small data) advantages

These hypotheses could subsequently be validated using small data techniques. Abduction allows scholarly imagination to address the "why" question behind observed patterns. It also complements the strengths of computationally intensive methods with scholarly creativity and thus facilitates elusive indigenous theorizing in the field (Lindberg, 2020). Table 3 provides a summary of several, possible BDR approaches. "Technique-driven" research is epitomized by the type of BDR that we have described throughout this paper, while "technique-dominated" research represents a view of BDR that applies some of the considerations with regard to abstraction that we have suggested throughout the discussion section. Finally, "symbiotic" research indicates how BDR may be combined with theory-based approaches and thus leverage the benefits of triangulation, structural frames, and abduction described above.

We recommend that, as a field, we strive to avoid purely technique-driven papers (Table 3, Column 1), unless the method itself offers some broader contribution to the field. For those researchers that are less theoretically inclined, a technique-dominated approach (Column 2) would be more suitable, as it leverages the advantages of big data and analytics but makes a conscious investment in linking to the literature and the general problem, ex ante or ex post. For those that are more theoretically inclined, the symbiotic approach (Column 3) allows BDR to do what it does best—discovering interesting frames or relationships—in synergy with small data approaches (theorizing and testing "why" questions). This would require care by important stakeholders in the field, senior researchers, dissertation chairs, and editors, to not endorse practices in which data access takes precedence over important research questions, BDR and non-BDR are considered isolated from each other, theory is completely abdicated or considered unnecessary, and accurate prediction replaces the need for explanation. Institutionalization of such practices, in our opinion, not only fails to leverage the advantage of big data but can also be detrimental to the field's long-term health.

7 Closing Remarks

As BDR gains a stronger foothold in our outlets and research community, new and critical issues to be debated are emerging. Given current projections that digital trace data will double every two years,[15] issues related to how our research field can best leverage such a trend are becoming as salient, if not more so, than the debate around the IT artifact was in the early 2000s. Our argument largely revolves around the fulcrum of theory, i.e., the idea that abstractions are critical for distinguishing ourselves from corporate research as well as to build a cumulative tradition.[16] If our field fails to continue to engage in powerful abstractions, our ability to think as a community about the rich phenomena surrounding IS and the multifarious relationships between the social and the technical will grow increasingly circumscribed. We, as IS researchers, will risk becoming a field of particulars with few connections across such particulars. This would lead to fragmentation within our field, which in turn would result in us wielding less institutional power. There are manifold opportunities to leverage BDR in our field and to engage with new abstractions that call for genuine and multifaceted researcher skills and propensities. We hope that this editorial helps frame the necessity of harnessing the strengths of both big and small data, investigating tactical and abstract problems, developing both theoretical and practical implications, and finding a common and stronger way forward for advancing the field of IS.

[15] https://www.idc.com/research/viewtoc.jsp?containerId=US43171317

[16] Consider Jorge Luis Borges' (1998) short story "Funes, His Memory" about a man that could not forget anything and therefore remembered everything: "Funes not only remembered every leaf of every tree in every patch of forest, but every time he had perceived or imagined that leaf … I suspect, nevertheless, that he was not very good at thinking. To think is to ignore (or forget) differences" (pp. 136-137). Forgetting differences, obviously, is abstraction. What Borges illuminates is that in order to be able to think, one must learn to make proper abstractions; otherwise, one's mind remains filled with particulars, rather than being complemented with generalizations, discriminations, patterns, regularities, connections, linkages, and causalities.

References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association of Information Systems*, *17*(2), 1-32.

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, *25*(3), 443-448.

Bapna, R., Ramaprasad, J., & Umyarov, A. (2018). Monetizing freemium communities: does paying for premium increase social engagement? *MIS Quarterly*, *42*(3), 719-735.

Bell, G., Hey, T., & Szalay, A. (2009). Computer science: Beyond the data deluge. *Science*, *323*(5919), 1297-1298.

Benbasat, I., & Zmud, R. W. (2003). The identity crisis within the IS discipline: Defining and communicating the discipline's core properties. *MIS Quarterly*, *27*(2), 183-194.

Berente, N., Seidel, S., & Safadi, H. (2019). Research commentary: Data-driven computationally-intensive theory development. *Information Systems Research*, *30*(1), 50-64.

Borges, J. L. (1998). Funes, his memory. In *Collected fictions* (pp. 131-137). Penguin.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. SAGE.

Cavusoglu, H., Phan, T. Q., Cavusoglu, H., & Airoldi, E. M. (2016). Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. *Information Systems Research*, *27*(4), 848-879.

Chae, I., Bruno, H. A., & Feinberg, F. M. (2019). Wearout or weariness? Measuring potential negative consequences of online ad volume and placement on website visits. *Journal of Marketing Research*, *56*(1), 57-75.

Davis, G. F. (2015). Editorial essay: What is organizational research for? *Administrative Science Quarterly*, *60*(2), 179-188.

Dennis, A. R. (2019). An unhealthy obsession with theory. *Journal of the Association for Information Systems*, *20*(9), 1404-1409.

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64-73.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, *59*(5), 1493-1507.

Grover, V., London, J., & Craig, K. (2016). A historical observation of the intellectual and institutional structures of the field. *Communications of the Association for Information Systems*, *38*(1), 432-476.

Grover, V., & Lyytinen, K. (2015). New state of play in information systems research: The push to the edges. *MIS Quarterly*, *39*(2), 271-296.

Hedman, J., Srinivasan, N., & Lindgren, R. (2013). Digital traces of information systems: Sociomateriality made researchable. *Proceedings of the International Conference on Information Systems*.

Hirschheim, R., Klein, H., & Lyytinen, K. (1996). Exploring the intellectual structures of information systems development: A social action theoretic analysis. *Accounting, Management and Information Technologies*, *6*(1-2), 1-64.

Hirschheim, Rudy. (2019). Against theory: With apologies to Feyerabend. *Journal of the Association for Information Systems*, *20*(9), 1338-1355.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, *12*(12), 767-797.

Johnson, S. L., Gray, P., & Sarker, S. (2019). Revisiting IS research practice in the era of big data. *Information and Organization*, *29*(1), 41-56.

Kling, R., Banville, C., & Landry, M. (1989). Can the field of MIS be disciplined? *Communications of the ACM*, *32*(1), 48-60.

Lazer, D., Pentland, A. A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., … Gutmann, M. (2009). Life in the network: The coming age of computational social science. *Science*, *323*(5915), 721-723.

Leidner, D. (2020). What's in a contribution? *Journal of the Association for Information Systems*, *21*(1), 238-245.

Lindberg, A. (2020). Developing theory through integrating human and machine pattern recognition. *Journal of the Association for Information Systems*, *21*(1), 90-116.

Lindberg, A., Berente, N., Gaskin, J., & Lyytinen, K. (2016). Coordinating interdependencies in online communities: A study of an open source software project. *Information Systems Research*, *27*(4), 751-772.

Lyytinen, K., & King, J. L. (2004). Nothing at the center?: academic legitimacy in the information systems field. *Journal of the Association for Information Systems*, *5*(6), 220-246.

Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems*, *19*(12), 1253-1273.

Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately Seeking the "IT" in IT research—A call to theorizing the IT artifact. *Information Systems Research*, *12*(2), 121-134.

Østerlund, C., Crowston, K., & Jackson, C. (n.d.). Building an Apparatus: Refractive, Reflective & Diffractive Readings of Trace Data. *Journal of the Association for Information Systems*, 1-43.

Rai, A. (2017). Editor's comments: Seeing the forest for the trees. *MIS Quarterly*, *41*(4), iii-vii.

Robey, D. (2003). Identity, legitimacy and the dominant research paradigm: An alternative prescription for the IS Discipline—A response to Benbasat and Zmud's Call for returning to the IT artifact. *Journal of the Association for Information Systems*, *4*(7), 352-359.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical "axis of cohesion" for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, 43(3), 695-719.

Simsek, Z., Vaara, E., Paruchuri, S., Nadkarni, S., & Shaw, J. D. (2019). New ways of seeing big data. *Academy of Management Journal*, *62*(4), 971-978.

Sudhir, K. (2016). Editorial: The exploration-exploitation tradeoff and efficiency in knowledge production. *Marketing Science*, *35*(1), 1-9.

Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social media affordances for connective action: An examination of microblogging use during the Gulf of Mexico oil spill. *MIS Quarterly*, *41*(4), 1179-1205.

Venturini, T., & Latour, B. (2010). The social fabric: Digital traces and quali-quantitative methods. *Proceedings of Future En Seine*.

Webb, E., & Weick, K. (1979). Unobtrusive measures in organizational theory: A reminder. *Administrative Science Quarterly*, *24*(4), 650-659.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75-89.

Appendix A: Our Coding Process

To identify BDR papers we coded all 392 papers to identify papers conforming to the following criteria: (1) they use digital trace data from users interacting with online platforms; (2) their abstract argues for the uniqueness of data and/or methods, thus suggesting that these aspects form a major component of the paper's contribution; and (3) their sample size is over 500. 41 papers were identified as BDR using these criteria. Such research often calls for the use of nontraditional estimation techniques, e.g., big data analysis approaches such as machine learning methods. Appendix B has a complete list of these articles.

Then, we randomly sampled a matching set of 41 non-BDR empirical papers to provide a comparison sample. This sample is defined as papers that do *not* meet the criteria for BDR stated above, and therefore reflect traditional social science and economics research that uses small samples of primary data that is mostly researcher-collected, either using a survey or an experimental approach. The analytical techniques used are often less computationally intensive and are more likely to, for example, consist of classic regression-based methods or qualitative studies. Appendix C has a complete list of these non-BDR articles. Each article in the sample was read in detail and coded. Several steps were taken to ensure the validity and reliability of the coding (Boyatzis, 1998). First, four coders experienced in IS research (the authors) each coded a subsample of five BDR and five non-BDR papers independently, and then collectively discussed the discrepancies encountered in the coding. Then, another subset of BDR and non-BDR papers were independently coded by each author and new minor discrepancies were discussed and reconciled. After the coding principles were determined and no issues in how to code specific articles remained, the final dataset was coded by a single coder. The final dataset was compared to a subsample of coding carried out by an independent coder. Interrater reliability was calculated with a value of 0.87, demonstrating high reliability of the coding.

We coded for the following qualitative aspects: tactical vs. abstract research, theory building (Grover & Lyytinen, 2015), type of IT artifact (Orlikowski & Iacono, 2001), and whether the paper deals with the design, management, and/or implications of IT artifacts (Benbasat & Zmud, 2003). We also counted the number of pages allocated to theoretical setup, method and findings, and discussion.

Tactical vs. Abstract Research

Tactical research uses raw variables in the data as stand-ins for concepts, and therefore has low "conceptual height" (Lindberg, 2020). Such research tends to deal with local issues, specific to a particular context, and therefore has low degrees of generalizability. For example, such research may relate logins on a platform to sales using digital traces of such actions, without trying to conceptualize the traces further. Often, the raw indicators (variables) from digital trace data are identified directly as theoretical constructs.

Abstract research defines abstract concepts and then seeks to formulate indicators of such concepts. This means that the concepts usually find themselves at a higher level of abstraction, compared to the empirical measurements of these constructs. Such research allows for generalization beyond specific measurements and contexts. Abstract research seeks to increase the "conceptual height" (Lindberg, 2020) of the claims so that that the indicators are treated as manifestations of constructs that are implicated in a higher-level theoretical discourse. In assessing the article, we ask if the studied phenomenon as represented by the constructs is a tactical problem or whether it is informed by a higher-level theoretical issue or principle that transcends the local context. This coding helped to provide evidence for Conjecture #1.

Relative Attention Paid to Theoretical Setup, Method, and Findings, and Discussion

Our goal is to observe links between a paper and the extant literature to assess whether the work draws from and builds upon past research. To do this, we assume that the number of pages used in a paper to theoretically build its argument offers a proxy representation of how much is invested in (i.e., the degree of focus or care placed on) such knowledge. Therefore, we count how many pages there are between: (1) the abstract and the beginning of the method section; (2) the method section and the results section; and (3) the discussion section and the end of the paper, not counting appendices or references. The number of pages in each of these sections indicates the scholars' relative investment in the method and results in comparison to the study's theoretical setup and discussion of findings and theoretical contributions. Integration with the extant literature calls for more investment in the frontend and backend of a paper, while emphasis on data and/or a technique and its novelty require more investment in the middle of the paper, namely, the method and findings sections. These proxy measures provide evidence for Conjecture #2 (increasing local diversity at the expense of a cumulative tradition) and Conjecture #5 (contribution of data/technique) and serve as a rough indication of Conjecture #4 (the importance of theory), as theoretical development usually calls for greater investment in the front end of a paper.

Theory Building

We measure whether or not hypothesis testing takes place, as hypotheses are often, but not always, based on articulated theory/logic. We also drew on Grover and Lyytinen (2015) to code for the type of theory building engaged in. We identified whether each article engaged in instantiation (i.e., no theory development) of theory, i.e., the borrowing of other theories to be instantiated wholesale in a new context, as opposed to modification or extension of theory. Modification of theory suggests the alteration of preexisting constructs and/or relationships between constructs. Extension of theory suggests the development of novel constructs and therefore also novel theory. This coding helped provide evidence for Conjecture #4.

Type of IT Artifact

We used Orlikowski and Iacono's (2001) taxonomy of the IT artifact: tool, proxy, ensemble, computational, and nominal. The tool view sees IT artifacts as tools being used for the intended purposes of their designers. The proxy view captures IT artifacts through surrogate measures, such as capital investments in IT. The ensemble view suggests that IT artifacts are bound together using both social and material resources. The computational view focuses on the use of algorithms and analytical models. Finally, the nominal view treats IT artifacts as effectively absent. This categorization helps to provide evidence for Conjecture #3.

Design, Management, and/or Implications of IT Artifacts

We drew on Benbasat and Zmud (2003) to classify whether each paper deals with IT artifacts at all, through identifying whether the research focuses on either the design, management, and/or implications of IT artifacts. This classification helped provide additional evidence for Conjecture #3.

Appendix B: BDR Articles

Abbas, A., Zhou, Y., Deng, S., & Zhang, P. (2018). Text analytics to support sense-making in social media: A language-action perspective, *MIS Quarterly*, *42*(2), 427-464.

Adamopoulos, P., Ghose, A., & Todri, V. (2018). The impact of user personality traits on word of mouth: Text-mining social media platforms. *Information Systems Research, 29*(3), 612-640.

Agarwal, A., & Mukhopadhyay, T. (2016). The Impact of Competing Ads on Click Performance in Sponsored Search. *Information Systems Research, 27*(3), 538-557.

Aguiar, L., Claussen, J., & Peukert, C. (2018). Catch me if you can: Effectiveness and consequences of online copyright enforcement. *Information Systems Research*, *29*(3), 656-678.

Arazy, O., Daxenberger, J., Lifshitz-Assaf, H., Nov, O., & Gurevych, I. (2016). Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. *Information Systems Research, 27*(4), 792-812.

Bapna, R., Ramaprasad, J., & Umyarov, A. (2018). Monetizing freemium communities: Does paying for premium increase social engagement. *MIS Quarterly, 42*(3), 719-735.

Benjamin, V., Zhang, B., Nunamaker, J. F., & Chen, H. (2016). Examining hacker participation length in cybercriminal internet-relay-chat communities. *Journal of Management Information Systems*, *33*(2), 482-510.

Bhatnagar, A., Sen, A., & Sinha, A. P. (2017). Providing a window of opportunity for converting eStore visitors. *Information Systems Research*, *28*(1), 22-32.

Burtch, G., Ghose, A., & Wattal, S. (2016). Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Information Systems Research*, *27*(3), 478-496.

Cavusoglu, Huseyin, Phan, T. Q., Cavusoglu, Hasan, & Airoldi, E. M. (2016). Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. *Information Systems Research, 27*(4), 848-879.

Fang, X., & Hu, P. J.-H. (2018). Top persuader prediction for social networks. *MIS Quarterly, 42*(1), 63-82.

Förderer, J., & Kude, T. (2018). Does platform owner's entry crowd out innovation? Evidence from Google Photos. *Information Systems Research*, *29*(2), 444-460.

Ge, R., Feng, J., Gu, B., & Zhang, P. (2017). Predicting and deterring default with social media information in peer-to-peer lending. *Journal of Management Information Systems*, *34*(2), 401-424.

Ghiassi, M., Zimbra, D., & Lee, S. (2016). Targeted Twitter sentiment analysis for brands using supervised feature engineering and the dynamic architecture for artificial neural networks. *Journal of Management Information Systems*, *33*(4), 1034-1058.

Ghose, A., Todri-Adamopoulos, V., Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Toward a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly, 40*(4), 889-910.

Gong, J., Abhishek, V., & Li, B. (2018). Examining the impact of keyword ambiguity on search advertising performance: A topic model approach. *MIS Quarterly, 42*(3), 805-829.

Gong, J., Hong, Y., & Zentner, A. (2018). Role of monetary incentives in the digital and physical inter-border labor flows. *Journal of Management Information Systems, 35*(3), 866-899.

Gunarathne, P., Rui, H., & Seidmann, A. (2017). Whose and what social media complaints have happier resolutions? Evidence from Twitter. *Journal of Management Information Systems*, *34*(2), 314-340.

Gunarathne, P., Rui, H., & Seidmann, A. (2018). When social media delivers customer service: differential customer treatment in the airline industry. *MIS Quarterly*, *42*(2), 489-520.

Han, S. P., Park, S., & Oh, W. (2016). Mobile app analytics: A multiple discrete-continuous choice framework. *MIS Quarterly, 40*(4), 983-1008.

Ho, Y. C. C., Wu, J., & Tan, Y. (2017). Disconfirmation effect on online rating behavior: A structural model. *Information Systems Research, 28*(3), 626-642.

Hoang, A.-P., & Kauffman, R. J. (2018). Content sampling, household informedness, and the consumption of digital information goods. *Journal of Management Information Systems, 35*(2), 575-609.

Hong, Y., & Pavlou, P. A. (2017). On buyer selection of service providers in online outsourcing platforms for it services. *Information Systems Research*, *28*(3), 547-562.

Hong, Y., Wang, C., & Pavlou, P. A. (2016). Comparing open and sealed bid auctions: Evidence from online labor markets. *Information Systems Research*, *27*(1), 49-69.

Huang, J., Boh, W. F., & Goh, K. H. (2017). A temporal study of the effects of online opinions: Information sources matter. *Journal of Management Information Systems*, *34*(4), 1169-1202.

Kanat, I., Hong, Y., & Raghu, T. S. (2018). Surviving in global online labor markets for IT services: A geo-economic analysis. *Information Systems Research*, *29*(4), 893-909.

Kane, G. C., & Ransbotham, S. (2016). Content and collaboration: An affiliation network approach to information quality in online peer production communities. *Information Systems Research*, *27*(2), 424-439.

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced customer analytics: Strategic value through integration of relationship-oriented big data. *Journal of Management Information Systems*, *35*(2), 540-574.

Kumar, N., Venugopal, D., Qiu, L., & Kumar, S. (2018). Detecting review manipulation on online platforms with hierarchical supervised learning. *Journal of Management Information Systems*, *35*(1), 350-380.

Kwon, H. E., Oh, W., & Kim, T. (2017). Platform structures, homing preferences, and homophilous propensities in online social networks. *Journal of Management Information Systems*, *34*(3), 768-802.

Lappas, T., Sabnis, G., & Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research*, *27*(4), 940-961.

Lee, G. M., Qiu, L., & Whinston, A. B. (2016). A friend like me: Modeling network formation in a location-based social network. *Journal of Management Information Systems*, *33*(4), 1008-1033.

Luo, X., Zhang, J. (Jennifer), Gu, B., & Phang, C. (2017). Expert blogs and consumer perceptions of competing brands. *MIS Quarterly*, *41*(2), 371-395.

Van Osch, W., & Steinfield, C. W. (2018). Strategic visibility in enterprise social media: Implications for network formation and boundary spanning. *Journal of Management Information Systems*, *35*(2), 647-682.

Qiu, L., Shi, Z., & Whinston, A. B. (2018). Learning from your friends' check-ins: An empirical study of location-based social networks. *Information Systems Research*, *29*(4), 1044-1061.

Samtani, S., Chinn, R., Chen, H., & Nunamaker, J. F. (2017). Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence. *Journal of Management Information Systems*, *34*(4), 1023-1053.

Shi, D., Guan, J., Zurada, J., & Manikas, A. (2017). A data-mining approach to identification of risk factors in safety management systems. *Journal of Management Information Systems*, *34*(4), 1054-1081.

Thies, F., Wessel, M., & Benlian, A. (2016). Effects of social interaction dynamics on platforms. *Journal of Management Information Systems*, *33*(3), 843-873.

Wang, Q., Li, B., & Singh, P. V. (2018). Copycats vs. original mobile apps: A machine learning copycat-detection method and empirical analysis. *Information Systems Research*, *29*(2), 273-291.

Xu, J. J., & Chau, M. (2018). Cheap talk? The impact of lender-borrower communication on peer-to-peer lending outcomes. *Journal of Management Information Systems*, *35*(1), 53-85.

Zhou, S., Qiao, Z., Du, Q., Wang, G. A., Fan, W., & Yan, X. (2018). Measuring customer agility from online reviews using big data text analytics. *Journal of Management Information Systems*, *35*(2), 510-539.

Appendix C: Sampled Non-BDR Articles

Alam, S. L., & Campbell, J. (2017). Temporal motivations of volunteers to participate in cultural crowdsourcing work. *Information Systems Research*, *28*(4), 744-759.

Anderson, E. G., Chandrasekaran, A., Davis-Blake, A., & Parker, G. G. (2018). Managing distributed product development projects: integration strategies for time-zone and language barriers. *Information Systems Research*, *29*(1), 42-69.

Ayabakan, S., Bardhan, I. R., & Zheng, Z. (Eric). (2017). A data envelopment analysis approach to estimate it-enabled production capability. *MIS Quarterly*, *41*(1), 189-205.

Bauman, K., & Tuzhilin, A. (2018). Recommending learning materials to students by identifying their knowledge gaps. *MIS Quarterly* (1247(1), 313-332.

Breuker, D., Matzner, M., Delfmann, P., & Becker, J. (2016). Comprehensible predictive models for business processes. *MIS Quarterly*, *27*(4), 72-88.

Burtch, G., Hong, Y., & Liu, D. (2018). The role of provision points in online crowdfunding. *Journal of Management Information Systems*, *35*(1), 117-144.

Dawande, M., & Janakiraman, G. (2017). Not just a fad : Optimal sequencing in mobile in-app advertising. *Information Systems Research*, *28*(3), 511-528.

Deng, S., Huang, Z. (James), Sinha, A. P., & Zhao, H. (2018). The interaction between microblog sentiment and stock returns: An empirical examination. *MIS Quarterly*, *42*(3), 895-918.

Furneaux, B., & Wade, M. (2017). Impediments to information systems replacement: A calculus of discontinuance. *Journal of Management Information Systems*, *34*(3), 902-932.

Greenstein, S., & Zhu, F. (2016). Open content, Linus' law, and neutral point of view. *Information Systems Research*, *27*(3), 618-635.

Hann, I. H., Koh, B., & Niculescu, M. F. (2016). The double-edged sword of backward compatibility: The adoption of multigenerational platforms in the presence of intergenerational services. *Information Systems Research*, *27*(1), 112-130.

Hao, H., Padman, R., Sun, B., & Telang, R. (2018). Quantifying the impact of social influence on the information technology implementation process by physicians: A hierarchical bayesian learning approach. *Information Systems Research*, *29*(1), 25-41.

Hardin, A., Looney, C. A., & Moody, G. D. (2017). Assessing the credibility of decisional guidance delivered by information systems. *Journal of Management Information Systems*, *34*(4), 1143-1168.

Hashim, M. J., Kannan, K. N., & Maximiano, S. (2017). Information feedback, targeting, and coordination: An experimental study. *Information Systems Research*, *28*(2), 289-308.

Heimbach, I., & Hinz, O. (2018). The impact of sharing mechanism design on content sharing in online social networks. *Information Systems Research*, *29*(3), 592-611.

Ho, S. Y., & Lim, K. H. (2018). Nudging moods to induce unplanned purchases in imperfect mobile personalization contexts. *MIS Quarterly*, *42*(3), 757-778.

Huang, P., & Zhang, Z. (2016). Participation in open knowledge communities and job-hopping: Evidence from enterprise software. *MIS Quarterly*, *40*(3), 785-806.

Huber, T. L., & Kude, T. (2017). Governance practices in platform ecosystems: Navigating tensions between cocreated value and governance costs. *Information Systems Research*, *28*(3), 563-584.

Kang, K., Hahn, J., & De, P. (2017). Learning Effects of Domain, Technology, and Customer Knowledge in Information Systems Development: An Empirical Study. *Information Systems Research*, *28*(4), 797-811.

Kwon, H. E., So, H., Han, S. P., & Oh, W. (2016). Excessive dependence on mobile social apps: a rational addiction perspective. *Information Systems Research*, *27*(4), 919-939.

Lankton, N. K., Harrison McKnight, D., Wright, R. T., & Thatcher, J. B. (2016). Using expectation disconfirmation theory and polynomial modeling to understand trust in technology. *Information Systems Research*, *27*(1), 197-213.

Li, S., Cheng, H. K., Duan, Y., & Yang, Y. C. (2017). A study of enterprise software licensing models. *Journal of Management Information Systems, 34*(1), 177-205.

Li, X. (2016). Could deal promotion improve merchants' online reputations? The moderating role of prior reviews. *Journal of Management Information Systems*, *33*(1), 171-201.

Li, X. (2018). Impact of average rating on social media endorsement: The moderating role of rating dispersion and discount threshold. *Information Systems Research*, *29*(3), 592-611.

Liang, N. (Peter), Biros, D. P., & Luse, A. (2016). An empirical validation of malicious insider characteristics. *Journal of Management Information Systems*, *33*(2), 361-392.

Lindberg, A., Berente, N., Gaskin, J., & Lyytinen, K. (2016). Coordinating interdependencies in online communities: A study of an open source software project. *Information Systems Research, 27*(4), 751-772.

Ludwig, S., van Laer, T., de Ruyter, K., & Friedman, M. (2016). Untangling a web of lies: Exploring automated detection of deception in computer-mediated communication. *Journal of Management Information Systems*, *33*(2), 511-541.

Mai, F., Shan, Z., Bai, Q., Wang, X. (Shane), & Chiang, R. H. L. (2018). How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems*, *35*(1), 19-52.

Mengxiang, L., Chuan-Hoo, T., Kwok-Kee, W., & Kanliang, W. (2017). Sequentiality of product review information provision: An information foraging perspective. *MIS Quarterly, 41*(3), 867-892.

Mithas, S., Whitaker, J., & Tafti, A. (2017). Information technology, revenues, and profits: Exploring the role of foreign and domestic operations. *Information Systems Research*, *28*(2), 430-444.

Niculescu, M. F., Wu, D. J., & Xu, L. (2018). Strategic intellectual property sharing: competition on an open technology platform under network effects. *Information Systems Research*, *29*(2), 498-519.

Pang, M. S. (2017). Politics and information technology investments in the U.S. federal government in 2003-2016. *Information Systems Research*, *28*(1), 33-45.

Roma, P., Gal-Or, E., & Chen, R. R. (2018). Reward-based crowdfunding campaigns: informational value and access to venture capital. *Information Systems Research 29*(3), 679-697.

Sambhara, C., Rai, A., Keil, M., & Kasi, V. (2017). Risks and controls in internet-enabled reverse auctions: Perspectives from buyers and suppliers. *Journal of Management Information Systems, 34*(4), 1113-1142.

Serrano, C., & Karahanna, E. (2016). The compensatory interaction between user capabilities and technology capabilities in influencing task performance: An empirical assessment in telemedicine consultations. *MIS Quarterly*, *40*(3), 597-621.

Shaikh, M., Vaast, E., & Shaikh, M. (2016). Folding and unfolding: Balancing openness and transparency in open source communities. *Information Systems Research*, *27*(4), 813-833.

Xue, L., Ray, G., & Zhao, X. (2017). Managerial incentives and IT strategic posture. *Information Systems Research*, *28*(1), 180-198.

Yan, J. I. E. K., Leidner, D. E., & Benbya, H. (2018). Differential innovativeness outcomes of user and employee participation in an online user innovation community. *Journal of Management Information Systems*, *35*(3), 900-933.

Ye, H. (Jonathan), & Kankanhalli, A. (2018). User service innovation on mobile phone platforms: Investigating impacts of lead userness, toolkit support, and design autonomy. *MIS Quarterly*, *42*(1), 165-187.

Zhao, L., Detlor, B., & Connelly, C. E. (2016). Sharing knowledge in social Q&A sites: The unintended consequences of extrinsic motivation. *Journal of Management Information Systems*, *33*(1), 70-100.

Zhou, W., & Duan, W. (2016). Do professional reviews affect online user choices through user reviews? An empirical study. *Journal of Management Information Systems, 33*(1), 202-228.

About the Authors

Varun Grover is the David D. Glass Endowed Chair and Distinguished Professor of IS at the Walton College of Business, University of Arkansas. He has published extensively in the information systems field, with over 400 publications, 250 of which are in major refereed journals. Over ten recent articles have ranked him among the top four researchers globally based on number of publications in the top IS journals, as well as citation impact. Dr. Grover has an h-index of 91 and over 37,000 citations on Google Scholar. Thompson Reuters recognized him as one of 100 Highly Cited Scholars Globally in business disciplines. He is currently senior editor for *MISQ Executive*, editor of the *Journal of the Association for Information Systems* section on Path Breaking Research, and has served as senior editor for *MIS Quarterly* (2 terms), the *Journal of the Association for Information Systems* (2 terms), and *Database*. Dr. Grover's current work focuses on the impacts of digitalization on individuals and organizations. He is the recipient of numerous awards from USC, Clemson, Academy of Management, AIS, DSI, the OR Society, Anbar, PriceWaterhouse, among others, for his research and teaching, and is a Fellow of the Association for Information Systems.

Aron Lindberg (aron.lindberg@stevens.edu) is an assistant professor of Information Systems at the School of Business, Stevens Institute of Technology. He received his PhD at Case Western Reserve University (2015), and primarily studies complex, distributed digital innovation processes, often using a combination of qualitative and computational methods. His research has been published or is forthcoming in major journals such as *MIS Quarterly*, *Information Systems Research*, and *Journal of the Association for Information Systems.*

Izak Benbasat (PhD University of Minnesota, 1974; Doctorat Honoris Causa, Université de Montréal, 2009) is a Fellow of the Royal Society of Canada and Sauder Distinguished Professor of Information Systems Emeritus at the Sauder School of Business, University of British Columbia, Canada. He received the LEO Award for Lifetime Exceptional Achievements in Information Systems from the Association for Information Systems, and was conferred the title of Distinguished Fellow by the INFORMS Information Systems Society. He has served as the editor-in-chief of *Information Systems Researc*h, editor of the IS&DSS Department of *Management Science*, and a senior editor of the *MIS Quarterly*. He was a Marvin Bower Fellow at the Harvard Business School, Shaw Visiting Professor at the National University of Singapore, and Otto Monsted Guest Professor at the Copenhagen Business School.

Kalle Lyytinen (PhD, Computer Science, University of Jyväskylä; Dr. h.c. Umeå University, Copenhagen Business School, Lappeenranta University of Technology) is Distinguished University Professor and Iris S. Wolstein Professor of Management Design at Case Western Reserve University and a distinguished visiting professor at Aalto University, Finland. He is among the top five IS scholars in terms of his h-index (89); he has the highest network centrality among IS scholars. He is the LEO Award recipient (2013), AIS Fellow (2004), and the former chair of IFIP WG 8.2: Information Systems and Organizations. He has published over 400 refereed articles and edited or written over 30 books or special issues. He conducts research on digital innovation concerning its nature, dynamics and organization, design work, requirements in large systems, and emergence digital infrastructures.