How Can Computational Social Science Motivate the Development of Theories, Data, and Methods to Advance Our Understanding of Communication and Organizational Dynamics?

Noshir Contractor

The Oxford Handbook of Networked Communication
Edited by Brooke Foucault Welles and Sandra González-Bailón

Print Publication Date: Feb 2020
Subject: Political Science, Political Behavior, Political Methodology
Online Publication Date: Mar 2019 DOI: 10.1093/oxfordhb/9780190460518.013.7

Abstract and Keywords

It is almost a decade since the article "Computational Social Science" was published in *Science* (Lazer et al., 2009). That article advocated for computational social science as a promising new arrow in the quiver for understanding and enabling social systems. The chapters in this section, and indeed in this book, are a testament to the decade-long trajectory of this movement. The chapters in this section also provide an opportunity to reflect on how computational social science can motivate the development of theories, data, and methods to advance understanding of current and emerging forms of communication and organizational dynamics. This chapter reviews some of the progress made on these dimensions and points to chapters in the section that serve as exemplars.

Keywords: computational social science, organizational dynamics, communication theory, social network theory, big data

1. Can Computational Social Science Motivate the Development of Theories of Communication and Organizational Dynamics?

THE advancement of organizational dynamics and communication theory is the most important benchmark by which the long-term impact of a new intellectual approach is evaluated here. In this section I outline four ways that computational social science is motivating developments that (1) test existing theories at scale; (2) extend existing theories to offer more nuanced insights; (3) generate new theories about existing (p. 114) phenomena by the inclusion and juxtaposition of concepts for which data were either unavailable or impractical to collect at scale; and (4) develop new theories about (relatively) new phenomena, such as the changing nature of organizing enabled by digital advances.

In its early stages, researchers were able to showcase the potential of computational social science to help test existing theories “at scale.” For instance, one of the best-known claims in network theory is that diverse social network ties provide individuals (and aggregates of individuals) greater access to social and economic opportunities (Burt, 2009). However, until the past decade these theories could only be empirically tested on relatively small networks, often made up of individuals within a single organization. These studies generated compelling evidence that organizational members who spanned “structural holes,” by connecting with others who were not directly connected, performed better than those who did not. Spanning structural holes gave those individuals access to social and economic opportunities for advancement. However, these ideas remained largely untested at the population level until a study conducted by Eagle and his colleagues (2010). Analyzing the call graph (network of who called whom on the phone) for the United Kingdom, they were able to demonstrate that individuals who had phone conversations with others who were not directly calling each other (e.g., those spanning structural holes) were more likely to reside in regions of higher social economic status. While the study left open the causal direction of this association—whether spanning structural holes leads to higher social economic status or vice versa—it provided an early example of how computational social science could be used to test existing theories at scale. The chapter by Benefield and Shen in this handbook utilizes the massively multiplayer online game (MMOG) EverQuest II to test several existing theories about gender roles and stereotypes. The chapter by Spiro in this handbook tests theories of social convergence that describe the coalescing of attention and people in the event of a crisis. These were previously used primarily tot study offline behavior, but they show that support for the theory is even more accentuated online. One of the theories that Hill and Shaw invoke in their chapter in this handbook is the well-established theory of the diffusion of innovation. They discuss how it is being used to study the diffusion of collaborative practices across peer production websites. In addition, they draw upon organizational population ecology theory, which posits that the fate of organizations is in large part determined not by what occurs inside them but by their position within the environment, including, for example, the carrying capacity of the niche they occupy. In his chapter in this handbook, Weber also builds on an ecology perspective but focuses on the community level, which posits that the fate of a population of organizations (in his case the traditional newspaper industry) is in large part determined by the community of industries in which they are embedded.

More recently, Aral and Nicolaides (2017) show how computational social science can be used not just to test existing theories but also to *advance* them by adding more nuance. Using data collected from over a million individuals over the course of five years, they showed that individuals’ exercise patterns were indeed influenced by those (p. 115) of others in their social networks, as predicted by theories of social contagion. More important, they were able to extend our understanding of the mechanisms by which social contagion operates. Prior research had argued that social contagion occurs as a result of the person potentially being influenced engaging in social comparison processes (Festinger, 1954) with the potential influencer. Aral and Nicolaides (2017) showed that individuals' social comparison processes led them to be more likely to engage in exercise activity to stay ahead of those slightly less active than they were, as compared to those who were slightly more active. The chapter by Benefield and Shen in this handbook explores how mentoring is impacted by mentors who are gender swappers. Gender swapping is by no means a new phenomenon; consider its deployment in no less than five of Shakespeare's plays. But Benefield and Shen showcase how digital trace data can offer new nuanced insights—especially about phenomena that are hard to observe (literally).

In addition to testing at scale and advancing existing theories, computational social science also has demonstrated the potential to *unleash new theories* that draw on explanatory variables and concepts that require leveraging and juxtaposing diverse data sources that were heretofore unavailable or impractical. One novel source of data, until recently *unavailable*, that shows considerable promise is functional magnetic resonance imaging (fMRI), which measures an individual's brain activity by detecting changes associated with blood flow. The approach is premised on the fact that cerebral blood flow reflects neuronal activation. For example, a recent study found that individuals whose friends were friends with each other were less likely to experience social exclusion (as measured by their fMRI) than individuals whose friends were less likely to be friends with each other (Schmälzle et al., 2017). Preliminary results from studies such as these hint at the prospect of building physiologically based, or at least physiologically informed, theories to explain the antecedents and outcomes of social networks. The power of gaining new insights by juxtaposing diverse disparate data led noted computer scientist Jim Hendler to herald the move away from big data to "broad" data (Hendler, 2013). The chapter by Spiro in this handbook describes the opportunities in government emergency response plans to link social media data posts by organizational entities with organizational-level features. In his chapter, Weber describes how he juxtaposed data from the Internet Archives with data from the *Editor and Publisher Yearbook*, as well as interviews, to turbocharge the explanatory power of theories of community ecology to explain the dynamics of change in the newspaper industry.

While access to fMRI data was until recently unavailable, other sources of data were available in principle but were *impractical* to encode at scale. One such example is the coding of group interaction data in a way that includes details about which individual is directing remarks to which other individual(s) in the group. Encoding these interactions requires painstakingly careful attention to various nonverbal cues such as eye gaze, body posture, and conversational distance. Today, thanks to advances in the capture of high-resolution video data and machine learning algorithms, we are able to automate the detection and use of nonverbal cues as a way of accurately determining which member(s) in the group were the senders and the intended recipients of (p. 116) interactions within the group (Mathur, Poole, Peña-Mora, Hasegawa-Johnson, & Contractor, 2012). The availability of these "big data from little teams" (Carter, Asencio, Wax, DeChurch, & Contractor, 2015) is leading to the development of new theories about how organizing in groups can be characterized by sequential structural signatures and to what extent the prevalence of distinct sequential structural signatures of interactions are systematically associated with how groups perform (Foucault Welles et al., 2014; Leenders, Contractor, & DeChurch, 2016; Schecter, Pilny, Leung, Poole, & Contractor, 2017).

In addition to testing, advancing, and developing new theories about extant phenomena, computational social science also has the promise of advancing our theoretical understanding of new phenomena. Many of the same digital forces that have propelled the emergence of computational social science as a promising mode of intellectual inquiry have inspired not only the “effervescence of collective behavior” (Gonzalez-Bailon, 2017) but also the emergence of disruptive novel forms of organizing such as peer production (Benkler, Shaw, & Hill, 2015; Hendler, Hall, & Contractor, 2018) and flash organizations (Valentine et al., 2017). These novel, agile, and often ephemeral forms of organizing are in turn inviting the development of a new generation of technologies to assemble (Asencio et al., 2015) and enable (Zhou, Valentine, & Bernstein, 2018) teams as well as understand and theorize about what explains their effectiveness (Contractor, 2013; Wax, DeChurch, & Contractor, 2017; Mukherjee et al. 2018). The chapter by Benefield and Shen in this handbook examines pick-up groups (PUGS), which are ad hoc teams that come together temporarily (typically for a few hours) to accomplish a specific task in the MMOG EverQuest II. Likewise, the chapter by Spiro in this handbook reports on research that can inform the design of tools for disaster management that facilitate automated discovery of potential collaborators in the midst of an emergency. The discussion of peer production in the chapter by Hill and Shaw in this handbook challenges conventional notions of what constitutes a “team.” Are two individuals who contributed independently and asynchronously (say, a year apart) to a joint Wikimedia page on the same team? Will they be considered as being on a team if one commented on and/or edited another person’s contribution? Irrespective of whether or not we label them as a team, there is no argument that this is a new form of organized, coordinated activity that invites new theorizing. One of the collateral opportunities afforded by these new technologies is that they allow us to employ computational social science methods to study at very high resolution the actions and interactions of individuals during the stage at which they search, court, invite, or decline requests to form into teams—a process that has historically been invisible until after the team is formed and only *if* it forms. Of particular note is Hill and Shaw’s call for moving from single-platform to multiplatform studies of online platforms. These are crucial in enabling us to generate new theories about how variations in technological affordances of platforms might shape the processes and outcomes of organizing on those platforms.

To summarize, computational social science over the last decade has demonstrated its potential to help us test existing theories at scale, extend these theories to offer more (p. 117) nuanced insights, develop new theories made possible by the juxtaposition of data that were either unavailable or impractical to collect at scale, and develop theories about new phenomena that are gaining salience in the wake of many of the same digital advances that are fueling computational social science.

2. Can Computational Social Science Motivate the Development of New Data Collection Instruments to Study Communication and Organizational Dynamics?

The growth of computational social science would have been impossible without the windfall of digital trace data. A growing proportion of the data currently being deployed in the study of communication and organizational dynamics is drawn from the Web. And in almost all cases, the data being analyzed were not collected for research purposes. In many instances these were either server-side logs made available to researchers, often via APIs and sometimes under nondisclosure agreements (NDAs), or were scraped off the Web using scripts. All of these *opportunistic* data collection efforts rely on what Salganik (2017) terms ready-made data, sometimes dismissively referred to as the inhalation of digital exhaust.

Remarkable insights have been gleaned by analyzing these opportunistic data sources. Conducting network and text analytics on situation reports published daily on the Web during natural disasters made it possible to automate the generation and evaluation of the interorganizational networks engaged in disaster response—in close to real-time and without having to impose on the already busy responders (Varda, Forgette, Banks, & Contractor, 2008). The chapter by Spiro in this handbook demonstrates the theoretical and analytical strides that continue to be made by leveraging opportunistic data to study online communication from 216 official emergency management–related Twitter accounts dealing with120 disaster declarations over the span of fifteen months. An early example of this effort was our ability to understand how individuals organized into guilds and went on quests in MMOGs such as Sony Online Entertainment's EverQuest II (Williams, Contractor, Poole, Srivastava, & Cai, 2011) and in virtual worlds such as Second Life (Foucault Welles & Contractor, 2015). These platforms also served as ideal crucibles to understand how the next generation of leaders, often as teens, were honing their teaming and leadership skills in these virtual environments (Reeves, Malone, & O'Driscoll, 2008). The chapter by Benefield and Shen in this book offers a compelling demonstration of the utility of such data to explore the impact of gender on networks. Weber (2018) was among the first communication scholars to see the research value of not just studying the Web as it is at a (p. 118) certain point but using the Internet Archives as the ultimate longitudinal opportunistic data source to study the dynamics of organizational—and indeed industry—changes. And the chapter by Hill and Shaw reports on their enormous success at curating one of the most definitive data sets from a population of peer-production sites based on the Wikimedia technology. Indeed, all four chapters in this section rely creatively and heavily on repurposing opportunistic online data.

Notwithstanding the unprecedented opportunities they offer, these data also surfaced some important limitations that discourage our sole reliance on ready-made data. Recent changes in the policies of social media sites such as Facebook in closing down API access first to personal pages and more recently to group pages are a harbinger of what Freelon (2018) has heralded as the "post-API age" for computational research. Aside from this potential shutout from digital trace data, most server-side logs were maintained by programmers for the primary purpose of debugging their software code. Organizations, increasingly recognizing the business potential of analyzing these data, are instrumenting the logs with those objectives in mind. Developers of the aforementioned MMOGs were among the first to recognize the potential of "re-instrumenting" server logs to include logging data that could provide insights for marketing, customer retention, and game design. More recently, developers of enterprise social media platforms such as Slack, Microsoft Teams, and Jive are also seeing the potential of conducting relational analytics using carefully instrumented logs to offer insights based on their clients' use of these platforms (Leonardi & Contractor, 2018). There is clearly an opportunity for researchers to engage closely with such platform developers in developing mutually beneficial collaborations. These collaborations will entail transferring current insights from research into, for instance, the implementation of algorithms on these platforms, but also providing the research community with the ability to purposively instrument these platforms to log digital traces that are geared to addressing research questions rather than only to help debug software or drive business goals.

These partnerships, while potentially promising, are not without risk. A collaboration can be abruptly terminated due to changes in key personnel, leadership, or ownership. In addition, the partnership will need to navigate significant intellectual property issues for the organization, and privacy issues for the users, that do not undermine the ability of the research to be published. King and Persily (2018) propose an innovative model that includes creating an entity, Social Science One (SS1), to explore a partnership between Facebook and universities brokered by the Social Science Research Council's Social Data Initiative to conduct research on the effects of social media on democracy and elections.

Alongside these approaches to engaging with organizations, it is also critical for the research community to innovate on the direct collection of data from participants unfettered by commercial constraints. Consider this the next generation extension of researchers designing carefully controlled experiments that relied on recruiting participants who came from primarily Western, educated, industrialized, rich, and democratic (WEIRD) populations (Henrich, 2010). Computational social scientists are **(p. 119)** increasingly relying on online platforms such as Prolific.ac and Mechanical Turk (Mason & Suri, 2012) to recruit a more egalitarian participant pool, carefully curated to minimize unfair labor practices (Semuels, 2018) and the increasing threat of bot-assisted participants or participant-assisted bots (Dreyfuss, 2018). In addition, efforts such as the development and deployment of experiments on the web-based Volunteer Science platform (Radford et al., 2016) have demonstrated the potential to not only scale up the participant pool but also engage in a concerted effort to build a community of researchers coordinating on broader questions (such as a fairer, safer, more understanding Internet, in the case of CivilServant.io) from a number of studies that can be conducted, collated, and compared across a common participant pool.

Beyond the Web, researchers are also recognizing the value of instrumenting humans directly in order to gain further insights about communication and organizational dynamics. The sociometric badge developed by Sandy Pentland and his team at the MIT Media Lab has been used to generate a new science of how to build teams (2012). Cattuto and colleagues (2010) have demonstrated, as part of the SocioPatterns project, the use of RFID technology to track collaboration networks, for instance at interdisciplinary scientific conferences.

In summary, while computational social science was catalyzed by the ability to opportunistically analyze large tracts of digital trace (or exhaust) data, the next generation of computational social science must consider more purposive instrumentation of online environments as well as personal wearable devices and apps offering what Salganik (2017) refers to as “custom-made” data.

3. Can Computational Social Science Motivate the Development of New Methods to Study Communication and Organizational Dynamics?

It is a well-established adage that the methods we are acquainted with shape the questions we ask (Monge, 1982). The high-resolution temporal data on actions, interactions, and transactions have challenged not only our theoretical explanatory frameworks, but also the limits of our methodological tools. For instance, until the past decade, our understanding of communication and organizational network dynamics was premised on the assumption that we had panels of longitudinal network data at discrete time intervals. The methods of choice to analyze these data were, for instance, stochastic-actor-oriented models (Snijders, 2005). However, the advent of time-stamped data chronicling every single relational event between a sender and a receiver propelled the development of a new approach to modeling network dynamics: relational event models (Brandes, Lerner, & Snijders, 2009; Butts, 2008). These models (p. 120) explain the timing as well as the sender and receiver of every relational event as a conditional function of all previous relational events in the organizational context (Leenders et al., 2016; Pilny, Schecter, Poole, & Contractor, 2016). Many extant theories of organizing posit macro-emergent states as being shaped, leveraged, and aligned with microprocess mechanisms (Kozlowski & Ilgen, 2006). However, as Kozlowski (2015) notes, we have stopped short of precisely articulating, let along testing, the temporal and sequential unfolding of these microprocess mechanisms. Relational event models provide a framework to posit these microprocess mechanisms as precise sequential structural signatures and test if the prevalence of these signatures is associated with certain emergent states. Consider the well-established body of research going back twenty-five years, relating boundary spanning in organizational teams to performance (Ancona & Caldwell, 1992). This research has generated mixed results on the impact of boundary spanning on performance (Marrone, 2010). Relational event models have the potential of taking collapsed data on boundary spanning and parsing it as a sequence of directed interactions: for instance, who spoke when with whom outside the team, and was it preceded or followed by an interaction within the team? Sequential structural signatures, such as these, have the potential to theoretically enrich our understanding of boundary spanning and potentially disambiguate the mixed results found in prior research. In her chapter, Spiro notes that one of the more striking results was the fact that the underlying social network among the emergency management organizations did not change during a crisis, even one that was severe. Interestingly, she has the data that will enable us to consider the possibility that while the "snapshot" structure of the network might not have changed, the sequential structure of how the network links unfolded—inferred using relational event models—might look very different in a crisis.

These sequential structural signatures are being augmented and enriched by methodological advances in text analytics. While content analysis (Krippendorff, 1980) has been a mainstay of social science research for decades, the increasing availability of text data in digital form and novel computational techniques are changing the scale and scope of our ability to utilize them as useful "telescopes" to probe human attitudes and behavior (Gonzalez-Bailon & Paltoglou, 2015). The turn of the century witnessed the development of several topic-modeling techniques such as latent semantic analysis (Dumais, 2004) and latent Dirichlet allocation (Blei, Ng, & Jordan, 2003). These were developed by computer scientists who were "much better at building powerful telescopes than at knowing where to point them" (Golder & Macy, 2014, p. 146). Meanwhile, social scientists such as Pennebaker and his colleagues (2015) developed tools such as Linguistic Inquiry and Word Count (LIWC) that were less computationally sophisticated but easier to use and interpret (they were word counts) by the social science community. For instance, these analyses revealed that leadership is closely related to the use of collective pronouns such as "we" and "us" rather than "I" or "me." More recently there has been a move from "frequency counts" to mapping meaning. These employ vector space models (VSMs) of semantics (Turney & Pantel, 2010) that leverage a large corpus of text from locations such as Google (Le & Mikolov, 2014; (p. 121) Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). The chapter by Spiro discusses how topic modeling can illustrate differences in the content of communication among emergency management organizations between emergency and nonemergency event days. It can also glean the differences in topics communicated between organizations representing different functional roles, thereby adding content to what was previously an interorganizational contact network.

The prevalence of large volumes of high-resolution data has also accelerated methodological developments in computational modeling of the dynamics of communication and organizational systems. When there was a dearth of dynamic empirical data, computational (and more specifically agent-based) models focused, by necessity, on developing simple, stylized models of social phenomena to explore how changes in inputs or mechanisms might impact emergent outcomes. For instance, simple computational models were able to demonstrate the plausibility of preferential attachment as a theoretical mechanism to explain the widespread prevalence of scale-free social networks (Wilensky, 2005). These were often referred to as *intellective* computational models (Pew & Mavor, 1998). The parameters in these computational models were often arbitrarily chosen and defended on theoretical grounds and/or resulted in emergent outcomes that were robust to modest changes in the values of these parameters. Today, with the availability of temporal data, we are witnessing a surge in the development of *emulative* computational models (Carley & Hirshman, 2011). These much larger models seek to emulate in substantial detail the dynamic features and characteristics of a specific team or organization (Carley, 2009). They have a much larger number of parameters; in the past, the modeler would have to specify values of the parameters informed by theories or the context being modeled. However, spurred by the availability of large amounts of dynamic empirical data, recent advances obviate the need for modelers to specify parameters. Instead we are able to leverage novel genetic algorithms and optimization techniques to empirically estimate these parameters (Stonedahl & Wilensky, 2010; Sullivan, Lungeanu, DeChurch, & Contractor, 2015; Thiele, Kurth, & Grimm, 2014). Using empirical data to estimate parameters in a computational model makes it analogous to a statistical (e.g., regression) model. This semblance has the potential to assuage the skepticism of traditional social science researchers, who have been understandably wary of deriving insights from computational models in which the specification of the parameters was (arguably) arbitrary.

The proliferation of data available on the formation and performance of millions of overlapping teams on online platforms, such as Wikipedia, Github, and Kaggle, has also motivated a renewed interest in the development of methodologies leveraging hypergraph methods that represent teams as hyperedges rather than a collection of edges that fails to preserve the team's entitativity (Lickel, Hamilton, & Sherman, 2016). Put simply, a team of three individuals can be represented in a network by three nodes connected by three edges. However, this representation loses information about whether this is one team of three individuals or three teams of pairs of individuals. If our goal is to study why individuals assemble into teams, there is a fundamental difference between explaining why A, B, and C (a hyperedge) assembled into one (p. 122) team, versus A and B, A and C, and B and C pairing (as edges) into separate teams. Likewise, if our goal is to understand the impact of communication on an organizational outcome such as performance, it is fundamentally different to assess the impact of a (face-to-face or email) private interaction between A and B and another between A and C (both edges) versus a joint interaction involving A, B, and C (a hyperedge). Clearly, edges and graph theory are not the most appropriate way to analyze how collectives form and perform from a network perspective. In response we have seen advances leveraging the study of hypergraphs in which a hyperedge, unlike an edge, is not confined to connecting only two nodes (Berge, 1979; Ghasemian, Ghasemian, Zamanifar, Zamanifar, & Ghasem-Aghaee, 2017; Taramasco, Cointet, & Roth, 2010). Hypergraphs also enable us to measure the overlap between two teams that are defined as team interlocks. Team interlocks have been shown to be important predictors of the success of scientific teams (Lungeanu, Carter, DeChurch, & Contractor, 2018). While not explicitly invoking hypergraph methods, the chapter by Hill and Shaw in this handbook invokes hypergraph thinking by explaining the success (and failure) of peer production sites as well as the spread of ideas and practices across sites, based on overlapping membership across communities. More recently, there has also been an effort to extend relational event models, discussed previously, to model not just an edge (a dyadic relationship between two nodes) but a hyperedge event that represents, for instance, an email sent by one person to two or more others (Kim, Schein, Desmarais, & Wallach, 2018).

In addition to propelling new methodological advances in areas such as relational event models, text analytics, computational agent-based models, and hypergraphs, computational social science has also invited careful re-examination of the classic approaches to research design and causal inference (Cook, Campbell, & Shadish, 2002). The excitement associated with the influx of computational social science scholarship from multiple (including non-social-science) disciplines rushing into this uncharted territory is reminiscent of what then Federal Reserve Bank chairman Alan Greenspan referred to as the "irrational exuberance" associated with the dot-com bubble in the 1990s. This excitement has to be tempered with careful reflection on how these new modes of asking and answering questions require us to "modernize—but not replace" (Salganik, 2017, p. 6) the classic approaches. For instance, we know that in large samples, *p*-values rapidly go toward zero, thereby impacting our traditional norms of using them to conduct tests of statistical significance. These have led, for instance, to a call for also reporting effect sizes in addition to *p*-values to safeguard against making intellectual claims that are statistically significant but have no "practical significance" (Lin, Lucas, & Shmueli, 2013, p. 906) or, more radically, change the default *p*-value from 0.05 to 0.005 (Benjamin et al., 2018).

More generally, computational social science is prompting us to reconsider the classical debates between theory driven research (TDR) and data driven research (DDR). Over the past half a century, there has been a strong preference for theory-driven research (TDR) over data-driven research (DDR) to advance our understanding of communication and organizational dynamics. Often referred to pejoratively as "dust (p. 123) bowl empiricism," in hindsight, part of the skepticism about DDR might have been grounded in the paucity of, or the inability to procure, large numbers of independent data sets where exploratory insights from one data set could be confirmed on another. Pure DDR has been fairly criticized for asking questions driven by the availability of data. This has led to comparisons with the drunken man who looked for his keys under the lamppost, not because that was where he lost them, but because that was where it was illuminated. Indeed, given the attention focused on data collected from Twitter, the *New York Times* (Zimmer, 2011) asked with tongue in cheek if "Twitterology" was a new science. Unfortunately, the same can also be said for much of pure TDR. We have a propensity to ask questions that lend themselves to be addressed by leveraging existing theories: looking for answers under the proverbial theoretical lamppost. This has led to calls for "taking off the theoretical straight jacket" (Schwarz & Stensaker, 2014). Benefield and Shen's chapter in this volume explicitly pursues a dual TDR and DDR approach. And the chapters by Hill and Shaw, Weber, and Spiro all report research using multiple methods to address the TDR-DDR cycle.

The debate between TDR and DDR has been joined by those who champion the value of phenomenon-driven research (PDR) (Schwarz & Stensaker, 2014). The importance of attempting to understand and help improve organizational phenomena was well captured in Lewin's classic adage "nothing is quite so practical as a good theory" {Lewin, 1951, p.

486}. PDR is motivated by a desire to solve problems associated with real-world phenomena (Watts, 2017). Advances in digital technologies have triggered a slew of new, or at least dramatically scaled-up, organizational phenomena, leading to novel communication and organizational dynamics that need to be understood and enabled. Understanding these phenomena should not rely on existing TDR, nor should it seek to de novo generate insights solely from DDR. Instead, PDR invites a delicate iterative waltz between TDR and DDR leveraging existing theory, modifying it to better fit the data, and sometimes necessitating the development of a new, more parsimonious theory, which must then be tested with new data (Mathieu, 2016). This iterative dance is distinct from deductive or inductive inference and is referred to as *abductive inference* (Haig, 2005; Halas, 2011; Meyer & Lunnay, 2013; Ren et al., 2018). The data driven segment of this iterative loop lends itself well to the utilization of data mining and machine learning techniques to classify and predict certain outcomes. These approaches will benefit immensely from recent interest in the interpretability of machine learning algorithms, which seek to uncover the logic in the algorithms making the classifications and predictions (Wang, Rudin, Doshi-Velez, Liu, Klampf, & MacNeille, 2017; Vellido, Martin-Guerrero, & Lisboa, 2012). Interpretable algorithms have the potential to inform the TDR segment of this iterative loop. They will also help address an enduring debate in the social sciences about the relative merits of prediction versus explanation (Hofman, Sharma, & Watts, 2017).

A decade after the essay on computational social science in *the journal Science* (Lazer et al., 2009), and in part propelled by its growth, we are seeing a concerted and organized effort by journals, institutions, and funding agencies to help evolve the norms and (p. 124) incentives associated with social science inquiry. Journals, such as *Nature Human Behavior*, solicit as one form of submission a “registered report” in which methods and proposed analyses are preregistered and reviewed prior to data collection. If the review is favorable, and the research is conducted as proposed, the results are guaranteed to be published irrespective of the findings, thereby obviating the bias that, for instance, statistically significant results are three times more likely to be published than papers with null results (Dickersin, Chan, Chalmersx, Sacks, & Smith, 1987). Institutions such as the Center for Open Science are creating platforms like the Open Science Framework (https://osf.io/) to serve as “a scholarly commons to connect the entire research cycle” with the goal of promoting transparency, openness, and reproducibility (Nosek et al., 2015) By way of incentives, they award “badges” to articles for preregistering a research plan to ward off accusations of *p*-hacking (Simonsohn, Simmons, & Nelson, 2015) and HARKing (Kerr, 1998), as well as for making available on the platform full sets of data and the code used. Some have compared this form of accreditation to the LEED certification for environmentally designed buildings. Finally, it is therefore not surprising that many of these ideas are at the core of two major funding initiatives, Next Generation Social Science, and Ground Truth, at the US Defense Advanced Research Project Agency (DARPA) (Rogers, 2017), and The Future of Work at the Human Technology Frontier, one of the 10 “Big Ideas” initiatives at the US National Science Foundation (NSF).

Acknowledgments

Preparation of this chapter was supported by the following research grants: US Army Research Laboratory (W911NF-09-2-0053), National Science Foundation (IIS-1514427), DARPA (HR001118C0022), and NASA (NNX15AM32G). My thanks to Yessica Herrera for providing me a digital recording of my presentation on this topic at the Conference on Complex Systems 2017 in Cancun, Mexico.

References

Ancona, D. G., & Caldwell, D. F. (1992). Bridging the boundary: External activity and performance in organizational teams. *Administrative Science Quarterly*, *37*(4), 634. **http://doi.org/10.2307/2393475**

Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature Communications*, *8*, 14753. **http://doi.org/10.1038/ncomms14753**

Asencio, R., Huang, Y., DeChurch, L. A., Contractor, N. S., Sawant, A., & Murase, T. (2015). *The MyDreamTeam builder: A recommender system for assembling & enabling effective teams*. Paper presented at the Interdisciplinary Network for Group Research, Pittsburgh, PA.

Benefield, G. A., & Cuihua Shen. (2018). Gender and networks in virtual worlds. In B. Foucault Welles & S. Gonzales-Bailon (Eds.), *Handbook of communication in the networked age*. Oxford: Oxford University Press.

(p. 125) Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., et al. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10. **http://doi.org/10.1038/s41562-017-0189-z**

Benkler, Y., Shaw, A., & Hill, B. M. (2015). Peer production: A form of collective intelligence. In Malone, T. W. & Bernstein, M. S. (Eds.) *Handbook of Collective Intelligence* (pp. 175–204). Cambridge, MA: MIT Press.

Berge, C. (1979). *Graphs and hypergraphs*. Amsterdam: North Holland Publishing Company.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(4–5), 993–1022. **http://doi.org/10.1162/jmlr.2003.3.4-5.993**

Brandes, U., Lerner, J., & Snijders, T. A. B. (2009). *Networks evolving step by step: Statistical analysis of dyadic event data* (pp. 200–205). Paper presented at the International Conference on Advances in Social Network Analysis and Mining (ASONAM).

Burt, R. S. (2009). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, *38*(1), 155–200. **http://doi.org/10.1111/j.1467-9531.2008.00203.x**

Carley, K. M. (2009). Computational modeling for reasoning about the social behavior of humans. *Computational & Mathematical Organization Theory*, *15*(1), 47–59.

Carley, K. M., & Hirshman, B. (2011). Agent based model. In G. A. Barnett (Ed.), *Encyclopedia of social networks* (Vol. I, pp. 11–16). Thousand Oaks, CA: SAGE Publications.

Carter, D. R., Asencio, R., Wax, A., DeChurch, L. A., & Contractor, N. S. (2015). Little teams, big data: Big data provides new opportunities for teams theory. *Industrial and Organizational Psychology*, *8*(04), 550–555. **http://doi.org/10.1017/iop.2015.82**

Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.-F., & Vespignani, A. (2010). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS One*, *5*(7), e11596–9. **http://doi.org/10.1371/journal.pone.0011596**

Contractor, N. (2013). Some assembly required: Leveraging Web science to understand and enable team assembly. *Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences*, *371*(1987), 20120385–20120385. **http://doi.org/10.1098/rsta.2012.0385&domain=pdf&date_stamp**

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Dickersin, K., Chan, S. S., Chalmersx, T. C., Sacks, H. S., & Smith, H. Jr. (1987). Publication bias and clinical trials. *Controlled Clinical Trials*, *8*(4), 343–353.

Dreyfuss, E. (2018, September 23). *A bot panic hits Amazon's Mechanical Turk*. Retrieved from **https://www.wired.com/story/amazon-mechanical-turk-bot-panic/**

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology.*, *38*(1), 188–230.

Eagle, N., Macy, M., & Claxton, R. (2010). Network diversity and economic development. *Science*, *328*(5981), 1029–1031. **http://doi.org/10.1126/science.1186605**

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. **http://doi.org/10.1177/001872675400700202**

Foucault Welles, B., & Contractor, N. S. (2015). Individual motivations and network effects: A multi-level analysis of the structure of online social relationship. *Annals of the American Academy of Political and Social Science*, 659(1), 180–190. **http://doi.org/10.1177/0002716214565755**

Foucault Welles, B., Welles, B. F., Vashevko, A., Vashevko, A., Bennett, N., Bennett, N., et al. (2014). Dynamic models of communication in an online friendship network. *Communi-*

cation Methods and Measures, *8*(4), 223–243. **http://doi.org/10.1080/19312458.2014.967843**

(p. 126) Freelon, D. (2018). Computational Research in the Post-API Age, Political Communication, doi: 10.1080/10584609.2018.1477506

Ghasemian, F., Ghasemian, F., Zamanifar, K., Zamanifar, K., & Ghasem-Aghaee, N. (2017). Composing scientific collaborations based on scholars' rank in hypergraph. *Information Systems Frontiers: A Journal of Research and Innovation*, *36*(1), 66. **http://doi.org/10.1007/s10796-017-9773-z**

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*(1), 129–152. **http://doi.org/10.1146/annurev-soc-071913-043145**

Gonzalez-Bailon, S. (2017). *Decoding the social world*. Cambridge, MA: MIT Press.

Gonzalez-Bailon, S., & Paltoglou, G. (2015). Signals of public opinion in online communication: A comparison of methods and data sources. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 95–107. **http://doi.org/10.1177/0002716215569192**

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, *10*(4), 371–388. **http://doi.org/10.1037/1082-989X.10.4.371**

Halas, M. (2011). Abductive reasoning as the logic of agent-based modelling. In T. Burczynski, J. Kolodziej, A. Byrski, & M. Carvalho (Eds.), *Proceedings of the 25th European Conference on Modelling and Simulation,* (pp. 105–111). Digital Library of the European Council for Modelling and Simulation. **http://doi.org/10.7148/2011-0105-0111**

Hendler, J. (2013). Broad data: Exploring the emerging web of data. *Big Data*, *1*(1), 18–20. **http://doi.org/10.1089/big.2013.1506**

Hendler, J. A., Hall, W., & Contractor, N. (2018). Web science—now more than ever. *IEEE Computer*, *51*(6), 12–17. **http://doi.org/10.1109/MC.2018.2701637**

Henrich, Joseph, S. J. H., A. N. (2010). The weirdest people in the world? *Doi.org*, *33*(2–3), 1–75. **http://doi.org/10.1017/S0140525X0999152X**

Hill, B. M., & Shaw, A. (2018). Studying populations of online communities. In B. Foucault Welles & S. Gonzales-Bailon (Eds.), *Handbook of communication in the networked age*. Oxford: Oxford University Press.

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, *355*, 486–488.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217. **http://doi.org/10.1207/s15327957pspr0203_4**

Kim, B., Schein, A., Desmarais, B. A., & Wallach, H. (2018). The Hyperedge Event Model. *arXiv preprint arXiv:1807.08225*, *1–26*.

King, G., & Persily, N. (2018, October 8). A new model for industry-academic partnerships. Working paper. Cambridge, MA: Harvard University.

Kozlowski, S. W. J. (2015). Advancing research on team process dynamics. *Organizational Psychology Review*, *5*(4), 270–299. **http://doi.org/10.1177/2041386614533586**

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Association for Psychological Science*, *7*(3), 77–124.

Krippendorff, K. (1980). *Content analysis* (Vol. 7, pp. l–84). Beverly Hills, CA: Sage Publications.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Life in the network: The coming age of computational social science. *NIH Public Access*, *323*(5915), 721–723. **http://doi.org/10.1126/science.1167742**

Le, Q., & Mikolov, T. (2014). *Distributed representations of sentences and documents* (pp. 1188–1196). Paper presented at the ICML ‘14: Proceedings of the 31st International Conference on Machine Learning.

(p. 127) Leenders, R. T. A. J., Contractor, N. S., & DeChurch, L. A. (2016). Once upon a time: Understanding team processes as relational event networks. *Organizational Psychology Review*, *6*(1), 92–115. **http://doi.org/10.1177/2041386615578312**

Leonardi, P., & Contractor, N. S. (2018). Better People Analytics: Measure who they know, not just who they are. *Harvard Business Review*, *96*(6), 70–81.

Lickel, B., Hamilton, D. L., & Sherman, S. J. (2016). Elements of a lay theory of groups: Types of groups, relational styles, and the perception of group entitativity. *Personality and Social Psychology Review*, *5*(2), 129–140. **http://doi.org/10.1207/S15327957PSPR0502_4**

Lin, M., Lucas, H. C. Jr., & Shmueli, G. (2013). Research commentary—too big to fail: Large samples and the p-value problem. *Information Systems Research*, *24*(4), 906–917.

Lungeanu, A., Carter, D. R., DeChurch, L. A., & Contractor, N. S. (2018). How team interlock ecosystems shape the assembly of scientific teams: A hypergraph approach. *Communication Methods and Measures*, *12*(2–3), 174–198. **http://doi.org/10.1080/19312458.2018.1430756**

Marrone, J. A. (2010). Team boundary spanning: A multilevel review of past research and proposals for the future. *Journal of Management*, *36*(4), 911–940. **http://doi.org/10.1177/0149206309353945**

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. **http://doi.org/10.3758/s13428-011-0124-6**

Mathieu, J. E. (2016). The problem with [in] management theory. *Journal of Organizational Behavior*, *37*(8), 1132–1141.

Mathur, S., Poole, M. S., Peña-Mora, F., Hasegawa-Johnson, M., & Contractor, N. (2012). Detecting interaction links in a collaborating group using manually annotated data. *Social Networks*, *34*(4), 515–526. **http://doi.org/10.1016/j.socnet.2012.04.002**

Meyer, S. B., & Lunnay, B. (2013). The application of abductive and retroductive inference for the design and analysis of theory-driven sociological research. *Sociological Research Online*, *18*(1), 1–11. **http://doi.org/10.5153/sro.2819**

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–119). MIT Press.

Mukherjee, S., Huang, Y., Neidhardt, J., Uzzi, B., & Contractor, N. (2018). Prior shared success predicts victory in team competitions. *Nature Human Behaviour*, *11*, 1. **http://doi.org/10.1038/s41562-018-0460-y**

Monge, P. R. (1982). Systems theory and research in the study of organizational communication: The correspondence problem. *Human Communication Research*, *8*(3), 245–261.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Scientific standards: Promoting an open research culture. *Science*, *348*(6242), 1422–1425.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (**www.LIWC.net**).

Pentland, A. S. (2012, April). The new science of building great teams. *Harvard Business Review*, 60–70.

Pew, R. W., & Mavor, A. S. (1998). *Modeling human and organizational behavior*. Washington, DC: National Academies Press. **http://doi.org/10.17226/6173**

Pilny, A., Schecter, A., Poole, M. S., & Contractor, N. (2016). An illustration of the relational event model to analyze group interaction processes. *Group Dynamics: Theory, Research, and Practice*, *20*(3), 181–195. **http://doi.org/10.1037/gdn0000042**

Radford, J., Pilny, A., Reichelmann, A., Keegan, B., Welles, B. F., Hoye, J., et al. (2016). Volunteer science. *Social Psychology Quarterly*, *79*(4), 376–396. **http://doi.org/10.1177/0190272516675866**

(p. 128) Reeves, B., Malone, T. W., & O'Driscoll, T. (2008). Leadership's online labs. *Harvard Business Review*, *86*(5), 58–66.

Ren, Y., Cedeno-Mieles, V., Hu, Z., Deng, X., Adiga, A., Barrett, C. L., et al. (2018). *Generative modeling of human behavior and social interactions using abductive analysis* (pp. 1–8). Paper presented at the International Conference on Advances in Social Network Analysis and Mining (ASONAM).

Rogers, A. (2017, July 30). *Darpa wants to build a BS detector for science*. Retrieved from **https://www.wired.com/story/darpa-bs-detector-science/**

Salganik, M. J. (2017). Bit by bit: Social research in the digital age. Princeton, NJ: Princeton University Press.

Schecter, A., Pilny, A., Leung, A., Poole, M. S., & Contractor, N. (2017). Step by step: Capturing the dynamics of work team process through relational event sequences. *Journal of Organizational Behavior*, *21*(1), 1–19. **http://doi.org/10.1002/job.2247**

Schmälzle, R., Brook O'Donnell, M., Garcia, J. O., Cascio, C. N., Bayer, J., Bassett, D. S., et al. (2017). Brain connectivity dynamics during social interaction reflect social network structure. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), 5153–5158. **http://doi.org/10.1073/pnas.1616130114**

Schwarz, G., & Stensaker, I. (2014). Time to take off the theoretical straightjacket and (re-)introduce phenomenon-driven research. *The Journal of Applied Behavioral Science*, *50*(4), 478–501. **http://doi.org/10.1177/0021886314549919**

Semuels, A. (2018, September 23). *The Internet is enabling a new kind of poorly paid hell*. Retrieved from **https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/**

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a reply to Ulrich and Miller. *Journal of Experimental Psychology. General*, *144*(6), 1146–1152. **http://doi.org/10.1037/xge0000104**

Snijders, T. A. B. (2005). Models for longitudinal network data. *Models and Methods in Social Network Analysis*, *1*, 215–247.

Stonedahl, F., & Wilensky, U. (2010). *Finding forms of flocking: Evolutionary search in ABM parameter-spaces* (pp. 61–75). Paper presented at the International Workshop on Multi-Agent Systems and Agent-Based Simulation, Springer.

Sullivan, S. D., Lungeanu, A., DeChurch, L. A., & Contractor, N. S. (2015). Space, time, and the development of shared leadership networks in multiteam systems. *Network Science*, *3*(01), 124–155. **http://doi.org/10.1017/nws.2015.7**

Taramasco, C., Cointet, J.-P., & Roth, C. (2010). Academic team formation as evolving hypergraphs. *Scientometrics*, *85*(3), 721–740. **http://doi.org/10.1007/s11192-010-0226-4**

Thiele, J. C., Kurth, W., & Grimm, V. (2014). Facilitating parameter estimation and sensitivity analysis of agent-based models—a cookbook using NetLogo and “R”. *Journal of Artificial Societies and Social Simulation*, 17(3), 11 **http://doi.org/10.18564/jasss.2503**

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.

Valentine, M. A., Retelny, D., To, A., Rahmati, N., Doshi, T., & Bernstein, M. S. (2017). *Flash organizations* (pp. 3523–3537). Paper presented at the 2017 CHI Conference, New York, New York. **http://doi.org/10.1145/3025453.3025811**

Varda, D. M., Forgette, R., Banks, D., & Contractor, N. (2008). Social network methodology in the study of disasters: Issues and insights prompted by post-Katrina research. *Population Research and Policy Review*, *28*(1), 11–29. **http://doi.org/10.1007/s11113-008-9110-9**

(p. 129) Vellido, A., Martin-Guerrero, J. D., & Lisboa, P. J. (2012). *Making machine learning models interpretable*. (Vol. 12, pp. 163–172). Paper presented at the ESANN, Citeseer.

Wang, T., Rudin, C., Doshi-Velez, F., Liu, Y., Klampfl, E., & MacNeille, P. (2017). A bayesian framework for learning rule sets for interpretable classification. *The Journal of Machine Learning Research*, *18*(1), 2357–2393.

Watts, D. J. (2017). Should social science be more solution-oriented? *Nature Human Behaviour*, *1*(1), 1–5. **http://doi.org/10.1038/s41562-016-0015**

Wax, A., DeChurch, L. A., & Contractor, N. S. (2017). Self-organizing into winning teams: Understanding the mechanisms that drive successful collaborations. *Small Group Research*, *48*(6), 665–718. **http://doi.org/10.1177/1046496417724209**

Weber, M. S. (2018). The new dynamics of organizational change. In B. Foucault Welles & S. Gonzales-Bailon (Eds.), *Handbook of communication in the networked age*. Oxford: Oxford University Press.

Wilensky, U. (2005). *NetLogo preferential attachment model*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, Illinois. Retrieved from **http://Ccl.Northwestern.Edu/Netlogo/Models/PreferentialAttachment**.

Williams, D., Contractor, N., Poole, M. S., Srivastava, J., & Cai, D. (2011). The virtual worlds exploratorium: Using large-scale data and computational techniques for communication research. *Communication Methods and Measures*, *5*(2), 163–180. **http://doi.org/10.1080/19312458.2011.568373**

Zhou, S., Valentine, M., & Bernstein, M. S. (2018). *In search of the dream team* (pp. 1–13). Paper presented at the 2018 CHI Conference, New York, New York. **http://doi.org/10.1145/3173574.3173682**

Zimmer, B. (2011, October 30). Twitterology: A new science? *New York Times*, p. SR9. Retrieved from **https://www.nytimes.com/2011/10/30/opinion/sunday/twitterology-a-new-science.html**

Noshir Contractor

Noshir Contractor is the Jane S. & William J. White Professor of Behavioral Sciences in the McCormick School of Engineering & Applied Science, the School of Communication and the Kellogg School of Management at Northwestern University, Evanston, IL.