

Five Considerations for Engaging With Big Data From a Rhetorical-Humanistic Perspective



Zoltan P. Majdik

*North Dakota State University
Fargo, ND*

Poroi 16,1 (May 2021)

Abstract: This essay offers five conceptual entry points for engaging with Big Data from a rhetorical perspective. These five concepts—data in/as relationships, observability/action, patterns, diachronicity, and audience—serve as points of deep conceptual commonality between definitions of Big Data and principles in rhetorical studies and are offered here as considerations for critiquing uses of Big Data from a rhetorical-humanistic perspective, as well as for guiding rhetorical work that uses Big Data.

Introduction

The term “Big Data” means nothing. And if it means something, its meaning is unrecognizable across the contexts in which it is used. As much as it may function as a God term everywhere from genetics to finance to what TV shows get produced, the meaning of what (as this special issue’s CFP puts it) “makes big data *big*” remains ambiguous at best. In an article discussing conceptual heuristics for Big Data in journalism, Lewis and Westlund (2015) call the concept of Big Data “plastic,” “assuming different meanings in different contexts for different purposes” (p. 448).

The statement that “Big Data” means nothing may seem like a facile point to make, more useful for circuitous discussion than productive engagement. As Boellstorff (2013) argues, there may be “no unitary phenomenon ‘big data’ [but] the impact of big data is real and worthy of sustained attention.” Still, the emptiness of the signifier matters, because the definitional debates it draws point to *stases* that can crystallize tensions and act as “process[es] of invention” (Graham & Herndl, 2011, p. 154). This is true not only in those disciplines that have leaned strongly on Big Data—bioinformatics, for example, or finance, or the emerging

computational social sciences—but also for rhetorical studies: for how we think critically about the uses of data in Big Data applications, as well as for how we, as scholars of rhetoric, use Big Data.

In this paper, I engage the definitional stasis over what makes Big Data *big* as an entry point for considering the relationship between Big Data and rhetorical studies. Broadly¹ motivated by Sullivan’s (2013) call for “reaffirming a humanistic understanding of technology,” I start with axiomatic industry definitions² by Laney (2001, in an unpublished research note discussed by Diebold, 2012) and the NIST Big Data Public Working Group Definitions and Taxonomies Subgroup (2015) to show that the *goals* of Big Data analysis are in specific definitional aspects similar to those of rhetorical practices. This similarity suggests that the study of rhetoric can offer a strong basis in which critiques and correctives of Big Data can be grounded. But I also discuss what Big Data as a *source for* rhetorical analysis—rather than as an *object of* analysis—may look like. Hence, I start with the assumption that like most technologies, Big Data is value-neutral; that only in its uses do valuable or problematic applications emerge.

I argue that definitions of Big Data yield five points of conceptual similarity between the goals of Big Data analysis and the study of rhetoric: discovering unstructured relationships; observability as an emergent property of unstructured relationships; patterns as the accretion or calcification of relationships in rhetorical ecologies; diachronicity as a property of patterns that connects the generalized to local contexts; and the implications of these concepts on “audience” as a source of rhetorical practices. Following the design of a paper by Boellstorff (2013) that seeks to “explore four conceptual interventions that can contribute to the ‘big theory’

¹ Sullivan’s guest editorial focuses on technology, not data, but his call for rhetoric to serve as a mediator between us and a technological landscape that is neither good nor bad, just foreign to our lived experience, mirrors this essay’s argument for rhetorical concepts to serve as sense-making lenses for Big Data, offering entry points for critique and for engagement.

² Throughout this paper, I weave together quotes by industry practitioners talking about how they use Big Data with the traditional peer-reviewed sources typical for academic papers. This is because, as I argue below, useful definitions of Big Data are functional, based on use-contexts (“how do we use Big Data?”), rather than quantitative (“how big is Big Data?”).

sorely needed in regard to Big Data” and Lewis and Westlund (2015), these are concepts that can serve as conceptual “entry points” (Lewis & Westlund, p. 455) through which applications of Big Data can be (critically) read. They can also serve as entry points to analysis, offering guidance for engaging in rhetorical analysis with (not only of) Big Data. In arguing that these concepts can serve as entry points for scholars of rhetoric to work with Big Data, I will not sidestep the important critical work on big data and algorithmic judgment/decision making that others undertake, including those cited in the CFP, raised in work like boyd and Crawford’s (2012) or Noble’s (2018), and explored in this special issue. But I argue that engaging with conceptual definitions about what Big Data is can help (and has helped) guide the work of rhetorical analysis *as well as* offer *topoi* for critique. As Lewis and Westlund (2015) state, this claim should not “celebrate or fetishize big data,” nor does it assume that Big Data in rhetorical studies is the same as Big Data in industry or the life sciences (p. 450), but that the “implications of big data ... require conceptual heuristics that can guide future research” (p. 451).

In the following sections, I work through these five points of conceptual overlap between definitions of Big Data and concepts in rhetorical studies. Offering these conceptual entry points as considerations for scholars who want to engage in critique *of* Big Data from a rhetorical-humanistic perspective and those whose rhetorical scholarship and analysis is guided *by* Big Data, I demonstrate that Big Data in rhetorical studies does not need to be imagined as stacks of hard drives filled with data. Defining it functionally (what does Big Data-driven inquiry *do*?) rather than quantitatively (*how much* data is in Big Data?) shows the value of rhetorical studies for the critique of uses of Big Data, as well as the potential for Big Datasets to serve as sources for rhetorical analysis.

Consideration 1: Are Data Relational, or Are Data Points Conceptualized In/As Relationships?

Quantitative definitions of Big Data (“how much storage does it take up?”)—that is, definitions centered too literally on what makes Big *big*—are one of the quickest ways to demonstrate the challenges of incommensurability (see also Graham et al. (2015), who argue that definitions of Big Data that rely on “specific storage thresholds” are “untenable” (p. 71)). Different spheres of professional practice vary too widely in what constitutes “Big” for fixed quantitative definitions to have value. In bioinformatics, “Big”

may be measured in petabytes; in finance, in terabytes; in consumer behavior, in gigabytes. One use-context's "Big" may take up more storage space because it works with very large numbers of data points (e.g., many different molecules), but in another context, fewer data points may be changing rapidly (e.g., finance), or a limited number of data points may be unstructured and unrelated to each other, adding complexity (e.g., consumer preferences). As boyd and Crawford (2012) argue, "there is little doubt that the quantities of data now available are often quite large, but that is not the defining characteristic of this new data ecosystem" (p. 663).

More useful approaches to defining Big Data depart from the purely quantitative (the "Big") and instead turn to the functional nature of "Data." For boyd and Crawford (2012), "Big Data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets" (p. 663; I will return to these definitional points in the later subsections on relationships (cross-referencing), observability (search), and patterns and audience (aggregation)). Industry users of Big Data often refer to Doug Laney's 2001 definition, which defines big data across three Vs: volume, but also velocity and variety. For data to be "Big," the dataset needs to have specific characteristics, namely, volume, but also speed at which data are generated (and with a later addition of "volatility" to the definitional scope, the speed at which data changes) and variety of sources from which data points are drawn. The three Vs also ground the National Institute for Standards and Technology's definition, which cautions that the "big" in Big Data cannot simply be reduced to volumes that are "bigger" than before, or bigger than current techniques can efficiently handle" (NIST, 2015, p. 4).

Both Laney's and the NIST's definitions of Big Data use the concept of relationships to define the term instead of relationality. Datasets are most commonly understood as relational: records "conceptualized as the rows in a table where data elements are in the cells" (NIST, 2015, p. 12) are built on inherent relations between data points in what amounts to a set of structured data.³ When data are conceptualized in and as relationships, on the other hand, then what matters, definitionally speaking, is not what aspects inherent to the dataset relate data points to each other, but what

³ Relationality is a central organizing principle in grammar, molecular geometry, or even human networks, where relations reflect natural kin, while relationships are social constructions.

relationships between characteristics external to and not inherent in the data connect individual data points. Datasets defined by relationships are often also unstructured, a term that refers to data that are not pre-linked to other elements of data. The text of a tweet, for example, could be relationally situated relative to a date, a unique ID, etc. But juxtaposed with other tweets (or images, websites, videos, unrelated textual corpora like Congressional hearings, or even other relational datasets⁴), the tweet exists in relationships to other data points that cannot be pre-structured. This latter view of data increasingly is becoming the primary focus of Big Data. As the NIST report notes, “the Big Data paradigm shift has increased the emphasis on extracting the value from unstructured or relationship data” (2015, p. 13).

The implication of data defined through relationships rather than relationality and of datasets being unstructured is twofold. One aspect is that actionable information needs to be extracted from data that are inherently contingent and uncertain. Unlike in a relational database, data in Big Data is unbounded, part of an open network that neither begins nor ends with the dataset⁵; the data always could be otherwise and do not cohere along distinct pre-structured lines. There are no necessary boundaries to the data both because of their velocity (and later—see the section on diachronicity—volatility) and variety; hence, actionable value emerges from data that are contingent on the artificial limits of the dataset(s) and are therefore necessarily incomplete and uncertain. Volume, variety, and velocity are analog measures (Watzlawick et al., 1967), offering definitional vectors with an infinite capacity for variation, for “being otherwise:” there always could be more data points, more variety of sources, more fine-grained resolution of the velocity and volatility captured by the dataset at rest. The second aspect is that by having these characteristics, a goal of Big Data applications rests on the ability of data to surprise. As one professional user of Big Data puts it, using Big Data is about “the search for surprising insights,” discoveries that are “unexpected, non-obvious” (Kobielus, 2016). This is not to say that relational data cannot yield surprising, unexpected results, but that Big Data is *defined* by its capacity to do so.

⁴ The Berkeley School of Information (2019) defines “variety” not only as variety of sources but also as variety of data structures.

⁵ See also my later discussion of how this mirrors a “rhetorical ecology.”

It should be evident at this point how much core concepts of rhetoric can serve to mediate (see Sullivan (2013), above) between the world of actual humans and the world of Big Datasets. Rhetorical practices are, after all, those communicative engagements through which judgments and decisions emerge from necessarily contingent and uncertain information. If the data in Big Data are conditional and contingent (see also the sections on patterns and diachronicity below) because the data contained in a Big Dataset could always also be otherwise, then those data are probabilistic. Such a view of data aligns with “humanistic methods [that] ‘are necessarily probabilistic rather than deterministic,’” and suggests that “each reading of a text may produce a ‘new’ text” (Drucker, 2012, qtd. in Hart, 2015, p. 154). One of the more destabilizing experiences of using Big Data is when one realizes that with some analytical techniques, like certain machine learning approaches, every “observation” of the data—every output after a new run—produces slightly different results. Data in Big Data so function similarly to a rhetorical act, which draws contingent relationships between materially dissimilar domains; a speaker using a specific metaphor like “tipping point” to direct attitudes toward climate change, for example, takes a tactile, localized, physical feeling ingrained in all humans and uses it to make sense of a global, invisible, and imperceptible climate phenomenon by fixing the semiotic relationship between these two materially distinct domains (see e.g., van der Hel et al., 2018). These relationships always could be otherwise (Rapp, 2010, paraphrasing Aristotle). Other tropes, other figures, different ways of arranging the information would have been available but for the rhetor’s judgment to use the rhetorical construct she used. The “data” of rhetorical practices, in other words, is unstructured and unbounded: it always could be otherwise, existing in ecological relationships (Edbauer, 2005) unbounded by firmly structured situational constraints and relations, and instead “infected” by “things that have absolutely nothing to do with each other” (p. 15). Rhetoric’s artistry comes from drawing relationships between non-inherently linked concepts—it lies in its ability (to reprise a term from the preceding paragraph) to surprise.

Hence, both in functional definitions of Big Data and in the relationships that characterize the functions of rhetorical practices, insights are designed to emerge from relationships in ways that surprise and are unexpected. This is different in systems of relationality, where “entities—individuals, actors—do not generate and create the relationship. Instead, they take their cues from some pre-existing connection that links individuals” (Whimster, 2018).

Querying an SQL database (a relational data structure) means that while I may not know the answer to my query, I know what *kind* of answer I will receive. The same is not true—cannot be true—when seeking answers in unstructured, unbounded data.⁶

My goal in this section and subsequent ones is not to overstate similarities or to suggest that because rhetorical practices and Big Data share certain functional characteristics in how they conceive of “data”, one is just like the other. But because prominent definitions of Big Data often center around the unstructured, relationship-over-relational nature of data, and because the functional goals of such a definition of data cluster around their possibilities for surprising information to emerge, rhetoric—with its long history of thinking through precisely these kinds of issues—offers a strong grounding for mediating our understanding of the uses and applications of Big Data.

Consideration 2: Do Data Create Observability (and Action)?

A function of relationships is action. Relationships link entities that lack inherent pre-defined relations; as such, relationships are *generative*, acting on their environments by drawing new linkages. In rhetoric, drawing relationships between materially distinct domains generates symbolic action because it invites situated judgments about the suitability of these new linkages and perspectives to the moment at hand. Kenneth Burke differentiates between motion and action as the capacity of language to affect attitudes, offer correctives, craft perspectives, and motivate, by selecting some perspectives—making judgments about the validity of one type of symbolic information—over others.

For many Big Data applications, the distinction between motion and action as it relates to unstructured relationships is signified in an emerging term: observability. Observability comes from the need to differentiate between monitoring, sensing, and measuring on the one hand and acting on data on the other. Barnaghi et al. (2013) note that the goal of moving from sensing through metadata and pattern recognition is “actionable knowledge and decision-support mechanisms” (p. 10) in use-contexts where large, unstructured, but interconnected datasets require monitoring (like,

⁶ This is not to say that relationality is absent in either Big Data or in rhetoric; more on that in the section on patterns.

for example, complex computing systems, where different layers of logs and error traces require constant monitoring). This, however, is not “Big Data,” even if the volume of data is very large. Only once data streams become too complex, too large, too fast, and too unordered (i.e., too varied in source and/or structure) to monitor do these use-contexts meet the definitional thresholds of Big Data. That is the point at which a shift from monitoring to finding actionable information must take place. Observability points at this distinction. An industry paper contrasts monitoring (which tells us “when something is wrong”) with observability (which tells us “why something is wrong”), suggesting that while “monitoring is about measuring what you decide in advance is important,” observability “is the ability to ask questions that you don’t know upfront”, and (to return closer to the term’s roots in engineering) “a measure of how well internal states of a system can be inferred from knowledge of its external outputs” (New Relic, 2020).

Key to these definitions is that observability is a kind of layer, a mediator between people who need to make decisions and judgments and complex technologies. In a chapter on organizational transparency, Flyverbom (2019) argues that “efforts to create observability [ask] how can we ensure that what is disclosed is the right information” (p. 86). This connection between data and human judgment is mirrored in a Big Data product description by Tanya Bragin at Elastic, for whom observability connects relationships (not relations) between logs, metrics, and traces. Monitoring complex systems or data sets can generate too much information, creating a need for the “detection of ‘interesting’ events” (Bragin, 2019, note the return of “surprising” via the term “interesting” here). Observability is less concerned with observing everything; it is, instead, concerned with locating those emergent bits of knowledge on which we can act, given the requirements of a present situation. Observability draws focus on action (what can be done) instead of motion (what is).

These definitions of observability suggest that a guiding principle in Big Data is what rhetoric calls *phronesis*. *Phronesis* mediates, argues Farrell (1995): it is “the mediation of generalizable principle with unique particulars” (p. 226; see also Sullivan, 2013), of *all* data with those aspects of data that can help make actionable judgments. Much like Burke’s symbolic acting, observability eschews the need to account for *all* information in favor of identifying *workable* information. It is selection and de-selection, the “dynamic of concealment and unconcealment,” of “*doxa* and *aletheia* [that] are different stages in the production of meaning” (Hariman, 1986, p.

50). The ability to defend our judgments and actions (the why) matters more than the ability to marshal all the information that pertains to a judgment, because *phronesis* operates in situations where there is a stronger exigency for the former than the latter. It is “prudential reason,” argues Farrell (1995), but prudential reason “in the world of public affairs, where it is not neutral at all” (p. 326), where the “urgency of the moment” invites rhetorical practices that conform to the norms required by the situation (the “virtue of propriety” (p. 236)) and generate shared judgment. Information matters insofar as it can create action; the focus is on practice and what is practical, over an exhaustive accounting of all information. Action and observability, as properties of the relationship-centric nature of both rhetorical language and data in Big Data, offer conceptual considerations through which we can make sense of, critique, and mediate between humanistic practices and the uses of Big Data.

Consideration 3: Are Patterns Sufficiently Responsive to Local Contexts?

So far, I have focused on concepts that ground practices whose characteristics are lively, dynamic, emergent: drawing unexpected relationships from contingent and unbounded data, the emergent property of knowledge from these relationships, the symbolic actions and judgments generated by them. But there exists a necessary counterpart to this liveliness, evidenced in Farrell’s quote on *phronesis* above, linking rhetoric and eloquence to the “virtue of propriety.” Sometimes, lively, dynamic practices calcify into patterns.⁷ Indeed, an interesting property of rhetorical practices is the relationship between inventiveness and patterns. This is true for the humanities more broadly (see e.g., Bod, 2013a), but in some ways it is more evident in rhetoric, where repetition (of entire words or on a figural level) remains one of its most effective strategies.

Patterns are central to Big Data analysis, where the discovery of not only existing but also emergent patterns is a primary goal. Patterns are often seen as an index of the “things that worked,” of the pre-defined dominant relations of the past. But unlike in a flat

⁷ This may be particularly when their initial uses are effective and therefore, through repetition, increase in prevalence over time; I have argued elsewhere for such a pattern/prevalence-based definition of “effective.” See Majdik, 2019.

relational system (where such patterns are, correspondingly, flat, i.e., part of pre-defined relations), in unstructured systems of relationships, patterns are multidimensional, subject to change from the push and pull of different sources. These patterns are often challenging to see because they are hidden, particularly when data are varied (as they need to be to meet the definitional scope of Big Data) or when patterns are hidden behind preconceived notions or pre-determined relations. Nowhere is this more evident than in deep learning/neural network uses of Big Data, which start with the assumption that a “hidden layer” (Montañez, 2016) stands between the input of our human perceptions and an output that reflects more nuanced patterns. A Big Data application that seeks to discern how we balance risk to self and responsibility to others in today’s time of COVID-19 could, for example, draw on tweets, interviews, visual memes, and letters to editors simultaneously, where volume, velocity of change, and variety of sources would challenge any patterns that are reducible to pre-defined relations. Patterns instead would need to emerge from the data in ways that “surprise” and reflect a certain dependency on (and, correspondingly, instability from) the varied particularities of the contexts and experiences from which they emerge. As Bod (2013b) argues, patterns “can range from the particular to the universal,” they can be context-, time-, or location-dependent; they are, in other words, compatible with the work of the humanities (p. 173).⁸

For Bod (2018), the act of compiling patterns historically has served to discern something original, to allow a kind of knowledge of a text to emerge from compilations, dictionaries, or patterns whose constituent parts were often not directly related (in the strict sense of relationality described above). These patterns serve as an index of a kind of sum-total of lived human experiences, yet simultaneously are removed from them. As Bod (1998) argues

⁸ In a later essay, he takes this argument further, contending that the “opposing scholarly practices [of] searching for patterns versus understanding the unique” historically cannot be attributed to a division between sciences and humanities (as a reductive understanding of patterns as universalizing may suggest). Instead, the division originates from “opposing approaches in the humanities” in early antiquity (Bod, 2018, p. 15). As an example, he cites Aristophanes of Samothrace, who sought to recreate the original Homeric texts using analogy as a device for distinguishing “original” words from errors (p. 17). Analogy, of course, is a rhetorical device; as Bod argues, it was used by “analogists” to “search for patterns across words and texts,” in contrast to the “anomalist focus on the unique and the exceptional” (p. 18).

(summarized in a review by Manning (2002)), “human language cannot be described via a competent *grammar* [author’s emphasis], but rather should be described via ‘a statistical ensemble of language experiences’ (145) remembered by each language user—a corpus, if you will—which the language user draws on and productively recombines” (p. 441). Patterns calcify into resources, repositories. Rhetorical and communicative patterns so serve not only as an index of rhetorical acts long past but as resources *for* rhetorical practices. A metaphor like the aforementioned “tipping point” would, after its invention, effectively serve as a resource in the context of a growing need to deliberate about climate change, for example. In an earlier essay, I cite Luhmann’s (1992) use of the term “redundancy” as a kind of “memory that can be called on by many persons” (Majdik, 2019; see also Mays, 2017) as an example of how Big Data can reveal emerging patterns. There, the movement of two competing agentic frames across time represents the texture of rhetorical ecologies that “recontextualize rhetorics in their temporal, historical, and lived fluxes” (Edbauer, 2005, p. 9): as Edbauer argues, citing Shapiro, in any network, “the force of all messages, *as they accrete over time*, determines the very shape of the network” (pp. 9–10).

Hence, patterns inhabit a liminal space in humanistic approaches to inquiry, at once reflecting the practices and experiences of a people writ large while simultaneously staying connected to the shifting situations and events from which those practices and experiences emerge. Hart (2015) invokes this liminality when, in a paper advocating for computational approaches to rhetorical scholarship, he cautions that quantification of rhetorical text always must remember that “rhetoric is a text in bondage to an audience” (p. 155). This dynamic between the situatedness of rhetorical practices and their accretion into networks or ecologies over time remains a challenge for Big Data. On one hand, as I argued at the beginning of this section, definitions of Big Data push to move past patterns that are flat, static, reified relations, and that lack the ability for the kinds of new realities we see in scholarship from Bod to Luhmann to emerge. But normative definitions are not always aligned with the realities of practice, where, as Floridi (2012) argues, “the real epistemological problem with big data is *small patterns*. Precisely because so many data can now be generated and processed so quickly [...]” (p. 436). Unlike for universal patterns, the velocity of Big Data, the variety of sources from which they are derived—their time- and context-dependence—make “small patterns” both critical to the ethics of using Big Data and critical to questions of observability and action discussed above. Sentiments

like this are echoed in industry papers as well. To cite one example, Microsoft’s documentation on “Big data architecture style” references the need for detecting patterns *and* anomalies, with the latter a necessary counterpart to keep the former from becoming too detached from the situatedness of its underlying data (Microsoft, 2019).

Consideration 4: Do Data Account for Temporality?

The relationship between Big Data and patterns draws temporality into focus. Diachronic movement emerges from patterns—it cannot be visible from anything but patterns—and in so doing shows the responsiveness (or rather, the *capacity for* responsiveness) of patterns to external events. As definitions of Big Data evolved away from the original Vs, additional Vs like “value“ and “volatility” were added. “Volatility,“ defined by NIST as the “tendency for data structures to change over time” (2015, p. 7), recognizes that a definition of Big Data ought to capture diachronic perspectives beyond mere “volatility” from Laney’s original definition.

The evolution of the definitional scope from “velocity” to “volatility” is an essential step for mediating between the world of humans and their representation in Big Datasets. While “velocity” presents a challenge (how do we capture data that is generated very quickly?), “volatility” draws attention to the fact that if actionable value is to be extracted from Big Data, analyses need to recognize the changing nature of that data. “Small patterns,” in other words, a sensitivity toward the responsiveness (and fragility) of patterns to *contexts* (of time, space, place), matter not just as an ethical imperative to capture as much of the actual people that make up the Big Dataset, but also for capturing the best actionable insight.⁹

The importance of temporality is evident in Big Data’s focus on metadata. Metadata “describes additional information about the data such as how and when data was collected and how it has been processed” (NIST, 2015, p. 12). Metadata seeks to make visible and fix in place the kinds of relationships in the unstructured variety of

⁹ The same is true for mathematical models, often a product of Big Data analyses, which, to cite a recent “manifesto” in Nature, must be assessed, critiqued, and used with an eye toward how they represent “uncertainty and sensitivity” and unknowns if they are to be useful for policy (Saltelli et al., 2020).

data sources I described in the first section of this paper. Semantic metadata, for example, “represents a Big Data attempt to provide cross-cutting meanings for terms [and] is especially important for linked data¹⁰ efforts” (ibid.). Here, metadata can also serve temporal functions. It can describe “the history of a dataset,” its “provenance,” because in Big Data, “it is increasingly important to have information about how data was collected, transmitted, and processed” (ibid.).

This sense of metadata reflects historical concepts of the term “meta.” Boellstorff (2013) shows that though frequently read as a marker of hierarchical difference (“between”, “above”), the prefix *meta-* etymologically also refers to a “laterality”, a “‘before’ and ‘after’” (p. 5). Metaphors are one example of this laterality; they “carry across” meaning from one contextual space or semiotic use-context to another, unrelated one. Metadata are functionally related to metaphors in precisely this sense of laterality. There is a lateral temporality to *meta*, as it fixes meaning in context and in time. The defining use of metadata in Big Data is an attempt to fix meaning in time given the challenge presented by the unstructured, changing, and volatile nature of its data. Boellstorff (2013) cites Snowden arguing that “data collection always takes place in an interpretive frame—even one applied after the fact, so that a government could ‘go back in time and scrutinize every decision you’ve ever made’” (p. 8). Metadata tries to fix that contextual frame in place by linking the location in which a photograph was taken, the time of day a phone call was placed, etc., to the original bit of data. This practice would not be necessary in relational systems, where linkages pre-exist and everything there is to know is contained in the relational regime. But in systems of relationships, built on unstructured, unbounded, and variable data, these linkages need explication, and—should the data be useful beyond its immediate (temporally speaking) audience—preservation. Much like, for example, tropes like *meta*-phors fix volatile relationships between dissimilar meaning-contexts and make them usable over time, metadata in Big Datasets fixes relationships in contexts and in time.

Recognizing the diachronic dimension of (broadly speaking) “things” built from unstructured relationships is, as I have argued thus far, the domain of both rhetoric and the definition, if not

¹⁰ “Linked data” here refers to the “variety” aspect of Big Data definitions, where Big Data is often made up of a variety of sources from dissimilar domains.

always the practice, of Big Data. This may be particularly true in the area of rhetorics of science, technology, and medicine (see e.g., Kessler & Graham, 2018; Koteyko et al., 2013; Majdik, 2019; Schwartzman et al., 2011; Taylor, 2010), where our objects of analysis must always be understood in their “temporal unfolding” (Pickering, 1993). Temporality—the dynamics of rhetorical practices as they respond to and shape emergent events, captured in all its complexity in the concept of *kairos*—resolves the seemingly paradoxical relationship of rhetoric (and other humanistic practices) as simultaneously novel and inventive, and patterned and repetitive (Greene, 1984).¹¹ Time represents a patterning, an “accretion” of messages (Edbauer, 2005, pp. 9–10) in data assemblages volatile and varied enough to form ecologies, textual or otherwise. It also represents the potentiality of reaction in these ecologies, where patterns become exposed as temporally, contextually, and locally contingent, with new patterns emerging from old ones in response to ecological pressures, and always challenging their patternal progenitors.

Consideration 5: How Do Data Represent and Reflect an “Audience”?

In many ways, the preceding sections all work toward the centrality of “audience” (or rather, the textual manifestations of audiences) as source for rather than target of a study of rhetoric that makes space for Big Data. Big Data is *big* because it is drawn (in the humanities, behavioral, and social sciences) from groups of people the data try to reflect. The emotions, attitudes, perspectives, and beliefs of people exist in complex relationships and cannot be pre-determined by boxing parameters like psychological traits or demographics into flat data structures. Patterns discerned as representative of a group of people cannot be universalized but instead reflect the temporal and contextual contingency, dependency, and fragility of change over time. All this fits within the definitional scope of Big Data but not always into its practices.

These concepts are central to rhetoric, which, to return to Sullivan (2013), acts as a mediator between the worlds of particularity and situatedness, and the sometimes necessary need to move from there to a world of (as Farrell (1995) puts it), “collaborative agents [...] enacting phronesis through the timely choice (*kairos*)” (p. 326), a

¹¹ See also Bod’s differentiation between universal and particular patterns.

“more public, collectivized sense” (p. 228). An “audience” here is neither a set of expected behaviors targeted by a rhetorical act nor a counterfactual to which a rhetorical act may appeal, but an already-existing, agentic social field filled with rhetorical practices, enactments, and norms. It is “audience” in the sense of an ecological perspective, in which the rhetorical enactments and practices of people emerge “already infected by the viral intensities that are circulating in the social field,” crafting relationships that were not previously there and a reality that emerges from unbounded networks (Edbauer, 2005, p. 15). From this ecological vantage point, rhetorical practices become norms and resources, which advocate for, test, contest, challenge, or to cite Edbauer again, “infect” existing perspectives, attitudes, opinions, motivations. The promise of a rhetorical ecology with its unbounded entanglements of audiences and rhetors and external events and exigences is, just like the promise inherent in definitions of Big Data, the (ideal) *possibility* of seeing the motivations and attitudes and perspectives of people in all their complexity and tangled interrelations.

Of course, the inverse can also be true. Rhetorics also constitute their audiences (Charland, 1987), and datasets also reify the complexities of human relationships and engagements, representing pre-determined categories and abstractions as the lived realities and experiences of the people the data ought to represent (Poster, 2004; see also Noble, 2018). In data and computation-centric circles, these practices are called ontologies that “help integrate disparate or unorganized data [author’s note: hence, definitionally speaking, Big Data] to produce meaning, sort of ‘like a thesaurus, a finite set of terms, organized as a hierarchy’” (Iliadis, 2018, p. 223; citing Pomerantz, 2015, who shows that metadata fixes meaning in situational context but in so doing also sets meanings in seemingly stable relations). Big Data frequently constitutes its own audience, holding up a mirror in which we see ourselves as a one-dimensional “trait” or “category” or “demographic,” and—knowing that for better or worse, this reflection is how others see us—we act in alignment with the reductionist self reflected by the data (Poster, 2004). There is, for example, little nuance or complexity in a low credit score, and no matter how trustworthy or hardworking the person branded by it may be, many will act toward her as if that score *were* her lived reality and *represented* her character, and they will constrain her own behavioral agency in turn. Hence, these ontologies are not just ontologies as defined within a narrow technical industry context. They are also ontologies in the broader philosophical sense,

seemingly making “social spaces quantifiable” where “in reality, working with Big Data is still subjective, and what it quantifies does not necessarily have a closer claim on objective truth” (boyd & Crawford, 2012, p. 667). The unconcealing of such ontologies is precisely what a deep theory of rhetoric does: as Hariman (1986) argues, rhetoric’s “dynamic of concealment and unconcealment” situates “ontological claims within a social history,” or, more strongly (here citing Valesio), “is the key to ontology because it is the most concrete and precise tool that can be used in order to show that every positive ontology is an ideological construction” (p. 50–51).

These last two paragraphs reflect the introductory challenge of this essay. Rhetoric offers a deep conceptual toolset for critiquing Big Data, in large part because the definitional tenets of Big Data are reflected in many of the central conceptual principles of rhetoric. Hence, rhetorical studies offers a fitting lens for critique of Big Data because¹² it is conceptually closer to Big Data than many perspectives that critique from external vantage points. Consequently, there are also opportunities *in* Big Data *for* rhetorical analysis. While it remains important to resist the siren call of Big Data that “offers the humanistic disciplines a new way to claim the status of quantitative science” (boyd & Crawford, 2012, p. 667; see also Hariman, 1986, on the status of rhetorical discipline), responsiveness to the principles and considerations on which I center this essay can help guide uses of Big Data in rhetorical studies.

Implications and Conclusion

I open this section with the same caveat that I raised in the introduction. This essay is not *a priori* critical of Big Data. It uses key definitional arguments for what makes Big Data “Big Data” to draw entry points for a humanistic-rhetorical consideration of Big Data, and so to offer considerations for critique of, analysis grounded in, and engagement with Big Data from a rhetorical-humanistic perspective. These considerations include the importance of *relationships* in unstructured and often unrelated types of data; *observability* as a means of mediating between *all* available information and that which matters to take action in the moment; *patterns* that walk the thin line between the world of universals and the worlds of particulars; an awareness of

¹² See also Sullivan’s (2013) argument for rhetoric as a mediator.

diachronicity through which patterns link up to the dynamism of a living world; and finally, *audience*, and the various ways in which Big Data emerges from or can (often problematically) constitute its audience.

These conceptual entry points represent normative ideals for Big Data practices, drawn from the definitions of Big Data industries and practitioners themselves. Actual uses of Big Data frequently fall short as much of the critical literature on the topic persuasively demonstrates and as is evident in the horrors depicted in reporting where, to show one example, governments use what superficially appears to be Big Data to draw actionable inferences with real lived consequences about people (Byler & Boe, 2020). It should be evident, based on the considerations offered in this essay, that often such practices at best fall short of Big Data's functional definitions, and at worst make use of a willfully reductive understanding of Big Data.

I want to close with one final consideration, perhaps more fitting for a separate essay, and offered here more as a provocation than a sustained argument. What a foundational textbook in rhetoric defines as “rhetorical thinking”—bringing together the uncertain and contested particularities of a situation “with existence and human experience” (Hauser, 2002, p. 37) — resembles forms of Bayesian inference that are increasingly prominent in Big Data analytics. Consider McCloskey (1983), who, citing a review by Sims of a text by Leamer, points out that “the rhetorical context that creates such skepticism can be called a priori beliefs and can be analyzed in Bayesian terms” (p. 495). *All* rhetoric invites skepticism: the Aristotelian proving of opposites, the Burkean selection and de-selection of pieces of reality, the nature of knowledge in rhetorical contexts as emergent and contingent and uncertain, all openly suggest that the communicative engagements in rhetorical contexts always are enthymematic, always—as I discussed in the section on relationships—could be otherwise (Rapp, 2010).

Developments in Bayesian approaches have occurred “almost in parallel” with the rise of Big Data (Allenby et al., 2014, p. 169). In their “ability to deal with heterogeneity” (ibid.), the advantage of Bayesian approaches lies in being able to deal with uncertainty, “combining prior knowledge and uncertain evidence,” as well as their adaptability to changing contexts (Zhu et al., 2017, p. 2), evoking definitional concepts similar to those Hauser used to discuss rhetorical thinking. Bayesian statistics have been called

“common sense” in both industry guides (Kana, 2019) and academic literature because their approach to probabilities lies closer to how most humans think about probabilities than how probabilities are conceptualized in the frequentist paradigm. The role of prior probabilities as a “good guess [...] based on our real-life domain knowledge and common sense” (ibid.) has the distinct feel of a rhetorical culture (Farrell, 1995) where norms, patterns, contextually contingent knowledge, and commonplaces set the stage for engagement, advocacy, and debate. “Our brain,” argues Whimster (2018), “is an instrument that learns cumulatively on Bayesian principles—it understands on the back of accumulated experience and knows what to expect” (p. 4). Whimster cites this argument to make a point about relationality in communication: that the taking “of cues from some pre-existing connection that links individuals” (p. 4) is central to how we communicate. Memes, patterns, networks (ibid.) all represent constitutive concepts of communication that relate the general (the accretion of experiences, rhetorical practices, perspectives, etc.) to the particular lived communicative engagements from which they emerge and which they facilitate. As Whimster summarizes it, “minds can communicate with very little sweat on the brow through the use of memes” (ibid.).

A Bayesian approach is not Big Data, and Big Data is not all Bayesian. But I raise this topic as a concluding thought to demonstrate, along with the main part of the essay, some common conceptual grounds between the domain of rhetoric (with its focus on the humanistic, lively, inventive, engaging, and uncertain) and a domain of science and technology that often to many of us (and rightly so) connotes the opposite.¹³ This common ground cuts both ways: it can guide our own rhetorical analyses if the questions we raise in our research require the kind of voluminous, varied, and volatile textual data that define Big Data approaches (see e.g. Bod, 2013a; Graham et al., 2015; Gray & Holmes, 2020; Hart & Lind, 2011; Koteyko, 2015; Majdik, 2019; Waisanen & Kafka, 2020). It can also serve as a resource for critique, acting as a lens through which uses and applications of Big Data can be assessed from a rhetorical-humanistic perspective based on Big Data’s own functional definitions about what Big Data is and ought to be.

Copyright © 2021 Zoltan P. Majdik.

¹³ See also the opening paragraph of Hart (2001).

Reference List

- Allenby, G. M., Bradlow, E. T., George, E. I., Liechty, J., & McCulloch, R. E. (2014). Perspectives on Bayesian methods and big data. *Customer Needs and Solutions*, 1(3), 169–175.
- Barnaghi, P., Sheth, A., & Henson, C. (2013). From data to actionable knowledge: Big data challenges in the web of things. *IEEE Intelligent Systems*, 28(6), 6–11.
<https://doi.org/10.1109/MIS.2013.142>
- Berkeley School of Information. (2019, April 12). Big data isn't a concept—it's a problem to solve. *Berkeley School of Information*. <https://datascience.berkeley.edu/blog/what-is-big-data/>
- Bod, R. (1998). *Beyond grammar: An experience-based theory of language*. CSLI Publications.
- Bod, R. (2013a). *A new history of the humanities: The search for principles and patterns from antiquity to the present*. Oxford University Press.
- Bod, R. (2013b). Who's afraid of patterns?: The particular versus the universal and the meaning of humanities 3.0. *Low Countries Historical Review*, 128(4), 171–180.
- Bod, R. (2018). Has there ever been a divide? A *longue durée* perspective. *History of Humanities*, 3(1), 15–25.
<https://doi.org/10.1086/696299>
- Boellstorff, T. (2013). Making big data, in theory. *First Monday*, 18(10), 1–17. <https://doi.org/10.5210/fm.v18i10.4869>
- boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679.
<https://doi.org/10.1080/1369118X.2012.678878>
- Bragin, T. (2019, February 28). Observability with the elastic stack. *Elastic*. <https://www.elastic.co/blog/observability-with-the-elastic-stack>
- Byler, D., & Boe, C. S. (2020, July 24). Tech-enabled 'terror capitalism' is spreading worldwide. The surveillance regimes

must be stopped. *The Guardian*.
<https://www.theguardian.com/world/2020/jul/24/surveillance-tech-facialrecognition-terror-capitalism>

- Charland, M. (1987). Constitutive rhetoric: The case of the *peuple québécois*. *Quarterly Journal of Speech*, 73(2), 133.
<https://doi.org/10.1080/00335638709383799>
- Diebold, F. X. (2012). A personal perspective on the origin(s) and development of ‘big data’: The phenomenon, the term, and the discipline, second version. *PIER Working Paper No. 13-003*. <https://doi.org/https://dx.doi.org/10.2139/ssrn.2202843>
- Drucker, J. (2012). Humanistic theory and digital scholarship. In M. K. Gold (Ed.), *Debates in the Digital Humanities* (pp. 85–95). University of Minnesota Press.
- Edbauer, J. (2005). Unframing models of public distribution: From rhetorical situation to rhetorical ecologies. *Rhetoric Society Quarterly*, 35(4), 5–24.
<https://doi.org/10.1080/02773940509391320>
- Farrell, T. B. (1995). *Norms of rhetorical culture*. Yale University Press.
- Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy & Technology*, 25(4), 435–437.
- Flyverbom, M. (2019). *The digital prism*. Cambridge University Press.
- Graham, S. S., & Herndl, C. G. (2011). Talking off-label: The role of stasis in transforming the discursive formation of pain science. *Rhetoric Society Quarterly*, 41(2), 145–167.
<https://doi.org/10.1080/02773945.2011.553764>
- Graham, S. S., Kim, S.-Y., DeVasto, D. M., & Keith, W. (2015). Statistical genre analysis: Toward big data methodologies in technical communication. *Technical Communication Quarterly*, 24(1), 70–104.
<https://doi.org/10.1080/10572252.2015.975955>
- Gray, K., & Holmes, S. (2020). Tracing ecologies of code literacy and constraint in emojis as multimodal public pedagogy.

Computers and Composition, 55, 102552.
<https://doi.org/10.1016/j.compcom.2020.102552>

- Greene, J. O. (1984). A cognitive approach to human communication: An action assembly theory. *Communication Monographs*, 51, 289–306.
- Hariman, R. (1986). Status, marginality, and rhetorical theory. *Quarterly Journal of Speech*, 72(1), 38–54.
<https://doi.org/10.1080/00335638609383757>
- Hart, R. P. (2001). Redeveloping DICTION: Theoretical considerations. In M. D. West (Ed.), *Theory, method, and practice in computer content analysis* (pp. 43–60). Greenwood Publishing Group.
- Hart, R. P. (2015). Genre and automated text analysis: A demonstration. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 152–168). University of Chicago Press.
- Hart, R. P., & Lind, C. J. (2011). The rhetoric of Islamic activism: A DICTION study. *Dynamics of Asymmetric Conflict*, 4(2), 113–125. <https://doi.org/10.1080/17467586.2011.627934>
- Hauser, G. A. (2002). *Introduction to rhetorical theory* (2nd ed.). Waveland Press.
- Iliadis, A. (2018). Algorithms, ontology, and social progress. *Global Media and Communication*, 14(2), 219–230.
<https://doi.org/10.1177/1742766518776688>
- Kana, M. (2019, October 12). Bayesian nightmare. Solved! *Towards Data Science*. <https://towardsdatascience.com/bayesian-nightmare-how-to-start-loving-bayes-1622741fa960>
- Kessler, M. M., & Graham, S. S. (2018). Terminal node problems: ANT 2.0 and prescription drug labels. *Technical Communication Quarterly*, 27(2), 121–136.
<https://doi.org/10.1080/10572252.2018.1425482>
- Kobielus, J. (2016). *Surprise in the narrative flow of data science*. Retrieved July 21, 2020, from <https://www.linkedin.com/pulse/surprise-narrative-flow-data-science-jameskobielus>

- Koteyko, N. (2015). Corpus-assisted analysis of Internet-based discourses: From patterns to rhetoric. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 184–198). University of Chicago Press.
- Koteyko, N., Jaspal, R., & Nerlich, B. (2013). Climate change and ‘climategate’ in online reader comments: A mixed methods study. *The Geographical Journal*, 179(1), 74–86.
<https://doi.org/10.1111/j.1475-4959.2012.00479.x>
- Lewis, S. C., & Westlund, O. (2015). Big data and journalism. *Digital Journalism*, 3(3), 447–466.
<https://doi.org/10.1080/21670811.2014.976418>
- Luhmann, N. (1992). What is communication? *Communication Theory*, 2(3), 251–259.
- Majdik, Z. P. (2019). A computational approach to assessing rhetorical effectiveness: Agentic framing of climate change in the Congressional Record, 1994–2016. *Technical Communication Quarterly*, 28(3), 207–222.
<https://doi.org/10.1080/10572252.2019.1601774>
- Manning, C. D. (2002). Review: Rens Bod, beyond grammar: An experience-based theory of language. *Journal of Linguistics*, 38(2), 441–462.
<https://doi.org/10.1017/S0022226702211639>
- Mays, C. (2017). Writing complexity, one stability at a time: Teaching writing as a complex system. *College Composition and Communication*, 68(3), 559.
- McCloskey, D. N. (1983). The rhetoric of economics. *Journal of Economic Literature*, 21(2), 481–517.
- Microsoft. (2019). *Big data architecture style - Azure application architecture guide*.
<https://docs.microsoft.com/enus/azure/architecture/guide/architecture-styles/big-data>
- Montañez, A. (2016, May 20). Unveiling the hidden layers of deep learning. *Scientific American*.
<https://blogs.scientificamerican.com/sa-visual/unveiling-the-hidden-layers-of-deep-learning/>

- New Relic. (2020). *The age of observability*.
<https://newrelic.com/resources/ebooks/what-is-observability>
- NIST Big Data Public Working Group Definitions and Taxonomies Subgroup [NIST]. (2015, October). *NIST big data interoperability framework: Volume 1, definitions* (NIST SP 1500-1). National Institute of Standards and Technology.
<https://doi.org/10.6028/NIST.SP.1500-1>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Pickering, A. (1993). The mangle of practice: Agency and emergence in the sociology of science. *American Journal of Sociology*, 99(3), 559–589.
- Pomerantz, J. (2015). *Metadata*. MIT Press.
- Poster, M. (2004). The mode of information and postmodernity. In R. Blom, E. Karvonen, H. Melin, K. Nordenstreng, E. Puoskari, F. Webster, & P. F. Webster (Eds.), *The information society reader* (pp. 398–410). Routledge.
- Rapp, C. (2010). Aristotle's rhetoric. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2010).
<https://plato.stanford.edu/archives/spr2010/entries/aristotle-rhetoric/>
- Saltelli, A., Bammer, G., Bruno, I., Charters, E., Fiore, M. D., Didier, E., Espeland, W. N., Kay, J., Piano, S. L., Mayo, D., Jr, R. P., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., ... Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. *Nature*, 582(7813), 482–484.
<https://doi.org/10.1038/d41586-020-01812-9>
- Schwartzman, R., Ross, D. G., & Berube, D. M. (2011). Rhetoric and risk. *Poroi*, 7(1). <https://doi.org/10.13008/2151-2957.1087>
- Sullivan, D. L. (2013). A call for reaffirming a humanist understanding of technology. *Programmatic Perspectives*, 5(1), 153–157.

- Taylor, C. (2010). Science in the news: A diachronic perspective. *Corpora*, 5(2), 221–250.
<https://doi.org/10.3366/cor.2010.0106>
- van der Hel, S., Hellsten, I., & Steen, G. (2018). Tipping points and climate change: Metaphor between science and the media. *Environmental Communication*, 12(5), 605–620.
- Waisanen, D. J., & Kafka, J. (2020). Conflicting purposes in U.S. education reform: The paradoxes of Arne Duncan's educational rhetoric. *Rhetoric & Public Affairs*, 23.
- Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). Some tentative axioms of communication. In *Pragmatics of human communication: a study of interactional patterns, pathologies and paradoxes* (pp. 48–71). W. W. Norton & Company.
- Whimster, S. (2018). Pure relationality as a sociological theory of communication. *Frontiers in Sociology*, 3.
<https://doi.org/10.3389/fsoc.2018.00001>
- Zhu, J., Chen, J., Hu, W., & Zhang, B. (2017). *Big learning with Bayesian methods*. <http://arxiv.org/abs/1411.6370>