Abstract
We are awash in predictions about our data-driven future. Enthusiasts believe it will offer new ways to research behavior. Critics worry it will enable powerful regimes of institutional control. Both visions, although polar opposites, tend to downplay the importance of communication. As a result, the role of communication in human-centered data science has rarely been considered. This article fills this gap by outlining three perspectives on data that foreground communication. First, I briefly review the common social scientific perspective: “communication as data.” Next, I elaborate on two less explored perspectives. A “data as communication” perspective captures how data imperfectly carry meanings and guide action. “Communication around data” describes communication in organizational and institutional data cultures. I conclude that communication offers nuanced perspectives to inform human-centered data science. Researchers should embrace a robust agenda, particularly when researching the relationship between data and power.
Introduction
The past few years have seen an increasing recognition that data has consequences in the everyday lives of individuals. Along with this recognition has come a concern about the relationship between data and power. Power asymmetries are exacerbated through data abstraction, storage, and interpretation that reside outside of the realm of individual awareness [1]. For these reasons, data is thought by critical scholars to be a macro force directly guiding action. Data has also been argued to undergird and shape communication. For example, Rosenberg [2] argued that data precedes interpretation and argumentation. Finally, the typical critique of “big data” is that it offers a false vision of total knowledge and scientific control [3]. Data is thought to enable a new “divide” that exacerbates power differentials between those who produce it and those who benefit through its interpretation [4].

You might wonder: why does communication matter? Why should we care (as has been famously said) “who said what to whom and to what effect?” What might human-centered data science scholars learn from communication?

In this essay I argue that communication is more relevant than ever in an age of data. Communication scholars have rich and nuanced set of perspectives to deploy to investigate the relationships between data and power. Further, communication’s interdisciplinary and progressive social stance sensitizes us to the nature of data as empirically necessary, imperfect, and socially constructed. I describe three perspectives on data and communication that, although theoretically distinct, might productively blend through research and practice.

Communication as Data
Communication scholars have long researched communication by collecting data. Indeed, it would be impossible to perform empirical research without data of some kind. This perspective needs little introduction as it captures a common-sense understanding in social science as a whole.

“Communication as data” posits that variables residing at specific levels of analysis can be captured with particular instrumentation. Data captures phenomena of interest that are often not self-evident. Self-efficacy, for example, is both a latent variable and outlook held by individuals. Exactly where variables (such as, say, social capital) “reside” has been subject to vigorous debate.¹ Quantitative scholars might argue that the belief one’s actions have a positive outcome on one’s life is a way to measure self-efficacy. Because it’s relevant to individuals’ well being, they promote communication that leads to individuals having greater self-efficacy.

Communication theory has historically needed new data sets and methodologies to advance, even as data does not stand on its own. In 1970 Klaus Krippendorff [5] outlined a “theory of data.” He argued that data should contain features necessary to justify the phenomena under observation as communication. Although provocative in its time, this perspective has not been influential as other theories. Public opinion surveys, for example, allowed Elihu Katz’s and his team to theorize how opinion leaders relay information to a wider public.

¹ For this reason, to scholars with more clearly defined levels of analysis (e.g. psychology), communication is unbearably meso-level.
Meanings are encoded in, transmitted through, and decoded from [10] data. Much as with other media there is significant slippage in the meanings that are inscribed in it and those that are interpreted by receivers. Still, historically, data and the tools used to collect it have shaped social imaginaries in unintended ways. Thinking about even the mundane survey in the United States, categories for race in census forms shaped how we understood diversity. A notion of an “average American” further altered how a nation viewed itself [11]. The way data sets are structured – and what they elide or leave out – communicates and persuades. Advocates of “critical code studies” suggest that code can be analyzed as cultural “texts” much as movies or other media.

In recent years communication scholars have imported perspectives from neighbor disciplines. Latour’s material semiotics has been influential, as has computer science. Yet, communication scholars may not need adopt actor-network theory nor become computer scientists to become sensitized to data as communication. Communication’s rich history in infrastructures [12], for example, has been largely forgotten in the decades since transportation separated from communication.

Data infrastructures may not exactly be interlocutors, but they are certainly alive. The first two perspectives in this essay are necessarily intertwined. Scholars often wish to collect data from platforms that act as unruly mediators. For example, Kevin Driscoll and Shawn Walker [13] make clear that data-driven platforms such as Twitter are unstable. They require astute analysis of how they shape communication. Much to this point, Thorson and Wells [14] argue that a “media effects” of violence on television gave us cultivation theory and “mean world syndrome” [7]. In each case, scholars developed theories supported by new ways of capturing communication processes.

Communication scholars rarely adopted a feared “big data” perspective where they place wholesale faith in pure empiricism. As far back as 1956 Herbert Blumer suggested that variables were constructed based on instrumentation, doctrine, or ingenuity [8]. Craig Calhoun [9] traced communication’s interdisciplinary history, suggesting that to view communication simply as a social science would be to “reduce it in unfortunate ways.” What seems more helpful in the current moment is to recognize that communication scholars attach methodologies to measure phenomena to bolster arguments and develop theory. Here it is necessary to turn to two further perspectives that inform how scholars might approach communication and data.

**Data as Communication**

To put it coarsely, data is a language that computers "speak." Many humans are limited in this regard. Still, I argue interrogating data should not be outside the purview of communication scholars. A “data as communication” perspective regards data itself as a form of communication. Data are constructed with particular audiences in mind, but can be intercepted and re-interpreted. Data journalists unearth stories in data to connect the public with issues of importance. Data can be treated as communication that makes systems interoperable even as their operations become frustratingly opaque.

- the “two-step” flow [6]. Analyzing incidents of

- Communication scholars rarely adopted a feared “big
data” perspective where they place wholesale faith in

- Meanings are encoded in, transmitted through, and
decoded from [10] data. Much as with other media
there is significant slippage in the meanings that are
inscribed in it and those that are interpreted by
receivers. Still, historically, data and the tools used to
collect it have shaped social imaginaries in unintended
ways. Thinking about even the mundane survey in the
United States, categories for race in census forms
shaped how we understood diversity. A notion of an
“average American” further altered how a nation
viewed itself [11]. The way data sets are structured –
and what they elide or leave out – communicates and
persuades. Advocates of “critical code studies” suggest
that code can be analyzed as cultural “texts” much as
movies or other media.

In recent years communication scholars have imported
perspectives from neighbor disciplines. Latour’s
material semiotics has been influential, as has
computer science. Yet, communication scholars may
not need adopt actor-network theory nor become
computer scientists to become sensitized to data as
communication. Communication’s rich history in
infrastructures [12], for example, has been largely
forgotten in the decades since transportation separated
from communication.

Data infrastructures may not exactly be interlocutors,
but they are certainly alive. The first two perspectives
in this essay are necessarily intertwined. Scholars often
wish to collect data from platforms that act as unruly
mediators. For example, Kevin Driscoll and Shawn
Walker [13] make clear that data-driven platforms such
as Twitter are unstable. They require astute analysis of
how they shape communication. Much to this point,
Thorson and Wells [14] argue that a “media effects”
tradition needs to consider how algorithms and social media control the information people receive. Communication scholars need to be astute interpreters to connect communication theory to the current day.

**Communication Around Data**

A "communication around data" perspective refers to human communication in environments, cultures, and organizations that work with data. Nick Seaver [15] persuasively argued that data and code alone cannot be considered a ground truth. He pushed scholars to "examine the logic that guides the hands, picking certain algorithms rather than others, choosing particular representations of data, and translating ideas into code" (p. 10). Much to this point, boyd and Crawford [3] drew attention to the value of "small data" and practices of information workers. To extend Seaver, boyd, and Crawford, inquiry of beliefs and practices might occur at the individual and meso (community or organizational) levels.

On the individual level communication scholars can research how people come to share understandings about data. People might develop beliefs about appropriate uses of data through upbringing, education, and collaboration. I term these beliefs, drawing on Ilana Gershon [16], "data ideologies." In this perspective data operates as what Giddens terms a double hermeneutic. It is not only a scientific term, but various communities construct their own concepts of what data is and does. Although research in this area is still emerging, I suspect that beliefs about data are less unified than boosters and critics alike claim.

Organizations and collectives that use data reflect particular "data cultures." Here, too, communication takes center stage. For example, Gideon Kunda’s work on the culture of software developers in a large organization takes communication as a research subject and focal point for ethnography [17]. Production studies, too, unpacks how media are produced through the actions of many hands [18]. How individuals collaborate on data-driven communication technologies is also a pressing question in human-centered data science.

The now ubiquitous civic hackathon is one example of where data ideologies and cultures are made visible through communication. Government officials are keen to position hackathons as a form of civic engagement for geeks. They promise to bring people together for collective action on pressing social problems [19]. The spectacle of data-driven innovation involves emotional appeals by public officials, group work, and the inevitable final "pitches." We can see and hear the times when groups have internal disagreements or a pitch doesn’t resonate with the crowd. Slight of hand and showmanship are imported from tech world presentations [21]. Officials espouse a friendly, collaborative notion of "open data." Yet, there are times when participants don’t toe the bureaucratic line and have quite different ideas about how data might be connected with political participation.

**Conclusion**

Researching data is not simply "researching data." Data is not an inert object for communication any more than a newspaper is simply "information." Although brief, I hope this essay provokes recognition and curiosity about the methodological and theoretical contributions of communication to human-centered data science. "Communication as data" draws attention to how
communication scholars have always been cautious data producers and interpreters. "Data as communication" captures how data itself might be treated as a form of communication. Despite its relevance, "communication around data" remains less unexplored. Particular community, organizational and institutional situations produce specific data ideologies and cultures. Only by judiciously employing multiple perspectives on data might we come to a more complete understanding of how data impacts everyday life.

References


