

What Communication Can Contribute to Data Studies: Three Lenses on Communication and Data

ANDREW SCHROCK
Chapman University, USA

We are awash in predictions about our data-driven future. Enthusiasts believe big data imposes new ways of knowing, while critics worry it will enable powerful regimes of institutional control. This debate has been of keen interest to communication scholars. To encourage conceptual clarity, this article draws on communication scholarship to suggest three lenses for data epistemologies. I review the common social scientific perspective of *communication as data*. A *data as discourse* lens interrogates the meanings that data carries. *Communication around data* describes moments where data are constructed. By employing multiple perspectives, we might understand how data operate as a complex structure of dominance.

Keywords: data, big data, open data, communication theory, computer-mediated communication

The past few years have seen increasing recognition that a systematic shift toward data has serious consequences for communication. Rob Kitchin (2014) defined the common understanding of data as “the raw material produced by abstracting the world into categories, measures, and other representative forms . . . that constitute the building blocks from which information and knowledge are created” (p. 1). Christine Borgman (2015) described an inclusive definition of data as “representations of observations, objects, or other entities used as evidence of phenomena” (p. 28). Both authors stress that data is a fluid concept that defies a universal understanding. Still, they share an understanding that data are abstractions or representations captured in digital format that are then repurposed.

The question of epistemologies has figured prominently in critiques of data, particularly big data. Backers of big data believe it will provide new ways to cure diseases and create new markets. Kitchin (2014) suggested that big data was ushering in “an entirely new epistemological approach” (p. 2). boyd and Crawford (2012) criticized big data as promising a totalizing vision of knowledge production and scientific control. Other scholars, such as Andrejevic (2014), worry that data exacerbate digital divides. New players such as data brokers collect, aggregate, triangulate, and package data that are used for targeted advertising. Individuals are rarely able to obtain and interpret data that they create. This epistemological challenge is severe for communication scholars as well, as online platforms continue to restrict data collection through application programming interfaces, even as digital tools to analyze data proliferate. This article responds to what Crawford, Miltner, and Gray (2014) describe as a need to

"collectively invest in an explicit epistemological pluralism" (p. 1669). But I wish to pull communication as a discipline a little closer to discuss how communication may lead in the area of data studies.

This commentary article approaches communication as an idea (Beniger, 1993; Peters, 2001) that cuts across different paradigms (Potter, Cooper, & Dupagne, 1993). As Calhoun (2011) noted, our field's heterogeneity requires "the production of more and better connections among different lines of work" (p. 1495). Accordingly, this article draws from an interdisciplinary history of communication that includes cultural studies, sociology, psychology, and media studies. Data complicate traditional models of human communication, such as encoding-decoding (S. Hall, 2006) and the Lasswell formula. Data infrastructures govern control—for example, who sees which messages on social network sites. Yet restricting discussion to a singular model might be constrictive, considering the media environments in which data operates are still emerging. Different actors and systems likely conceptualize and use data quite differently. Therefore, I am more concerned with conceptualizing ways communication scholars might think about data in their research.

For this task I draw on what Eviator Zerubavel (1999) termed "cognitive lenses"—hermeneutics for thinking about how communication scholars research data. The power of data does not reside solely in a singular moment of digitization, transfer, mutation, or interpretation. Rather, data-driven systems operate as what Stuart Hall (2006) termed a "complex structure of dominance" (p. 128). Meanings are encoded, circulated, and decoded through various social practices. It is only by employing multiple lenses at different moments in a data-driven system that we might more fully understand the relationship between data and power.

Debates about data in various formulations (big data, open data) have provoked timely, lively, and morally necessary discussions in communication-oriented publications (e.g., Bowker, 2014; boyd & Crawford, 2012; Crawford et al., 2014; Driscoll & Walker, 2014). These discussions tend to take stances on epistemology: how data affect production of knowledge and the beliefs that scientific and lay communities alike place in data. Given that this epistemological debate is still unfolding and involves a copious number of articles, this short commentary admittedly sacrifices completeness for brevity. Drawing on the question of epistemologies, it explores the following questions: How might communication scholars conceptualize the relationship between data and power? What are potential sites of research and intervention? In response, I outline three lenses that researchers can employ when researching data and communication that describe how we might approach this emerging research area.

Communication as Data

Communication scholars, unless their work is purely theoretical, research communication by collecting and interpreting data. It would be impossible to perform empirical research without data of some kind. A perspective of *communication as data* needs the least introduction, because it captures a commonsense quantitative understanding in social science as a whole. Variables residing at specific levels of analysis can be captured with particular instrumentation. Communication phenomena tend to reside between self, other, and environment. *Process* is often identified as what makes communication distinct

from other social sciences (Monge, Farace, Eisenberg, Miller, & White, 1984). Communication scholars are attuned to locating and describing latent constructs whose evanescence presents a challenge to research.

Communication as a discipline has certainly not been immune to searches for data purity. In 1970, Klaus Krippendorff, drawing on psychologist Coombs (1960), argued that data should contain sufficient features to justify processes under observation as communication—essentially a full cycle of observation, interpretation, and behavior change. He made a formalist argument that “what is needed is an extensive development of computational techniques for processing communication data” (Krippendorff, 1970, p. 267). Otherwise communication scholars might continue to rely on intuition and subjective beliefs. Communication would always be in the eye of the beholder. Some decades before the pitched fervor of big data, Krippendorff was interested in leveraging multiple forms of data to more fully capture data that objectively defined communication.

Although provocative in its time, communication scholars have not generally adopted a perspective where definitions of communication entirely emerge from data. There are several reasons to regard communication scholarship as benefiting from the interaction between reflexive researchers, methodological approaches, and theoretical traditions. When the communicative imagination is put in conversation with the literature, researchers create instrumentation to describe and capture what can only initially be intuited (Chaffee, 1991). Instinct working together with new instrumentation is nothing new. Analyzing incidents of violence on television gave us cultivation theory and “mean world syndrome” (Gerbner & Gross, 1976). Public opinion surveys encouraged Elihu Katz and his team to theorize how opinion leaders relay information to a wider public—the “two-step” flow (Katz, 1957). New ways to collect data enable new theories to develop in a synergistic cycle.

Another reason for social scholars to be concerned with data is that, as more communication is mediated, social media platforms increasingly shape communication. Facebook does not neutrally relay messages. It collects, organizes, and relays posts and advertisements based on internal analytics that maximize engagement and positive responses to advertisements. Collecting data from platforms that act as unruly mediators places scholars in a difficult predicament. Kevin Driscoll and Shawn Walker (2014) argued that data derived from a common platform such as Twitter are unstable even across different collection methods. Social media platforms are not just where communication happens—they shape what we might know of communication.

As early as the 1950s, Herbert Blumer (1956) suggested that variables were constructed based on instrumentation, doctrine, or ingenuity. These examples demonstrate communication’s history of cautious empiricism. Scholars have intuited and developed theories supported by data that captured communication patterns and processes. Communication has rarely regarded data as empirically pure—to us, “raw data” has always been an “oxymoron” (Bowker, 2014; Gitelman, 2013). Still, large data sets may indeed lead to new understandings of how media enter media ecosystems, are remixed, and serve political purposes (e.g., Freelon, 2015). We would gain little from declaring certain areas methodologically off-limits since data can be shaped to suit certain whims. Simply using large data sets should not be stigmatized as blind behaviorism that critics often describe as harkening the age of big data.

Data as Discourse

Communication scholars have often found themselves confronted with interpreting previously unrecognized and novel forms of communication. Nonverbal communication escaped widespread recognition until Edward Hall (1959) found that cultural norms dictated physical closeness. Communication seeps into the pores of society, often remaining hidden until revealed. The steady evolution of media ecosystems and technologies has led to perennial debates about what constitutes communication. Data is intimately involved with mediated communication, seeping into emergent control mechanisms and surveillance. Kitchin (2014) and Borgman (2015) suggest that data can be thought of as what Stuart Hall (2006) called a "form." In other words, the "content" of data has communicative characteristics, even if the receiver is more likely to be a computer than a human, and there is less semiotic slippage between sender and receiver than in other forms.

Advocates of critical code studies have long suggested that code can be analyzed as cultural "texts," much like other media. Mark Marino (2006), for example, regarded code as "a sign system with its own rhetoric, as verbal communication that possesses significance in excess of its functional utility" (para. 19). To Marino there were alternate meanings that are inscribed in data beyond those that are strictly interpreted. Less poetically, we can simply consider that data sets are crafted with particular audiences in mind. A nonprofit organization tasked to collect and aggregate data on community health will select different variables than an online platform interested in leveraging behavioral traces to extract advertising revenue. Data are inherently relational; thus, ongoing frictions around the ease with which identifying information can be extracted from data. Triangulation—leveraging one data set with another—can reveal more than just subtext. For example, nearly 90% of individuals could be reidentified using four spatiotemporal data points in a study of 1.1 million anonymized credit card transactions (de Montjoye, Radaelli, Singh, & Pentland, 2015).

Communication practitioners and scholars are often in the position of translating data for dominant and alternate meanings. Data journalists unearth stories in data that connect readers to timely issues of public importance. Digital ethnographer Stuart Geiger argued that Wikipedians used trace metadata—codes that represent when and how a Wikipedia entry was edited—as a form of communication (Geiger & Ribes, 2011). Geiger and Ribes argued that these traces have emic meaning. Although illegible to the untrained eye, they were the "primary mechanism in which users themselves know their distributed communities and act within them" (Geiger & Ribes, 2011, p. 1). John Cheney-Lippold (2016) took a critical perspective to argue that the National Security Agency (NSA) used communications data to determine degree of "foreignness," which was used to classify users as citizens or noncitizens.

Communication Around Data

Big data often elicits a staunchly cybernetic perspective. Langlois, Redden, and Elmer (2015) noted that, in the context of social media, "information is taken from individuals, processed through black-boxed algorithms that produce a certain kind of knowledge, and then some kind of solution is given" (p. 8). Left to their own devices, assemblages of data and algorithms may perpetuate disparities more handily than they change them. Simplistic demands for "algorithmic audits" and opening the "black box" of

technology, although popular, can prohibit more systematic understandings (Ananny & Crawford, 2016). A *communication around data* perspective looks not at the empirical value of data nor at data on its own. Rather, it approaches data as socially constructed to understand how meanings and values become embedded in it.

Data is what Giddens terms a “double hermeneutic” (Giddens, 1990, p. 15), with both scientific and extrascientific meanings. Nick Seaver (2013) argued that data and code cannot be considered a ground truth. He pushed scholars to “examine the logic that guides the hands . . . choosing particular representations of data, and translating ideas into code” (Seaver, 2013, p. 10). Software engineers and data scientists develop their own ideas about what data does and how it should be used. Even a cursory glance at citizen scientists, informational activists, and open-source data enthusiasts reveals a wide variety of beliefs about how to create and manipulate data (Schrock, forthcoming). Communication scholars might research how data workers develop beliefs about appropriate uses of data through the media, upbringing, education, and work environment. Their data ideologies may not mirror consensus perspectives on big data or open data. Charting beliefs about and practices with data can also help identify sites for intervention, because data intermediaries such as hackers, journalists, and activists can be particularly powerful players in a data ecosystem (Magalhaes, Roseira, & Strover, 2013).

Sites for observing communication around data include communities, bureaucracies, and organizations. Ethnographic approaches to researching communication have historically proven valuable to understanding how shared understandings and practices develop. Mundane talk in work conversations and meetings reveals important facets of how technologies are created and maintained (Kunda, 2006; Orr, 1996). Work ethnographies can use corporate environments as a way to interrogate power. For example, Alex Fattal (2012) argued that Facebook’s corporate headquarters reveals how the company conceptualizes data as a form of soft power aligning with expansionist goals. Ethnography is not the one methodology to rule them all, but it can assist our understanding of power in systems that are intentionally closed to other ways of knowing. We should seek ways individuals can participate, collaborate, and rectify unjust power distributions (Couldry & Powell, 2014).

Another route for communication scholars is researching how data are defined across various communities of practice (Puschmann & Burgess, 2014). We mostly know about how people come to think about the harms and benefits of data in moments of scandal. For example, the murky world of data infrastructures was illuminated by scandal when Edward Snowden revealed the extent of NSA surveillance. In response, many augmented their Internet use habits and started to associate data with harmful surveillance (Horrigan & Rainie, 2015). At other pivotal moments, data are rhetorically positioned as beneficial to civic life. For example, young, tech-friendly government officials espouse a collaborative notion of “open data” to garner public support for initiatives (Baack, 2015). Much like any technology, data is not just an object—it is a container for ideas.

Conclusion

Even a cursory glimpse into communication’s interdisciplinary history reveals that concerns about data go back more than 50 years. Communication journals have recently been the site of an

interdisciplinary debate about the politics of data. They are continuing a vibrant tradition of communication researchers grappling with changing media ecosystems, emerging methodologies, and turbulent political situations. However, it has been understated how data intersects with epistemological concerns of communication researchers. To provide clarity, this commentary suggests three productive lenses: communication as data, data as discourse, and communication around data. A robust agenda fitting with the discipline should employ multiple lenses to understand the impact of data in media ecosystems on human experience. Much has been made of the overambitious beliefs people place in big data. But we should not dismiss emergent quantitative methodologies as the devil's handmaiden. Overtheorizing, too, has clear limitations. Similar to other media such as newspapers and television, communication does not demand a unified theory of data. This extreme would sacrifice communication's rich history on the altar of perpetual newness. The option this commentary presents is to continue the discipline's tradition of aware empiricism by drawing on its ability to synthesize research and adapt to new terrains.

References

- Ananny, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*.
- Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, 8, 1673–1689.
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2).
doi:10.1177/2053951715594634
- Beniger, J. R. (1993). Communication—Embrace the subject, not the field. *Journal of Communication*, 43(3), 18–25.
- Blumer, H. (1956). Sociological analysis and the "variable." *American Sociological Review*, 21(6), 683–690.
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.
- Bowker, G. C. (2014). The theory/data thing. *International Journal of Communication*, 8, 1795–1799.
- boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679.
- Calhoun, C. (2011). Communication as a social science. *International Journal of Communication*, 5, 1479–1496.
- Chaffee, S. H. (1991). *Explication*. Newbury Park, CA: SAGE Publications.

- Cheney-Lippold, J. (2016). Jus algoritmi: How the National Security Agency remade citizenship. *International Journal of Communication, 10*, 1721–1742.
- Coombs, C. H. (1960). A theory of data. *Psychological Review, 67*(3), 143–159.
- Couldry, N., & Powell, A. (2014). Big data from the bottom up. *Big Data & Society, 1*(2), 1–5.
- Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication, 8*, 1663–1672.
- de Montjoye, Y. A., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science, 347*(6221), 536–539.
- Driscoll, K., & Walker, S. (2014). Working within a black box: Transparency in the collection and production of big Twitter data. *International Journal of Communication, 8*, 1663–1672.
- Fattal, A. (2012). Facebook: Corporate hackers, a billion users, and the geo-politics of the “social graph.” *Anthropological Quarterly, 85*(3), 927–955.
- Freelon, D. (2015). On the cutting edge of big data: Digital politics research in the social computing literature. In S. Coleman & D. Freelon (Eds.), *Handbook of digital politics* (pp. 451–472). Northampton, MA: Edward Elgar.
- Geiger, R. S., & Ribes, D. (2011, January). *Trace ethnography: Following coordination through documentary practices*. Paper presented at the 44th annual Hawaii International Conference on System Sciences, Kauai, HI.
- Gerbner, G., & Gross, L. (1976). Living with television: The violence profile. *Journal of Communication, 26*, 172–199.
- Giddens, A. (1990). *The consequences of modernity*. Stanford, CA: Stanford University Press.
- Gitelman, L. (2013). *Raw data is an oxymoron*. Cambridge, MA: MIT Press.
- Hall, E. T. (1959). *The hidden dimension*. Garden City, NY: Doubleday.
- Hall, S. (1980). Encoding/decoding. In S. Hall, D. Hobson, A. Lowe & P. Lewis (Eds.), *Culture, Media, Language* (pp. 128–137). New York, NY: Routledge.
- Horrigan, J. B., & Rainie, L. (2015). *Americans’ views on open government data*. Washington, DC: Pew Research Center. Retrieved from <http://www.pewinternet.org/2015/04/21/open-government-data/>

- Katz, E. (1957). The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly*, 21(1), 61–78.
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1–12.
- Krippendorff, K. (1970). On generating data in communication research. *Journal of Communication*, 20, 241–269.
- Kunda, G. (2006). *Engineering culture: Control and commitment in a high-tech corporation*. Philadelphia, PA: Temple University Press.
- Langlois, G., Redden, J., & Elmer, G. (2015). *Compromised data: From social media to big data*. New York, NY: Bloomsbury Academic.
- Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 243–276). New York, NY: Harper and Brothers.
- Magalhaes, G., Roseira, C., & Strover, S. (2013, October). *Open government data intermediaries*. Paper presented at the 7th International Conference on Theory and Practice of Electronic Governance, Seoul, South Korea.
- Marino, M. C. (2006, December 4). Critical code studies. *Electronic Book Review*. Retrieved from <http://www.electronicbookreview.com/thread/electropoetics/codology>
- Monge, P. R., Farace, R. V., Eisenberg, E. M., Miller, K. I., & White, L. L. (1984). The process of studying process in organizational communication. *Journal of Communication*, 34(1), 22–43.
- Orr, J. E. (1996). *Talking about machines: An ethnography of a modern job*. Ithaca, NY: ILR Press.
- Peters, J. D. (2001). *Speaking into the air: A history of the idea of communication*. Chicago, IL: University of Chicago Press.
- Potter, W. J., Cooper, R., & Dupagne, M. (1993). The three paradigms of mass media research in mainstream communication journals. *Communication Theory*, 4, 317–335.
- Puschmann, C., & Burgess, J. (2014). Metaphors of big data. *International Journal of Communication*, 8, 1690–1709.
- Schrock, A. (forthcoming). Data ideologies of an interested public: A study of grassroots open government data intermediaries. *Big Data & Society*.
- Seaver, N. (2013, May). *Knowing algorithms*. Paper presented at the Media in Transition 8: Public Media, Private Media Conference, Cambridge, MA.

Zerubavel, E. (1999). *Social mindscapes: An invitation to cognitive sociology*. Cambridge, MA: Harvard University Press.