EDITOR'S COMMENTS

Computationally Intensive Theory Construction: A Primer for Authors and Reviewers

By: Shaila Miranda, Senior Editor
Nicholas Berente, Senior Editor
Stefan Seidel, Associate Editor
Hani Safadi, Associate Editor
Andrew Burton-Jones, Editor-in-Chief

As outlined in our March 2022 editorial, we are using our editorials in 2022 to address a key question in the mind of every author: What are editors looking for when they assess a paper like mine? While we could offer this advice for every type of paper submitted to MISQ, we have identified several genres for which we believe feedback will be particularly helpful, and we are dedicating one editorial to each one.[1] This one is dedicated to computationally intensive theory construction—*using computationally intensive approaches in the service of building new theory*.

The editorial is not offered to provide the "one true view" of the topic, but simply to provide helpful advice from editors who conduct this type of work and handle these types of papers. If your view of how best to write, review, or edit these papers differs from ours, that does not mean that your approach would not be welcome at *MIS Quarterly*. The computationally intensive theory construction genre is diverse and evolving and, as our prior EIC Arun Rai (2018) stressed, genres can also be combined. For all these reasons, we expect and encourage further dialog on how best to conduct and review such work. With these caveats, let's get into the topic.

Computationally Intensive Theory Construction

Information systems (IS) is seeing growth in novel, computationally intensive approaches to research (Agarwal & Dhar, 2014; Berente et al., 2019; Gaskin et al., 2014; Grover et al., 2020; Levina & Vaast, 2016; Lindberg, 2020; Maass et al., 2018; Miranda et al., in press; Müller et al., 2016; Pentland et al., 2020; Rai, 2016). Computationally intensive theory construction promises to revitalize scholarly knowledge through the application of novel methods and data sources. It is inherently phenomenon driven, a trajectory of research that is attracting widespread support for its relevance (Fisher et al., 2021; Von Krogh, 2018). But this trailblazing opportunity is fraught with uncertainty, as community norms, standards, and conventions for computationally intensive theory construction are only now emerging (Berente et al., 2019; Lindberg, 2020; Shrestha et al., 2021). In the meantime, how do we—as authors, reviewers, and editors—undertake and assess the quality of computationally intensive theory construction? What constitutes an adequate contribution meriting publication? When are data and methods rigorous enough to publish in *MIS Quarterly*? There are no simple answers to these questions, but our goal in this editorial is to offer some helpful guidance.

The purpose of computationally intensive theory construction is to produce theoretical insights from patterns identified using computational techniques, including—but not limited to—those that reveal categories, category memberships, associations, networks, sequences, and complex system dynamics by providing simulated data, pattern visualizations, or quantifications (Berente et al., 2019). Despite increasing understanding of the application of computational methods to theory construction, this genre poses unique challenges. Such work involves creative yet disciplined inquiry with rigorous yet reasonable thoroughness. Striking this balance involves judgment calls rooted in community norms, standards, and conventions.

In this editorial, we offer guidance for conducting and assessing computationally intensive theory construction, without any

---

[1] For each editorial, we are following a three-stage process. The editor-in-chief (Andrew) first selects editors with substantial expertise in the genre/method to run a masterclass for our editorial board. This group of editors is selected to include scholars with different levels of experience on the board to allow for differences in opinion and perspective. In this case, Shaila Miranda is a second-term SE, Nick Berente is a first-term SE (with three prior AE terms), Stefan Seidel is a second-term AE, and Hani Safadi is a first-term AE. Based on the learning and feedback from that session, the same group then runs an online seminar for authors. We then incorporate the learning from these two sessions into the editorial.

ambition to prematurely forestall community debates surrounding this genre. We hope to reduce uncertainty for researchers and review teams by distilling the foundations of a framework from our collective experiences as authors, reviewers, and editors of work in this genre. This framework is offered as a scaffold and should not be applied dogmatically, since doing so would infringe on the creativity imperative to this genre. As computational tools and methods continue to evolve, so too will community norms, standards, and conventions, and we champion mindful departures from this and prior frameworks.

Lineage and Overview

The framework advanced in this editorial builds on and clarifies the approach synthesized by Berente et al. (2019) from grounded theory and computational theory discovery traditions (Džeroski et al., 2007; Glaser & Strauss, 1967). Their approach involved the following iterative activities: sampling, synchronic analysis, diachronic analysis, and lexical framing. *Sampling* entails constructing a dataset. Through *synchronic and diachronic analysis* of sampled data, researchers surface patterns—discernable regularities—representing aspects of the data in particular ways. These patterns are intermediate products between data and theory and do not themselves constitute a theoretical contribution until they are abstracted in a way that situates them within and extends a theoretical discourse. This process of situating patterns in a theoretical discourse is *lexical framing* and is a key part of crafting a contribution.

Computationally intensive theory construction draws from three key genres—mixed methods research, computational social science, and grounded theory methodology (GTM), but is not reducible to any one of them. *Mixed methods research* involves combining qualitative and quantitative methods (Tashakkori et al., 1998). Computationally intensive theory construction may involve mixing computational methods or combining qualitative or quantitative approaches, but does not require it. *Computational social science* applies computing in the service of science (Lazer et al., 2009). Computationally intensive theory construction is a subclass of that genre, focusing on theory construction. GTM specifies manually intensive sampling and coding practices for developing theory, typically from qualitative data (Glaser & Strauss, 1967). Although the quantitative variant of GTM offers guidance for those using legacy computational approaches (Glaser, 2008), computationally intensive theory construction also draws on modern computational techniques that often take on a different character (Berente et al 2019). In short, if you are a researcher who uses mixed methods research, computational social science, or GTM, we hope you will see the potential for taking your work even further in this new genre.

As authors, editors, and reviewers in computationally intensive theory construction projects, we have encountered three recurring challenges: the complexity of lexical framing, the uncertain relationship between pattern and theory, and confusion about how much is enough—when have you got enough data, analysis, and contribution to stop? In the following sections, we address each of these issues. Figure 1 depicts our revision of the Berente et al. (2019) framework, which we will use to guide our discussion.

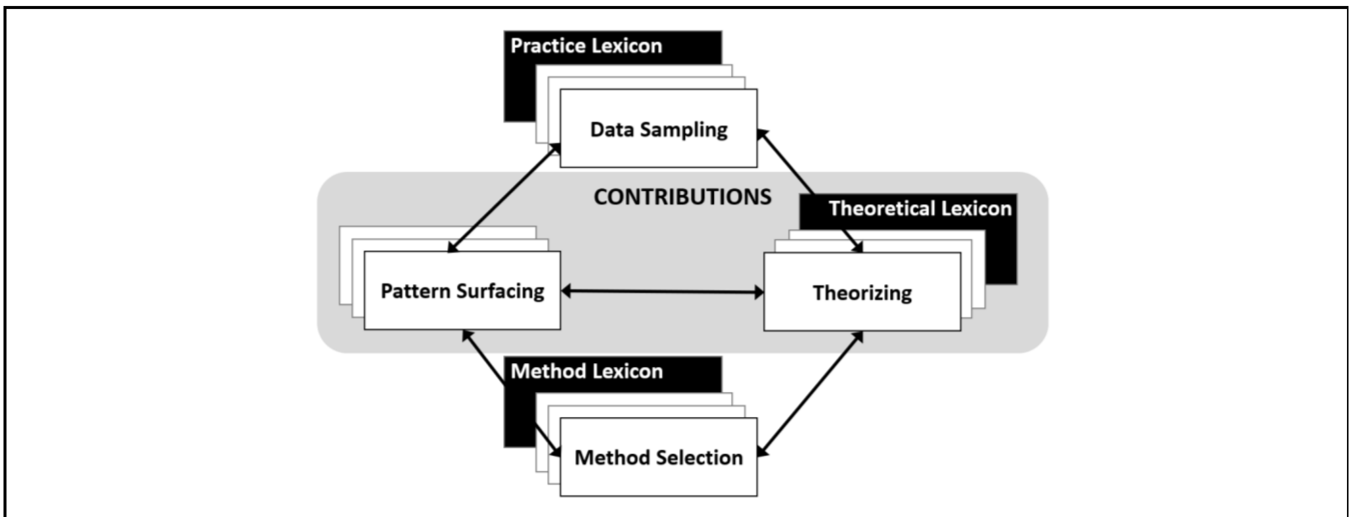


Figure 1. Overview of Computationally Intensive Theory Construction

Lexical Framing

A "lexicon" is the particular language that a community uses to represent its knowledge (Berente et al., 2019; Gaskin et al., 2014; Habermas, 1984). This language is theoretically laden and rife with assumptions and meanings. In contrast to high-paradigm fields that study few phenomena and use a few methods and established, integrated lexicons (Kuhn, 1962), IS researchers study diverse phenomena, draw upon disparate theories (from various reference fields), and use a range of methodologies. We therefore often draw on diverse lexicons and need to reconcile these lexicons. Our low-paradigm community thus faces unique challenges related to lexical framing. Lacking a single lexicon, we often struggle to situate our work in an ongoing discourse.

The researcher's knowledge of lexicons provides a scaffolding for identifying research problems, sampling data, selecting methods, and building theory. In any given computationally intensive theory construction endeavor, the researcher brings together and synthesizes three different sets of lexicons.

Practice lexicons are constitutive elements of the empirical phenomenon under study. In IS research, the phenomenon involves people and technologies as they act in the world. A key element of practice within social forms, such as communities, organizations, and fields, is the language used to communicate and coordinate action (Habermas, 1984). Each type of practice we study necessarily entails its own lexicon. This lexicon is theory laden in the sense that understandings of a community are embedded in their language (Habermas, 1984). For example, Pentland et al. (2021, p. 974) identified 300 *actions* in a hospital's electronic health record system with names such as "MR_REPORTS" and "ARVL_LST_DL_TIME" that convey specific meanings in that hospital community but mean little to those outside it. Communities thus have a highly specific language into which new members are socialized (Lave & Wenger, 1991) and which researchers must often learn.

Method lexicons accompany each methodological tradition. The methods that researchers use shape the inferences they make. Each computational technique derives from a different theoretical domain, bringing its own assumptions, meanings, and relations. Social network analysis, for example, has concepts such as nodes, edges, centrality, and density, rooted in communication and graph theories (Bavelas, 1950; Scott, 1988). Sequence analysis considers temporally ordered events, assessing the similarity of and distance among sequences (Abbott, 1995). Method lexicons, therefore, accompany methodological choices. Practice lexicons often need to be translated into a method lexicon. For example, Pentland et al. (2021, see footnote #4 on p. 970) referred to the system's "actions"—the practice lexicon—as "events," a term situated in the lexicon of the process mining method they used.

Theoretical lexicons are those of the theoretical discourses in which researchers situate themselves when crafting a theoretical contribution. Researchers problematize their research in relation to concepts and associations in the cumulative tradition of a scholarly community, drawing on and contributing to its lexicon. A theoretical contribution is a contribution *to* an ongoing scholarly conversation (Alvesson & Sandberg, 2011; Huff, 1999) and requires researchers to adopt the specialized lexicon of that discourse community. Their contribution may then extend or alter that lexicon—for instance, by adding novel concepts and relationships or by challenging established concepts and relationships. The theoretical lexicon of a scholarly community thus embodies its cumulative tradition. For example, Pentland et al. (2021) situate their research into hospital actions and events in relation to the theoretical lexicon they describe as "process theory" in organizational analysis.

A key task for researchers in this genre is creatively curating—assembling and aligning—these lexicons around the patterns that they identify. While lexical framing is important in any research genre, it is more complex in computationally intensive theory development because of the level of creativity required. We therefore call for those engaging with this genre to seek (and respect) *discursive flexibility*. In lieu of rigid adherence to a single discourse, discursive flexibility involves "strategically switching among multiple discourses which construe the necessary meanings and representations to achieve an objective" (Hunsberger & Alonso-Fradejas, 2016, p. 226). It fosters creative insight, especially when the lexicons are not immediately aligned (Zyphur, 2009). For example, by marrying events and event sequences from hospitals' practice lexicon with vertices and edges from the method lexicon of network analysis, Pentland et al. (2021) devised a new way of modeling process dynamics. Holding up the patterns from their analysis against the theoretical lexicon of process theory enabled them to show how their approach advanced knowledge of process variety, change, and iteration. Discursive flexibility does not mean using different lexicons indiscriminately. Rather, it requires that we use lexicons mindfully, remain cognizant of the compatibility of paradigms underlying those lexicons, and align them in ways that reconcile incompatibilities (Gioia & Pitre, 1990).

Discursive flexibility also entails "abandoning the inflexible pursuit of pre-conceived objectives" (Gibson, 2000, p. 368). Often, patterns surfaced across a study require researchers to realign their lexicons midstream. The mindful alignment of practice, method, and theoretical lexicons helps inform theory from patterns surfaced. We now turn to what we mean by "surfacing patterns."

Surfacing Patterns

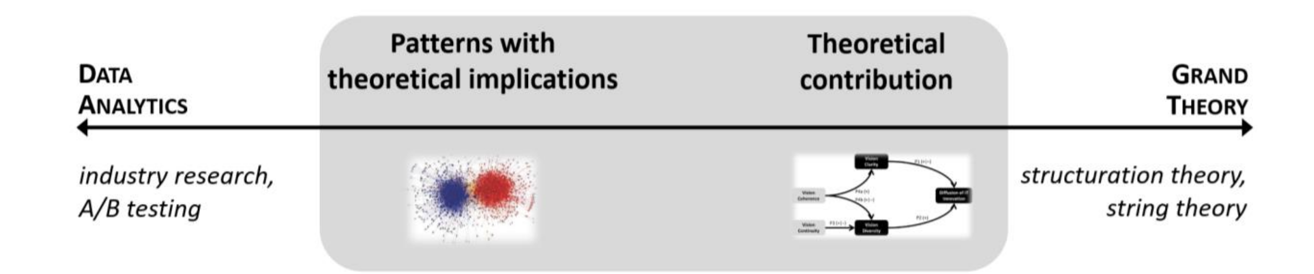
A *pattern* is a "regular and intelligible form or sequence discernible in certain actions or situations" (New Oxford American Dictionary, 2010). Patterns are generated from sensory information that is organized and classified according to regularities identified via a process called pattern recognition (Verhagen, 1975). Pattern recognition is an innate human faculty (Mattson, 2014), involving the perceiving and specification—or formulation—of a phenomenon (Sayre, 1965). For example, the photographer (Hani) who took the photograph in Figure 2 perceived the patterning of rocks and framed that pattern with his lens, thereby specifying it and giving it salience over other patterns in the botanical garden. Thus, pattern surfacing is a deliberate, selective, and generative human act. It is creative and requires interpretation in order to convert recognized patterns into a useful understanding or explanation (Maass et al., 2018). Because we interpret new information in the light of our prior knowledge (Strauss & Quinn, 1997), the framing of a pattern is shaped by the researcher's lexicons at hand. This lexical framing influences what patterns we notice, how we express them, and how we interpret them. Lexical framing also provides the language to relate the pattern to what we already know about a phenomenon (Gregor, 2006). Different researchers will therefore construct different patterns from the same/similar data based on their unique experiences, methodological choices, and theoretical lexicons. Mathematics and statistics provide many lexicons for quantifying and communicating patterns (Steen, 1988), but visual and graphical approaches also support recognizing and communicating patterns (Fox & Hendler, 2011).

Figure 2. A Pattern of Rocks in a Park (three big rocks followed by two smaller ones)

Pattern surfacing can be augmented by computational methods to identify regularities. Just as concepts sensitize researchers to the empirical world (Blumer, 1954), the methods selected shape the patterns researchers see and the inferences they draw from those patterns (Chalmers, 1999). For example, some researchers studying open source software projects focus on developers as nodes and developer collaborations as relations, or edges, in a network, thereby identifying network patterns (Maruping et al., 2019; Singh et al., 2011; Tang et al., 2020). Others emphasize sequences of activities among developers, taking the research in a much different direction (e.g., Bradley et al., 2020; Lindberg et al., 2016). Through synchronic analysis, researchers surface cross-sectional patterns, which they interpret as concepts and static relationships among concepts (Berente et al., 2019). Through diachronic analysis, researchers engage with temporal patterns, considering relationships like precedence and change (Berente et al., 2019). Such analyses may yield insights about causality through temporal precedence (Pearl, 2009) or about temporal mechanisms for change (Pentland et al., 2021). In addition to these chronotic views of time, diachronic analysis also can contribute to kairotic understandings, for example, of how a sense of "the right time" to adopt a technology unfolds (Czarniawska, 2004).

Computationally surfaced patterns are not themselves theories, but rather empirical generalizations (Handfield & Melnyk, 1998). Nor are new patterns that lead to new insight necessarily theoretical contributions. For example, Stokes et al. (2020) applied machine learning to a dataset of 61k molecules to surface patterns resulting in a new antibiotic—halicin. This was an important discovery but not itself a *theoretical* contribution. The scientists later sought to explain how halicin worked, which is necessary for attesting to the safety of the drug (Stokes et al. 2020). While the antibiotic itself was an important contribution, by situating the pattern within a cumulative tradition to provide the explanation for how it worked, the scientists theorized the pattern.

The search for patterns thus ranges across a continuum from the kind of data analytics tasks that are conducted in industry all the way to "grand theory" (see Figure 3). Rooted in specific phenomena, we do not anticipate that projects involving computationally intensive theory construction would generate "grand theory" (these may emerge from syntheses across multiple projects). Instead, we expect more modest, midrange theoretical contributions. Examples of such contributions are increasingly evident. For instance, Miranda et al. (2015) developed a theoretical model of innovation diffusion relative to the coherence, continuity, clarity, and diversity of the innovation's vision. Lindberg et al. (2016) constructed a theory of discursive open source coordination to deal with unresolved interdependencies among developers and software codebases. Between theoretical contributions and atheoretical data analytics, we believe that interesting patterns can have theoretical implications and often merit publishable contributions without necessarily forming specific theories. For example, in their study of political blogs, Adamic and Glance (2005) uncovered the phenomenon of polarization at a time when digital media were expected to have a connecting and unifying effect. This pattern subsequently had important implications for theories of political polarization. Consequently, we encourage our colleagues on editorial boards, reviewers, and authors to recognize the value not only of theoretical contributions, but also of such patterns with theoretical implications as potentially publishable contributions under certain circumstances (addressed later in this editorial).



Note: Patterns reproduced with permission from Adamic and Glance (2005, Figure 1) and from Miranda et al. (2015, Figure 6).

Figure 3. From Patterns to Theory

Table 1. Sample Patterns and the Method Families that Generate Them

Sample patterns	Method family: specific techniques	Examples
Latent categories and associations	*Dimensionalization and category surfacing*: Cluster analysis; multidimensional scaling; latent semantic analysis; relational class analysis; topic modeling; language and network embedding	Vaast et al. (2017); Miranda et al. (2015, in press); Yang et al. (in press)
Predictions of membership in known categories or of a numeric outcome	*Concept associations*: OLS regression; neural networks	Meyer et al. (2014); Miranda et al. (2015); Lindberg et al. (2016); Shrestha et al. (2021)
Process dynamics	*Process discovery*: Process mining; routine dynamics; sociotechnical sequence analysis	Recker et al. (2009); van der Aalst (2011); Gaskin et al (2014); Pentland et al. (2017); Pentland et al. (2021)
Patterns of association	*Social structure*: Social network analysis	Johnson et al. (2015); Zhang et al. (2016); Safadi et al. (2021)
Complex system dynamics	*Simulation*: Agent-based model; NK landscape; Generative adversarial model	Nan (2011); Brunswicker et al. (2019); Malgonde et al. (2020); Haki et al. (2020); Sturm et al (2021); Hahn and Lee (2022)

Researchers work with many different kinds of patterns. In Table 1, we identify a few archetypical patterns, the method families and techniques that produce them, and examples of their use. As the table shows, dimensionalization and category surfacing algorithms such as cluster analysis and topic modeling surface latent categories or continuous dimensions from associations. Legacy correlational techniques such as scatterplot matrices, partial correlations, ANCOVAs, and logistic regressions, as well as more recent machine learning techniques help researchers surface correlational insights. Process mining, sequence analysis, and narrative network techniques yield process insights. Social network analysis metrics and algorithms such as centrality, community detection, and motif analysis can yield insights not only into social structure, but also into discourse. Computational simulation illuminates complex system dynamics, especially when observational or experimental data are inaccessible.

Computationally intensive theory construction projects frequently combine patterns, possibly derived from different method families. For example, Hannigan et al. (in press) combined patterns from topic modeling of tweets with social network analysis to map "cultural holes" in Edmonton's artificial intelligence ecosystem. Researchers also combine patterns derived from qualitative and computational methods (e.g., Lindberg et al., 2016). Often method choices are only loosely coupled to the data. For example, relational class analysis (RCA) can be used to surface categories from coded texts (Miranda et al., 2015) as well as from attitudinal surveys (Goldberg, 2011). Similarly, topic modeling can be applied to texts as well as images (Bahmanyar et al., 2018).

But when does a pattern and its associated explanation or understanding merit publication? When is the data enough? When is the analysis through the methods applied sufficient? Next, we address these questions, offering some considerations for stopping rules.

Stopping Rules

Computationally intensive theory construction involves search processes through which a researcher surfaces patterns and situates those patterns in relation to relevant lexicons. Because such search processes are potentially interminable, they call for stopping rules—heuristics that ex ante specify when the search is to be terminated (Browne et al., 2007). Based on the challenges we have faced in this genre, we specify stopping rules for three questions: What constitutes a contribution? When is sampled data sufficient? How do we ensure methodological rigor? These stopping rules thus provide criteria that guide authors and reviewers to evaluate computationally intensive research in the three categories—theoretical contribution, data sampling, and methodological rigor. Importantly, these criteria are evaluated in relation to the relevant theoretical, methodological, and practice lexicons.

Stopping Rule 1: Do I have a Contribution Meriting Publication?

Researchers generate theoretical understanding, explanations, or hypotheses from some combination of experience, intuition, and logic through which they contribute to the scientific discourse (Mills, 1959; Peirce, 1903; Popper, 1959). Peirce (1903) described this as abduction, a process for generating scholarly insight through viewing observations in light of other, previously made observations. That is, the researchers infer their understanding of the phenomenon in terms of what seems to be most plausible, under consideration of previous observations. Viewing contributions as steps in a process of abduction paints a humble, fallibilist approach to scholarship, with full understanding that each contribution is imperfect. Each contribution is necessarily a simplification of reality, and these imperfect simplifications accumulate over time to form the tradition of a specific discourse.

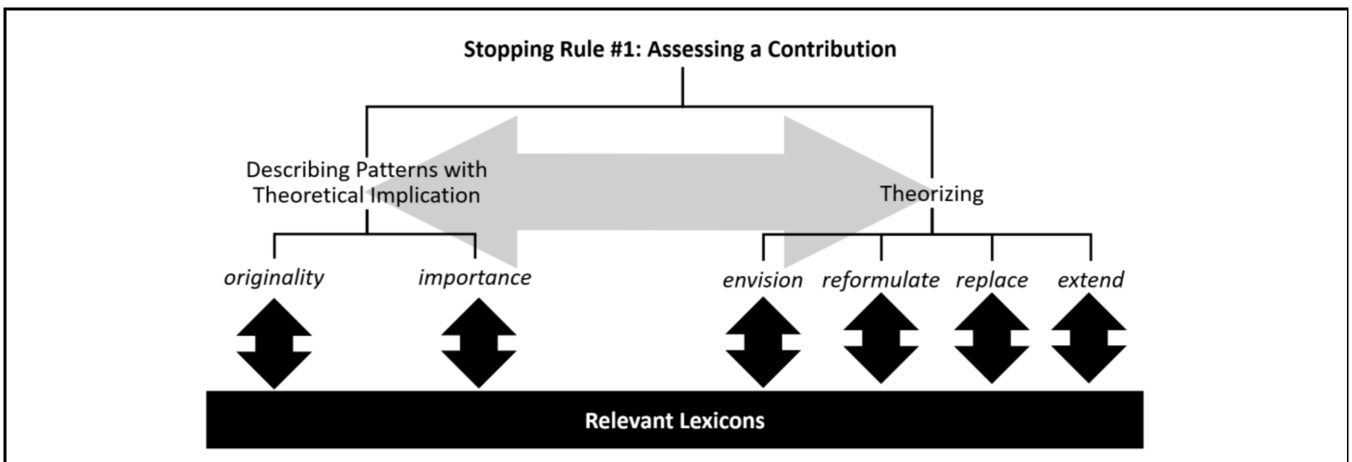


Figure 4. Have I Made a Contribution?

When describing scholarly contributions, Tiwana and Kim (2019) used the metaphor of bricks and an edifice. They argued that if each contribution (brick) stands independently, the field does not establish a strong cumulative tradition. Instead, studies should relate to and build on each other. Strong contributions are bricks in the edifice of a cumulative tradition. Although Berente et al. (2019) initially called for contributions to undertake both synchronic and diachronic analysis, a viable brick may involve one or the other alone. At some point in the investigation, researchers "freeze" the process (Weick, 1995b) and report on insights gleaned, recognizing that these insights constitute an imperfect and provisional contribution to an ongoing discourse. Consistent with this fallibilist, cumulative, bricks-and-edifice approach to scholarly contributions, there are two types of contributions from computationally intensive theory construction: theoretical contributions and patterns with theoretical implications (see Figure 4). For each, we can identify key criteria for establishing stopping rules in relation to the relevant lexicons on which the researcher draws.

Theorizing is the gold standard for scholarly contributions in this genre. Because different traditions can view theory differently, we define theory inclusively and simply as accounts of some phenomenon that provide explanation or understanding (Bacharach, 1989; Corley & Gioia, 2011; Schwandt, 1999; Van de Ven, 2007). There are several ways that research can theorize. First, the work may *envision* a new theory, providing a novel or an alternative account of the phenomenon. For example, Vaast et al. (2017) proposed a new theory of connective action through social media that attended to the particularities and roles of online interactions in social movements. Second, it may challenge, thereby *reformulate* or *replace*, an existing theory by revealing inadequacies of existing accounts. For example, Tiwana (2018) challenged one of the premises of modular systems theory by showing how a monolithic application architecture can lead to better performance. Third, it may enhance or *extend* an existing theory, for instance, by adding concepts and associations to enhance the theory's precision. For example, Lindberg et al. (2016) developed a theory of discursive coordination to account for how developers coordinate residual dependencies unaddressed by previously identified coordination mechanisms.

Describing patterns with theoretical implications can be strong contributions too. For instance, you may realize that a pattern you identified connects in interesting ways to many theoretical discourses in the field, but it may not be clear to you yet how it contributes to each and every one. Science need not wait for all these specific contributions to be identified; it can be important to simply get the message out (Wiesche et al., 2017, p. 695). Across the social sciences, the value of such contributions is gaining recognition, especially when studying emergent phenomena for which available theoretical lexicons are underdeveloped (Ågerfalk, 2014; Hambrick, 2007; Robinson, 2019).

But not all patterns automatically offer a contribution. We do not purport to have universal rules for when a pattern is or is not an acceptable contribution, nor when such patterns merit full papers or shorter contributions such as research notes. This needs to be a judgment among the review team and authors. Nevertheless, we offer two criteria for assessing theoretical contributions of all sorts, but which can especially guide the assessments of theoretical implications: the *originality* of the work and its *importance* to practice and to science (Corley & Gioia, 2011; Thatcher & Fisher, 2022). By the *originality* of the work, we refer to the novelty of the focal phenomenon or the surprise elicited by the surfaced patterns (Robinson, 2019). Findings that run counter to conventional wisdom, the prevailing discourse, or the conceptions of some communities, are particularly interesting (Davis, 1971). For example, amid utopian expectations of how digital media would democratize the public sphere, early research on political blogs spotlighted the polarization of political discourse online without advancing a theoretical explanation for that polarization (Adamic & Glance, 2005). Second, the contributions must be *important*—that is, likely to stimulate future research or alter theory or practice (Hambrick, 2007). If a novel pattern is important, i.e, likely to impact a theoretical discourse or practice, it should be considered for publication.

Both types of contributions—theorizing and describing patterns with theoretical implications—require mindfully relating one's insights to some theoretical lexicon. When analyzing patterns, the researcher neither has a proposition in mind nor is the explanation directly observed—it is abduced from both the empirical data and the theoretical lexicon. Moreover, patterns surfaced from a very novel phenomenon may not resonate immediately with an established theoretical lexicon, posing challenges to the researcher trying to situate it in a cumulative tradition. For example, Salge and Karahanna (2018) tackled the new phenomenon of social bots. Staying close to their data, they positioned the bots within the social network lexicon, noting their centrality to Twitter networks, and within the activism literature, noting their use to manipulate Twitter users and mobilize collective action. In a later study, Salge et al. (2021) uncovered further patterning of social bots and situated these patterns in the lexicon of conduit brokerage.

Grover and Lyytinen suggested we "encourage and value rigorous descriptive inquiries" in order to develop innovative theory (in press). We agree! Only when situated in an ongoing discourse though do such insights contribute to understanding. Our perspective stands in contrast to the unattainable nomothetic ideal of a theoretical contribution as fully generalizable, universal laws. The knowledge we generate is always provisional. The reality of what we study shifts—especially in the IS field. Patterns that inform theory at one time may not even exist at another. Better instrumentation and more comprehensive data may reveal novel patterns about previously studied phenomena. To render scholarly outcomes attainable, researchers bracket out portions of the phenomenon as they construct their explanations (Barrett & Orlikowski, 2021). Therefore, it is important to be mindful of such bracketing by specifying boundary conditions that establish the contextual and temporal limits of the theoretical statements and to invite further research to generalize and elaborate on those statements (Morgan, 1980). And once again, the relevance of a pattern and its associated theoretical insight is necessarily relative to the state of the discourse, as reflected by a theoretical lexicon.

Stopping Rule 2: Are We Done Sampling?

Data is a frequent starting point for computationally intensive work. Guided by general questions or a favorite computational approach, researchers seek out novel and interesting datasets. But when have they sampled enough data to justify their inferences? With computational trace data, in particular, one can always conceivably collect more. To guide researchers' engagement with such data, we draw upon the accumulated wisdom of two traditions—the statistical tradition and the case-based tradition. For each alternate path summarized in Figure 5, we identify key criteria for establishing stopping rules.

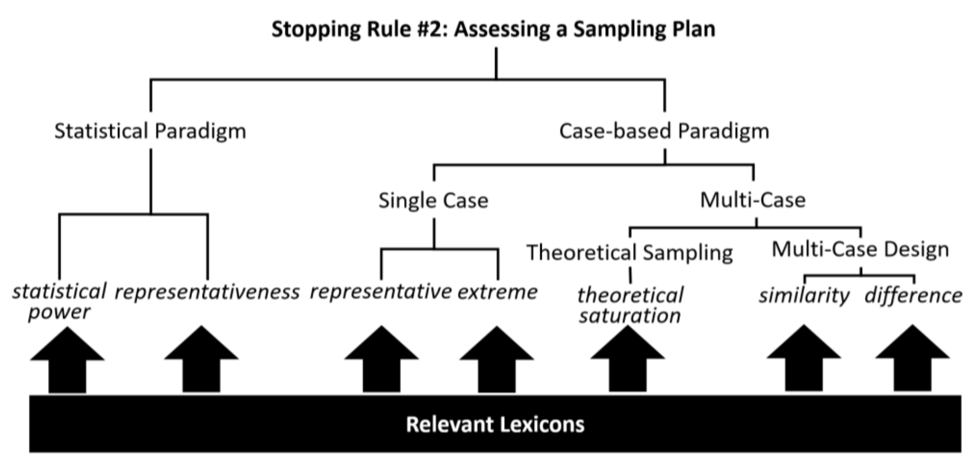


Figure 5. Am I Done Sampling?

Statistical paradigm: Within the statistical tradition, the hallmarks for sampling adequacy are representativeness and statistical power. Sample *representativeness* is typically justified by randomly sampling from the population. If relevant subgroups are known, stratified random sampling enables the representativeness of each subgroup. There are established statistical techniques for determining the adequacy of *statistical power* for generalizing inferences from empirical data to a population (Neter & Wasserman, 1974). Although one can justify a sample through statistical rationality, no sample is ever truly representative and no inference can be universally justified (Tsang & Williams, 2012). A sample drawn from one population does not apply to other populations. Further, populations change. Thus, even within a statistical paradigm, sampling requires judgment calls. One can think of a particular dataset as a limited set of cases, or perhaps even a case itself, and qualitative, case-based research traditions can help with justifying sampling decisions.

Case-based paradigm: The sampling strategy of qualitative research is case-based. A *single case*—an event, person, team, organization, or community—can be adequate for generating a contribution if selected specifically for its potential theoretical implications. There are two ways to leverage a single case: representative cases and extreme cases (Van de Ven, 2007; Yin, 2003). *Representative cases* are chosen because they are typical of a domain. Although some cross-case replication may be ideal, sometimes this may be prohibitively difficult. If a case is defensibly representative and interesting patterns are identified, then a single case can be adequate, particularly when the case is expected to be—and turns out to be—revelatory (Dubé & Paré, 2003). Examples of single cases include Adamic and Glance's (2005) political blogosphere study and Selander and Jarvenpaa's (2016) study of Amnesty International's initiatives. *Extreme cases* are unique in some way and are particularly descriptive of certain key phenomena. Consider, for example, Giddens et al.'s (in press) study of an organization tackling human trafficking. Such edge cases can spotlight a social problem that has eluded research attention to date.

One *multicase* approach is *theoretical sampling,* an approach situated within the grounded theory tradition, in which sampling and data analysis occur iteratively (Urquhart, 2013). Through constant comparison, the researcher moves between data, concepts, and relationships, and back to data to abstract insights from specific empirical indicators (Glaser & Strauss, 1967). Through such theoretical sampling, the researcher lets previously detected patterns and anomalies guide the selection of further data. This process continues until iterative sampling produces no new insights and *theoretical saturation* is attained (Glaser and Strauss 1967). This iterative process underlying traditional grounded theory is predicated on the cost of data collection, where obtaining access to an additional site or additional respondent can be challenging.

A second multicase strategy is *multicase design* (Eisenhardt, 1989). In pursuing this strategy, researchers may sample for similarities or differences among cases. When sampling for *similarity*, the goal is to replicate observed patterns. This strategy facilitates the discovery of robust patterns and the elimination of noise irrelevant to the pattern. When sampling for *difference*, researchers select specific cases based on theoretically informed criteria across which they wish to compare. This permits the researcher to surface contrasting patterns that offer greater theoretical insight (Eisenhardt & Graebner, 2007). In these designs, cases are sampled based on theoretical arguments of representativeness. Some studies may sample for both similarity and difference. For example, Tidhar and Eisenhardt (2020) applied machine learning to 40 cases of iOS apps across two product categories of varying levels of popularity and used different revenue models to glean insights into the fit between value capture and value creation. Thus, samples do not require statistical grounds (random sampling and statistical power) if they are defensibly and convincingly representative of the similarity and/or difference anticipated.

Regardless of the tradition informing one's sampling strategy, believing that large samples render the patterns identified more generalizable is a fallacy (Tsang & Williams, 2012). Instead, we urge review teams to consider existing quantitative and qualitative research guidance in judging the appropriateness of sampling choices.

Stopping Rule 3: Are the Methods Rigorous?

Researchers engaging in computationally intensive theory construction are like detectives using the methods and evidence at hand to construct patterns and inform theory. Sometimes they may surface patterns toward contributions using a single method. At other times, it may take multiple computational methods, perhaps in conjunction with traditional qualitative or quantitative approaches. Changing methods may allow researchers to discover different patterns in their data, thereby highlighting different aspects of the phenomenon. Researchers thus could conceivably engage with methods ad infinitum, but this would not be reasonable. The question in employing different methods is: What are some ways in which the researcher can ensure sufficient rigor while at the same time keep the investigation reasonable? Key stopping rules relate to when concepts are adequately operationalized and when inferences are sufficiently justified. Figure 6 summarizes some considerations to help with this determination.

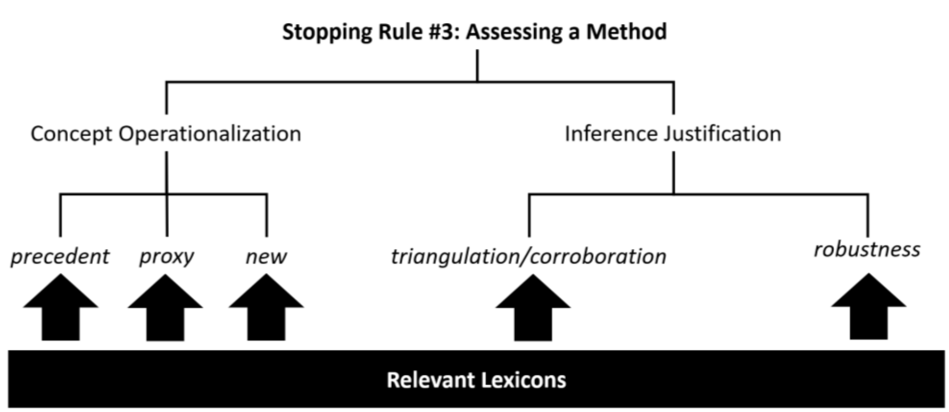


Figure 6. Am I Done Analyzing?

Concept operationalization, that is, linking an abstract concept to measurable observations, is the first methodological challenge facing the computationally intensive researcher. The researcher may adopt three approaches to establish this reasonability: precedent, proxy, or justifying the creation of a new measure. When a concept preexists one's study, that is, is already part of the theoretical lexicon or perhaps can be drawn from another lexicon, prior operationalizations provide a *precedent*. In this case, it is incumbent on the researcher to be aware of and use such precedents. For example, a rich diversity literature preceded Miranda et al. (in press), providing a Herfindahl-based measure of framing diversity. In drawing upon such precedents, the researcher should exercise caution to ensure that the prior context is sufficiently analogous to the present one and that the prior measure is not being misappropriated. Sometimes even with precedent, operationalizations need to be adapted. For example, Miranda et al. adapted the Herfindahl concentration index to measure diversity by subtracting it from one. If no viable precedent exists, the researcher can use a *proxy*. A classic example of a creative proxy involved using the replacement rate for floor tiles around museum exhibits to assess the popularity of the exhibits (Webb et al., 1966). Proxies that originate from the theoretical or another lexicon can be excellent ways to operationalize concepts, but must be well justified. For example, researchers may co-construct or vet the measure with a panel of experts. Or they may assess the extent to which the proxy covaries with measures of related concepts, but not with those for unrelated concepts (i.e., convergent and discriminant validity). A new concept is likely to require the researcher to design a *new* measure. Here too, the researcher may be able to leverage analogical reasoning for a viable metric. For example, Miranda et al. borrowed the cosine similarity measure used in information retrieval research (e.g., Piezunka & Dahlander, 2015) to assess their concept of foreshadowing. Proxies and new measures will inevitably invite challenges from the review team. When these occur, though, we encourage editors to engage with authors around their proposed measures. Possibly, the metric itself may be sound but inadequately justified. Or the dataset may permit a viable operationalization that the authors have not considered.

Inference justification is the second methodological challenge. The underlying question is, "How do I know this pattern is not just an artifact of current procedures?" This challenge can be addressed using triangulation and corroboration as well as robustness checks. *Triangulation and corroboration* entail using different types of data or different methods to assess the convergence of accounts (Denzin & Lincoln, 2007; Silverman & Marvasti, 2008). In quantitative grounded theory, triangulation is effected through the use of alternate indices for each concept; when associations among these alternate indices for two concepts are consistent, one can conclude that the inferred association between the concepts is justified (Glaser, 2008). Additionally, researchers can compare qualitative and quantitative views of their data, that is, they can compare inferences from their metrics with patterns they have manually identified in the data. For example, Miranda et al. (in press) viewed news and social media as one type of actor, and firms and governments as another. Cross-validation in large datasets can similarly provide corroboratory evidence in support of conclusions.

Robustness checks are standard practice in many traditions, permitting researchers to assess the sensitivity of their findings to different assumptions (Lu & White, 2014). Such practices are valuable also in computationally intensive theory construction and can be easily appropriated from the domain of testing theory to that of constructing theory. Often though, what begins as a "big data" investigation is progressively reduced to a small *n* problem as the unit around which inferences are drawn changes. For example, although Miranda et al. (in press) began with a large text corpus, many of their inferences were drawn around how seven types of actors were represented in that corpus across six years. To assess the robustness of their inferences, they built on statistical ideas of leverage points (Bollen & Jackman, 1985) and jackknifing (Efron & Stein, 1981), and they devised an approach to assess the robustness of their conclusions: the direction of an assessed association between two concepts should remain unchanged with the exclusion of each data point in turn.

In assessing methodological rigor, it is critical to remember that we are engaged in theory construction, not theory testing. As we adopt triangulation and corroboration, as well as robustness practices from hypothetico-deductive approaches, we urge review teams to be mindful of balancing rigor and reasonableness toward theory construction. Although we encourage authors and review teams to establish a valid trail of evidence for their inferences, these inferences should not be required to meet hypothesis testing standards. Theory construction is an exploratory, not (dis)confirmatory, endeavor. The contributions of computationally intensive theory construction involve constructing creative, novel, and impactful, yet tentative and fallible, statements that stimulate follow-on research. The idea is to provide a brick toward the construction of an edifice.

While we have used terminology implying certain epistemological and ontological assumptions—including notions of triangulation and robustness—our position is that computationally intensive theory construction can be undertaken within different paradigms and is ontologically and epistemologically agnostic. The key elements of lexical framing, pattern surfacing, and stopping rules pertain, regardless of one's ontological and epistemological positions. We address how researchers can deal with the challenges of this "big tent" approach by proposing three key principles for computationally intensive theory construction next.

Key Principles for Computationally Intensive Theory Construction

We have discussed and advocated for specific practices in the conduct and assessment of computationally intensive theory construction. These are by no means an exhaustive set of concerns. In our own work, we have found the three overarching research principles advocated by Bourdieu (Bourdieu & Wacquant, 1992; Wacquant, 2006) to offer invaluable guidance in filling the gaps. We, therefore, describe these three principles and why they are particularly salient to this approach to research.

Reflexivity: This principle entails an attitude of questioning and mindfulness. A danger of data-driven theory construction is that researchers run the risk of "merely 'rationalizing' what they see" in their data (Berente et al., 2019, p. 56). We draw attention to the need for reflexivity because it is neither possible nor desirable to specify a viable "script" for computationally intensive theory construction that precludes such rationalization. Instead, we call for reflexivity, through which the researcher is introspective about their experiences, ideas, and paradigms that may blind them to some categories in their data, thereby surfacing and challenging their tacit assumptions, beliefs, and expectations (Bourdieu & Wacquant, 1992). The purpose is not simply to justify the insights one offers, but also to elaborate on them to the extent permitted by the data and to consider the implications of one's work for those represented in it and those who might consume it (Creswell, 2007). Importantly, reflexivity "is a collective enterprise rather than the burden of the lone academic" (Bourdieu and Wacquant 1992, p. 36). This means that it is incumbent on a review team to ensure the authors have undertaken due diligence, but also to ensure that adherence to taken-for-granted notions about methods does not stymie the authors' ability to develop and share their insights.

Methodological pluralism: Bourdieu exhorted the researcher "to deploy whatever procedure of observation and verification is best suited to the question at hand and continually confront the results yielded by different methods" (Wacquant, 2006, p. 266). Similarly, Alford (1998, p. 52) noted that cognizant methodological pluralism "may help you reformulate your research questions, self-consciously locating them within foreground or background paradigms of inquiry." Such methodological pluralism clearly contributes to reflexivity. Methodological pluralism is not necessarily the same as mixed methods. Yes, researchers often combine computational and qualitative analyses quite effectively in their pursuit of theory construction (e.g., Lindberg et al., 2016). But methodological pluralism also embraces combining disparate computational techniques, possibly from different methods families and requiring different lexicons, toward gaining deeper insight. While this enriches the scholarly object, it introduces challenges in the review process, calling for reviewers with disparate methodological skills.

Epistemic attention to all research operations: Bourdieu et al. (2011, pp. 1-2) railed against the divides between the "prophets who fulminate against the original impurity of the empirical" and "the high priests of methods," noting that prioritizing either theoretical or empirical work was detrimental to advancing knowledge. Algorithmic work can assume a veneer of technological rigor that can "lead to a displacement of epistemological vigilance," wherein the meaningfulness or ethicality of a technically adept measure goes unquestioned (Bourdieu et al., 2011, p. 9). Bourdieu called for attentiveness to every research activity, even the most mundane. This requires the researcher to be diligent and meticulous in tracking their activities and holding these activities up for scrutiny. Transparency and traceability are essential for establishing any contribution and *MIS Quarterly*'s policy requires it (Burton-Jones et al., 2021). Because of the creative and iterative nature of the theory construction process, it is easy for some researchers to shortcut their empirical engagement in favor of a plausible story. On the other end, it is easy for the computationally driven researcher to shortchange the thought experiments and lexical engagement necessary for theoretical elaboration and establishing boundary conditions. Thus, to construct a coherent theory, the researcher needs to attend to both the theoretical lexicon, research operations, and their interplay.

Conclusion and Editorial Reflections

We have referred to this method as computationally intensive theory *construction* to highlight that the process and resulting product are social constructions. The use of computational techniques may cast an aura of objectivity on the theoretical product. However, "facts are necessarily suffused with theory"—formal as well as lay theory—that shapes the theoretical contribution (Wacquant, 2006, p. 265). Path dependencies throughout a research study compound the downstream impacts of subjectivity and non-determinism on the scholarly objects constructed. Nonetheless, this genre is consistent with alternate philosophical assumptions.

The framework, stopping rules, and key principles advanced in this editorial are not intended to be the last word. Our intention is to provide an accessible starting point for scholars who are interested in using computational methods to generate theory as well as a "brick" for the community of scholars interested in advancing our thinking about the "edifice" of computationally intensive theory construction as a new research genre—one that complements existing genres yet contributes to the cumulative tradition. We hope that our framework and the examples we offer will help ease some of the anxieties of this endeavor and encourage more scholars to take up the challenge of computationally intensive theory construction.

In the end, research is a conversation. We are excited to start this conversation with you and to see how you can take this conversation further. This editorial, likewise, resulted from a conversation among the editors—each with our own perspectives and priorities. It does not represent what any of us would have written on our own, but it is what we agreed to together. Accordingly, we conclude with one major idea that each of us hopes *you* will take away from this editorial.

Shaila: I exhort readers to internalize the need for discursive flexibility. Each core principle for this genre—reflexivity, methodological pluralism, and epistemic attentiveness to all research operations—necessitates discursive flexibility. Being reflexive means entertaining alternate lexicons as one surfaces one's own assumptions or faces review team challenges with openness. Embracing methodological pluralism entails assembling lexicons from disparate traditions in potentially unprecedented ways. Epistemic attentiveness to all operations necessitates iterating between practice, methods, and theoretical lexicons. For authors, I reference Weick's (1995a, p. 286) advice on flexibility: "Those who have more ideas often show less attachment to any one idea and are more receptive to suggestions about ways to improve specific ideas"; such authors, he suggested, are more apt to successfully navigate the review process. My request to review teams is this: Computationally intensive theory construction can promote bold and novel insights that elude the more structured research genres. To cultivate such possibilities, consider nurturing authors of promising works, whose "story" is yet unclear. Permit—even challenge—them to evolve their lexicons as their insights deepen through the review process.

Nick: The main notion that I hope readers take away from this editorial is that patterns—in all of their computational glory—can be contributions in their own right. Novel visualizations, observations, and narratives using computational approaches that give us insight into the world can and should be published. The abundance of data and novel methods offers opportunities to generate truly interesting and innovative insights. But this requires that we break out of established scripts that force us to package what we see in some tired theory from the 1970s, or force us to meet strict tests of unidirectional causation that limit the sorts of data admissible for research. We need to break out of this mold. This does not mean that any social network image or cool visualization merits a publication in *MISQ*. This is about theory construction, after all, and ignoring prior theory risks recreating it with a different label. But we might relax the need to problematize prior theory to motivate our work, and instead highlight the implications of our findings in a particular theoretical discourse—perhaps at the end of papers. In the end, whether some pattern identified merits publication is a judgment call on the part of the review team. But I hope that judgment will err on the side of the impactful, the novel, the interesting, and the bold.

Stefan: My hope is that computationally intensive theory construction will provide space for researchers' creativity in combining data sets, methods, and theories in unorthodox ways. These innovative approaches will be—and already are—our field's next-generation tools to produce novel, important, and interesting insights about the fast-paced, emergent phenomena of the digital age. To allow individual researchers as well as our field to fully capitalize on this new genre, we need to carefully address one important tension—that between embracing creativity, emergence, messiness, and intertwining of manual and algorithmic activity on the one hand, and science's requirement of faithful, transparent, and rigorous reporting on the other. The scripts, models, and novel lexicons we develop for computationally intensive theory construction must adhere to those requirements without infringing on creativity. Against this background, we offer a scaffold of practices and principles that we believe will be important to this genre but that allow for scholars' mindful and creative departures.

Hani: The computational era of research brings exciting opportunities to leverage novel computational tools, methods, and techniques. As the youngest of my colleagues authoring this editorial, I still remember the excitement of learning and reflecting upon then-new computational methods and trying to research and publish using them. IS researchers are at the forefront of these advancements and I hope to see my colleagues continue this tradition of introducing new computational methods to make the case for their use in the IS field and advance our theoretical understanding of digitally enabled sociotechnical phenomena. Reflecting upon C. P. Snow's "The Two Cultures," I hope that computationally intensive theory construction develops into a genre that enables researchers to be creative and bold but also attentive to the cumulative tradition, thereby advancing the theoretical discourse in novel, interesting, and impactful directions.

Andrew: My main hope with this editorial is to provide a common ground for discussion between authors, reviewers, and editors for papers that adopt this genre. Because of the open-endedness and novelty of computationally intensive theory construction, I find that the degree of uncertainty among authors and reviewers, and the degree of co-development of contribution between authors and reviewers, is higher in this genre compared to other research genres. This can make for an exhausting review process for everyone! That is why we entitled this editorial "for authors and reviewers." We want to empower authors to craft and submit these papers, we want to empower reviewers to give value-adding reviews of such work, and we want to give all of them some common principles and language to aid their discussions in the review process. We hope that will lead to more excellent examples of this genre in *MIS Quarterly*.

Acknowledgments

We thank the ~400 participants in two author development workshops for their valuable comments. We also thank the *MISQ* editorial board for their feedback in our masterclass with them. For helpful suggestions on this editorial, we thank Ahmed Abbasi, Malmi Amadoru, Kalle Lyytinen, Semi Min, Jan Recker, Suprateek Sarker, Susan Scott, Amrit Tiwana, and Youngjin Yoo.

References

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology*, *21*(1), 93-113.

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 US election: Divided they blog. *Proceedings of the 3rd International Workshop on Link Discovery*.

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

Ågerfalk, P. J. (2014). Insufficient theoretical contribution: A conclusive rationale for rejection? *European Journal of Information Systems*, *23*(6), 593-599.

Alford, R. R. (1998). *The craft of inquiry: Theories, methods, evidence*. Oxford University Press.

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review*, *36*(2), 247-271.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*(4), 496-515.

Bahmanyar, R., Espinoza-Molina, D., & Datcu, M. (2018). Multisensor earth observation image classification based on a multimodal latent Dirichlet allocation model. *IEEE Geoscience and Remote Sensing Letters*, *15*(3), 459-463.

Barrett, M., & Orlikowski, W. (2021). Scale matters: Doing practice-based studies of contemporary digital phenomena. *MIS Quarterly*, *45*(1), 467-472.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, *22*, 725-730.

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

Blumer, H. (1954). What is wrong with social theory? *American Sociological Review*, *18*, 3-10.

Bollen, K. A., & Jackman, R. W. (1985). Regression diagnostics: An expository treatment of outliers and influential cases. *Sociological Methods & Research*, *13*(4), 510-542.

Bourdieu, P., Chamboredon, J.-C., & Passeron, J.-C. (2011). *The craft of sociology: Epistemological preliminaries*. De Gruyter.

Bourdieu, P., & Wacquant, L. J. (1992). *An invitation to reflexive sociology*. University of Chicago Press.

Bradley, R., Mockus, A., Ma, Y., Zaretzki, R., & Bichescu, B. (2020). Coordinating interdependencies in an open source software project: A replication of Lindberg, et al. *AIS Transactions on Replication Research*, *6*(1), 14.

Browne, G. J., Pitts, M. G., & Wetherbe, J. C. (2007). Cognitive stopping rules for terminating information search in online tasks. *MIS Quarterly*, *31*(1), 89-104.

Brunswicker, S., Almirall, E., & Majchrzak, A. (2019). Optimizing and satisficing: The interplay between platform architecture and producers' design strategies for platform performance. *MIS Quarterly*, *43*(4), 1249-1277.

Burton-Jones, A., Boh, W. F., Oborn, E., & Padmanabhan, B. (2021). Editor's comments: Advancing research transparency at MIS Quarterly: A pluralistic approach. *MIS Quarterly*, *45*(2), iii-xviii.

Chalmers, A. F. (1999). *What is this thing called science?* (3rd ed.). Queensland University Press.

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: What constitutes a theoretical contribution? *Academy of Management Review*, *36*(1), 12-32.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches (2nd ed.)*. SAGE.

Czarniawska, B. (2004). On time, space, and action nets. *Organization*, *11*(6), 773-791.
Davis, M. S. (1971). That's interesting. *Philosophy of the Social Sciences*, *1*(2), 309-344.
Denzin, N. K., & Lincoln, Y. S. (2007). *Collecting and interpreting qualitative materials*. SAGE.
Dubé, L., & Paré, G. (2003). Rigor in information systems positivist case research: Current practices, trends, and recommendations. *MIS Quarterly*, *27*(4), 597-635.
Džeroski, S., Langley, P., & Todorovski, L. (2007). Computational discovery of scientific knowledge. In S. Dzeroski & L. Todorovski (Eds.), *Computational discovery of scientific knowledge* (pp. 1-14). Springer.
Efron, B., & Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics*, *9*(3), 586-596.
Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532-550.
Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25-32.
Fisher, G., Mayer, K., & Morris, S. (2021). From the editors—Phenomenon-based theorizing. *Academy of Management Review*, *46*(4), 631-639.
Fox, P., & Hendler, J. (2011). Changing the equation on scientific data visualization. *Science*, *331*(6018), 705-708.
Gaskin, J., Berente, N., Lyytinen, K., & Yoo, Y. (2014). Toward generalizable sociomaterial inquiry: A computational approach for zooming in and out of sociomaterial routines. *MIS Quarterly*, *38*(3), 849-871.
Gibson, D. R. (2000). Seizing the moment: The problem of conversational agency. *Sociological Theory*, *18*(3), 368-382.
Giddens, L., Petter, S., & Fullilove, M. H. (in press). Information technology as a resource to counter domestic sex trafficking in the United States. *Information Systems Journal*.
Gioia, D. A., & Pitre, E. (1990). Multiparadigm perspectives on theory building. *Academy of Management Review*, *15*(4), 584-602.
Glaser, B. G. (2008). *Doing quantitative grounded theory*. Sociology Press.
Glaser, B. G., & Strauss, A. L. (1967). *Grounded theory: Strategies for qualitative research*. Aldine.
Goldberg, A. (2011). Mapping shared understandings using relational class analysis: The case of the cultural omnivore reexamined. *American Journal of Sociology*, *116*(5), 1397-1436.
Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, *30*(3), 611-642.
Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The perils and promises of big data research in information systems. *Journal of the Association for Information Systems*, *21*(2), 268-291.
Grover, V., & Lyytinen, K. (in press). The pursuit of innovative theory in the digital age. *Journal of Information Technology*.
Habermas, J. (1984). *The theory of communicative action: Reason and the rationalization of society* (Vol. 1). Beacon Press.
Hahn, J., & Lee, G. (2022). The complex effects of cross-domain knowledge on is development: A simulation-based theory development. *MIS Quarterly*, *45*(4), 2023-2054.
Haki, K., Beese, J., Aier, S., & Winter, R. (2020). The evolution of information systems architecture: An agent-based simulation model. *MIS Quarterly*, *44*(1).
Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, *50*(6), 1346-1352.
Handfield, R. B., & Melnyk, S. A. (1998). The scientific theory-building process: A primer using the case of TQM. *Journal of Operations Management*, *16*(4), 321-339.
Hannigan, T. R., Briggs, A. R., Valadao, R., Seidel, M.-D. L., & Jennings, P. D. (in press). A new tool for policymakers: Mapping cultural possibilities in an emerging AI entrepreneurial ecosystem. *Research Policy*.
Huff, A. S. (1999). *Writing for scholarly publication*. SAGE.
Hunsberger, C., & Alonso-Fradejas, A. (2016). The discursive flexibility of "flex crops": Comparing oil palm and jatropha. *The Journal of Peasant Studies*, *43*(1), 225-250.
Johnson, S. L., Safadi, H., & Faraj, S. (2015). The emergence of online community leadership. *Information Systems Research*, *26*(1), 165-187.
Kuhn, T. (1962). *The structure of scientific revolutions*. University of Chicago Press.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., . . . Gutmann, M. (2009). Life in the network: The coming age of computational social science. *Science*, *323*(5915), 721-723.
Levina, N., & Vaast, E. (2016). Leveraging archival data from online communities for grounded process theorizing. In K. D. Elsbach & R. M. Kramer (Eds.), *Handbook of qualitative organizational research: Innovative pathways and methods* (pp. 215-224). Routledge.
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.

Lu, X., & White, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of econometrics*, *178*, 194-206.

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.

Malgonde, O., Zhang, H., Padmanabhan, B., & Limayem, M. (2020). Taming complexity in search matching: Two-sided recommender systems on digital platforms. *MIS Quarterly*, *44*(1), 49-84.

Maruping, L. M., Daniel, S. L., & Cataldo, M. (2019). Developer centrality and the impact of value congruence and incongruence on commitment and code contribution activity in open source software communities. *MIS Quarterly*, *43*(3), 951-976.

Mattson, M. P. (2014). Superior pattern processing is the essence of the evolved human brain. *Frontiers in Neuroscience*, *8*, Article 265.

Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., & O'Connor, P. J. (2014). A machine learning approach to improving dynamic decision making. *Information Systems Research*, *25*(2), 239-263.

Mills, C. W. (1959). *The sociological imagination*. Oxford University Press.

Miranda, S., Kim, I., & Summers, J. (2015). Jamming with social media: How cognitive structuring of organizing vision facets affects IT innovation diffusion. *MIS Quarterly*, *39*(3), 591-614.

Miranda, S., Wang, D., & Tian, C. (in press). Discursive fields and the diversity-coherence paradox: An ecological perspective on the blockchain community discourse. *MIS Quarterly*.

Morgan, G. (1980). Paradigms, metaphors, and puzzle solving in organization theory. *Administrative Science Quarterly*, *25*(4), 605-622.

Müller, O., Junglas, I., Brocke, J. v., & Debortoli, S. (2016). Utilizing big data analytics for information systems research: Challenges, promises and guidelines. *European Journal of Information Systems*, *25*(4), 289-302.

Nan, N. (2011). Capturing bottom-up information technology use processes: A complex adaptive systems model. *MIS Quarterly*, *35*(2), 505-532.

Neter, J., & Wasserman, W. (1974). *Applied linear statistical models*, McGraw-Hill.

New Oxford American Dictionary. (2010). Pattern. In *The New Oxford American Dictionary* (3rd ed.). Oxford University Press.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Peirce, C. S. (1903). Lecture VII: Pragmatism and abduction. In C. Hartshorne & P. Weiss (Eds.), *Collected papers of Charles Sanders Peirce*. Harvard University Press.

Pentland, B. T., Liu, P., Kremser, W., & Hærem, T. (2020). The dynamics of drift in digitized processes. *MIS Quarterly*, *44*(1), 19-47.

Pentland, B. T., Recker, J., & Wyner, G. (2017). Rediscovering handoffs. *Academy of Management Discoveries*, *3*(3), 284-301.

Pentland, B. T., Vaast, E., & Wolf, J. R. (2021). Theorizing process dynamics with directed graphs: A diachronic analysis of digital trace data. *MIS Quarterly*, *45*(2), 967-983.

Piezunka, H., & Dahlander, L. (2015). Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, *58*(3), 856-880.

Popper, K. R. (1959). *The logic of scientific discovery*. Routledge.

Rai, A. (2016). Editor's comments: Synergies between big data and theory. *MIS Quarterly*, *40*(2), iii-ix.

Rai, A. (2018). Editor's comments—Beyond outdated labels: The blending of IS research traditions. *MIS Quarterly*, *42*(1), iii-vi.

Recker, J., Rosemann, M., Indulska, M., & Green, P. (2009). Business process modeling-a comparative analysis. *Journal of the Association for Information Systems*, *10*(4), 333-363.

Robinson, S. L. (2019). What is a pre-theory paper? Some insights to help you recognize or create a pre-theory paper for AMD. *Academy of Management Discoveries*, *5*(1), 1-7.

Safadi, H., Johnson, S. L., & Faraj, S. (2021). Who contributes knowledge? Core-periphery tension in online innovation communities. *Organization Science*, *32*(3), 752-775.

Salge, C., Karahanna, E., & Thatcher, J. B. (2021). Algorithmic processes of social alertness and social transmission: How bots disseminate information on Twitter. *MIS Quarterly*, *46*(1), 229-260.

Salge, C. A. D. L., & Karahanna, E. (2018). Protesting corruption on Twitter: Is it a bot or is it a person? *Academy of Management Discoveries*, *4*(1), 32-49.

Sayre, K. M. (1965). *Recognition: A study in the philosophy of artificial intelligence*. University of Notre Dame Press.

Schwandt, T. A. (1999). On understanding understanding. *Qualitative Inquiry*, *5*(4), 451-464.

Scott, J. (1988). Social network analysis. *Sociology*, *22*(1), 109-127.

Selander, L., & Jarvenpaa, S. L. (2016). Digital action repertoires and transforming a social movement organization. *MIS Quarterly*, *40*(2), 331-352.
Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science*, *32*(3), 856-880.
Silverman, D., & Marvasti, A. (2008). *Doing qualitative research: A comprehensive guide*. SAGE.
Singh, P. V., Tan, Y., & Mookerjee, V. (2011). Network effects: The influence of structural capital on open source project success. *Mis Quarterly*, *35*(4), 813-829.
Steen, L. A. (1988). The science of patterns. *Science*, *240*(4852), 611-616.
Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., . . . Bloom-Ackermann, Z. (2020). A deep learning approach to antibiotic discovery. *Cell*, *180*(4), 688-702. e613.
Strauss, C., & Quinn, N. (1997). *A cognitive theory of cultural meaning*. Cambridge University Press.
Sturm, T., Gerlach, J. P., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., . . . Buxmann, P. (2021). Coordinating human and machine learning for effective organizational learning. *MIS Quarterly*, *45*(3), 1581-1602.
Tang, T. Y., Fang, E. E., & Qualls, W. J. (2020). More is not necessarily better: An absorptive capacity perspective on network effects in open source software development communities. *MIS Quarterly*, *44*(4), 1651-1678.
Tashakkori, A., Teddlie, C., & Teddlie, C. B. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. SAGE.
Thatcher, S. M., & Fisher, G. (2022). From the editors—The nuts and bolts of writing a theory paper: A practical guide to getting started. *Academy of Management Review*, *47*(1), 1-8.
Tidhar, R., & Eisenhardt, K. M. (2020). Get rich or die trying… finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, *41*(7), 1245-1273.
Tiwana, A. (2018). Platform synergy: Architectural origins and competitive consequences. *Information Systems Research*, *29*(4), 829-848.
Tiwana, A., & Kim, S. K. (2019). From bricks to an edifice: Cultivating strong inference in information systems research. *Information Systems Research*, *30*(3), 1029-1036.
Tsang, E. W., & Williams, J. N. (2012). Generalization and induction: Misconceptions, clarifications, and a classification of induction. *MIS Quarterly*, *36*(3), 729-748.
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), 1179-1206.
Van de Ven, A. H. (2007). *Engaged scholarship: A guide for organizational and social research*. Oxford University Press.
van der Aalst, W. M., Schonenberg, M. H., & Song, M. (2011). Time prediction based on process mining. *Information Systems*, *36*(2), 450-475.
Verhagen, C. J. D. M. (1975). Some general remarks about pattern recognition; its definition; its relation with other disciplines; a literature survey. *Pattern Recognition*, *7*(3), 109-116.
Von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, *4*(4), 404-409.
Wacquant, L. (2006). Pierre Bourdieu. In R. Stones (Ed.), *Key sociological thinkers* (pp. 261-277). MacMillan.
Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Rand McNally.
Weick, K. E. (1995a). Editing innovation into Administrative Science Quarterly. In L. L. Cummings & P. J. Frost (Eds.), *Publishing in the Organizational Sciences* (pp. 284-296). SAGE.
Weick, K. E. (1995b). What theory is not, theorizing is. *Administrative Science Quarterly*, *40*(3), 385-390.
Wiesche, M., Jurisch, M. C., Yetton, P. W., & Krcmar, H. (2017). Grounded theory methodology in information systems research. *MIS Quarterly*, *41*(3), 685-701.
Yang, K., Lau, R. Y., & Abbasi, A. (in press). Getting personal: A deep learning artifact for text-based measurement of personality. *Information Systems Research*.
Yin, R. K. (2003). Design and methods. *Case study research*, *3*(9.2).
Zhang, K., Bhattacharyya, S., & Ram, S. (2016). Large-scale network analysis for online social brand advertising. *MIS Quarterly*, *40*(4), 849-868.
Zyphur, M. J. (2009). When mindsets collide: Switching analytical mindsets to advance organization science. *Academy of Management Review*, *34*(4), 677-688.