RHETORICAL MACHINES

Writing, Code, and Computational Ethics

Edited by
John Jones
and
Lavinia Hirsu

THE UNIVERSITY OF ALABAMA PRESS
Tuscaloosa

11 / Good Computing with Big Data

Jennifer Helene Maher, Helen J. Burgess, and Tim Menzies

Numeric calculation has long been understood to potentially hold the power of overcoming the frailty of human judgment. For example, through his conception of the *calculus ratiocinator* (thought calculator) and a *lingua characteristica* (universal language), Gottfried Leibniz conjectured in the seventeenth century, "If controversies were to arise, there would be no more need for disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and say to each other (with a friend as witness, if they liked): calculemus—let us calculate."[1] The hopeful promise of mechanical inventions such as these points to the desire to cut through rhetorical uncertainty and replace human fragility with numerical certitude. To date, numeracy as computation in big data analytics (i.e., the computing of large-volume sets of information) appears closest to achieving Leibniz's dream. But discussions of goodness are often limited in certain ways in programming generally, and in data science specifically. To wit, software, according to Joseph Juran, is written well when it enjoys a "fitness for use" that can be measured through its "freedom from deficiencies."[2] In data analytics, technically good algorithms maximize the ability to process petabytes worth of data according to the three V's: volume, variety, and velocity.[3] Computing "good" is also often described as revealing otherwise hidden patterns of information that can positively address social issues such as healthcare in ways that would not otherwise be possible.[4] And less instrumental or consequential considerations of goodness often take shape in methodological discussions of how "ethical mining" can address privacy concerns[5] or how "moral mining" can reveal new information about human values.[6] Although examples such as these illustrate how *good* is often invoked in discussions of big data, they fail to capture, in their zeal to determine goodness as a kind of total valuation, the ways in which working with big data is often a complex negotiation of different kinds of good at different levels. Working with big data, it turns out, is a process that involves many layers of acquisition, cleaning, analysis, reporting, revising—any layer of which might be *good* in a different mode or multiple modes: moral/ethical, technical, ideological. And these different valuations of goodness might be formulated, negotiated, and revised multiple times during the process of working with a dataset.

To illustrate the ways in which goodness circulates in complex ways demands understanding the processing of large-volume sets of data as an inherently rhetorical enterprise rather than as the long-sought realization of *calculemus* through computation. To this end, we first unpack goodness in the context of big data and examine how the desire to render goodness in ways limited to conceptions of the technical, the methodological, or the consequential obfuscates the complex play of goodness that is always at work in computation because data analytics is not free of the kind of human deliberations and judgments that define rhetorical activity. Then, borrowing Benjamin Bratton's concept of "the Stack," we illustrate, through an analysis of a gender study formulated from pull requests on the open-source repository GitHub, how rhetorical/technological "stacking" functions in ways that circulate in a complex assemblage of various conceptions of goodness and its opposite. In doing so, big data, with its seductive promise to deliver numerical certitude, is understood to propagate rather than eliminate rhetoric.

Limited Goodness

Popular narratives highlight the power of big data to reveal otherwise hidden ways of improving everyday life through the computational ability to find and identify patterns in petabyte-size data. For example, between 2008 and 2014 Google Flu Trends tracked the real-time spread of the flu, in an attempt to control seasonal influenza epidemics that attack 5 to 10 percent of adults and 20 to 30 percent of children each year, causing 250,000 to 500,000 deaths[7]. To do so, Google Flu Trends aggregated search-term data determined to be "good indicators of flu activity" in an effort to improve upon the US Centers for Disease Control's (CDC) and European Influenza Surveillance Scheme (EISS) surveillance systems. As Google summarized, "Traditional flu surveillance is very important, but most health agencies focus on a single country or region and only update their estimates once per week. Google Flu Trends is currently available for a number of countries around the world and is updated every day, providing a complement to these existing systems."[8] In another example of claims for the good generated through big data analysis, Uber announced on January 13, 2015, that it was joining with the city of Boston "to help expand the city's capability to solve problems by leveraging data provided by Uber. The data will provide new insights to help manage urban growth, relieve traffic congestion, expand public transportation, and reduce greenhouse gas emissions."[9] Rather than treating this collaboration with Boston as simply a technical matter of calculating potential solutions to urban problems, Uber casts itself in decidedly moral terms in arguing that "our ability to share information . . . can serve a greater good."[10] Likewise, IBM's Big Data for Social Good Challenge, launched in 2015, invites data scientists to use its curated datasets to tackle "real world civic issues" and offer solutions with social impact. Successful competitors have thus far used open data sets for several purposes: Watch Flu Spread tracks historic and forecasted dispersion of flu incidents in the US in order to mitigate risk in specific areas, Juvo provides assistance in setting social and economic development plans with realistic achievable goals and metrics, and Oasis helps to address the problems of commercial deserts in Chicago.

The kinds of goodness that pervade many of these discussions focus in Aristotelian terms on big data as scientific knowledge (*episteme*) and as art or craft (*techne*). As Aristotle explains, "The origin of action . . . is choice, and that of choice is desire and reasoning with a view to an end."[11] But the Greek philosopher notes that both choices and the ends they ought to achieve depend upon the kind intellectual activity at work. Because science is that which "cannot be otherwise," this kind of knowledge has been constituted, since the advent of modern science, through the uncovering of those first principles and universals laws of necessity that can be demonstrated and replicated through analytic methods. "Good science," first of the natural world and now of a computable world, is meant to bring forth certainty without qualification through good scientific method, whether hypothesis- or data-driven. In doing so, data science is often understood as encapsulating a promised first principle of the digital age, as expressed by Andrew McAfee: "As the amount of data goes up, the importance of human judgment should go down."[12] But those techniques applied in data processing can lead to good data processing or bad data processing because "art . . . is a state concerned with making, involving a true course of reasoning, and lack of art on the contrary is a state concerned with making, involving a false course of reasoning: both are concerned with what can be otherwise."[13] As a result, the techniques by which data scientists address a problem have to be technically good, meaning that chosen methods aim to ensure the values of science: reliability and validity. Only then can the potential for consequential good achieved through science be actualized.

But whether as science or art, data science is too often limited to notions of goodness as either social consequences or matters of technique, the effect of which is to displace human conceptions of goodness onto those of the machine. In his discussion of cybernetics, Norbert Wiener acknowledged the potentially dangerous consequences that could result from cybernetic displacement of human values to that of the machine: "Those of us who have contributed to the new science of cybernetics thus stand in a moral position which is, to say the least, not very comfortable. We have contributed to the initiation of a new science, which, as I have said, embraces technical possibilities for good and for evil."[14] His answer as to how best to deal with the moral consequences of machine communication in light of such horrors as Bergen-Belsen and Hiroshima was to encourage a more enlightened public and to limit scientific and technical progress to those areas where the consequences of use might be much less drastic, such as in the fields of physiology and psychology, those "most remote from war and exploitation."[15] At the same time, Wiener held "very slight hope"[16] for the likelihood of implementing limitations to computational investigation. And in fact, sixty years later we find Grady Booch, chief scientist for software engineering at the IBM Almaden Research Laboratory, arguing in a 2008 issue of *IEEE Software* for a collective recognition of the important role of software engineers: "We as a professional community have developed technology that has changed the way individuals live, business operates, communities interact, and nations and civilizations thrive and expand. At that level of abstraction, a moral dimension is undoubtedly at play: when our technology touches the activities of the human spirit, then questions of social responsibility, individual rights, and goodness of fit to the moral atmosphere of the surrounding community come alive."[17] What both Wiener and Booch respectively advocate is the adoption of computing praxis, especially by those whose work is instrumental in the development of computing machines. Booch subsequently explained: "To say that algorithms are thoughtless is a reasonable and unemotional statement of fact. They have no moral center; they have no sense of right or wrong; they cannot take responsibility for their consequences. Bits cannot feel. However, we who craft such algorithms are expected to be thoughtful."[18] Thus, in the case of big data, the call for and implementation of ethical mining can be understood as the fulfillment of a choice to compute and interpret large data sets according to a positive value regarding privacy, for instance. But calculating the good or ill effects of computing in-use catalyzes a philosophical consequentialism that too often assigns to the data scientist the role of moral judge. Given the ongoing dominance of technological neutrality, the fact that morality could be baked in from the very start does not sit well, especially with those in the business of algorithmic modeling.[19] In order to maximize fitness of use in software, an emphasis is also placed on not just good, but *best* algorithmic practices. Through an emphasis on technical concerns such as security, scalability, reliability, efficiency, verifiability, and adaptability, *good* often denotes a measurable, instrumental quality.

However, if looked at from a systems approach point of view, data science as computation is removed completely from human judgment concerning goodness. In his theory of systems, Niklas Luhmann explicates a way to resist the ongoing privileging of human action.[20] Action must, in his view, be understood not as originating from human consciousness but rather from systems of communication that constantly remake themselves in a process known as *autopoiesis*. Originally described by Humberto Maturana and Francisco Varela as the self-reproduction of the living system of the cell, autopoiesis foregrounds the recursive regeneration of a system through its "network of processes of production,"[21] as well as its "operational closure" from other systems. As Maturana explains, "This circular organization constitutes a homeostatic system whose function is to produce and maintain this very same circular organization by determining that the *components* that specify it be those who synthesis and maintenance it secures."[22] In addition to the biological, Luhmann also includes in his systems model the social, psychic, and mechanistic.[23] But regardless of which model a system falls under, the autopoiesis of any system occurs through a self-reproduction of a binary code specific to that system's function. Thus, in the social system of the law, this binary is legal/illegal, the meaning of which serves both as the object of the system's self-reproduction and as a necessary distinction from other systems. In light of this binary, law as a social system is regenerated through "programmes" such as legislative acts and adherence to legal precedents[24] that, in turn, are regenerated through the binary. As a result, "Everything that cannot be brought under this controlling scheme of legal/illegal does not belong to the legal system but to its internal or external social environment."[25]

Given these descriptions of systems, one might rightly wonder how the autopoietic turn could be useful in a discussion of anything beyond technical good in computation, especially as Luhmann argues that system communication "does not permit code values of function systems to be identified with moral values—neither with good/bad nor with good/evil."[26] In the system of law, excluding its system communication from the moral might seem necessary, if not natural, as is the case in liberalism where the law is to be neutral in regards to morality in order to treat all citizens fairly and equally.[27] But even in the system of computation, any operation cast in terms of good or bad, except in the instrumentalist measure of technical efficacy and effect that we discussed earlier as *good computing*, can seem unnecessary. But this is not to say that notions of goodness have no place in systems of communication, including the law. While functional codes such as legal/illegal exists on a level of "higher amorality,"[28] Luhmann explains that goodness comes into play through a kind of binary code coupling: "Moral communication has to be framed within a specific binary code which opposes a positive and a negative value. This code can be supposed to be invariant because it is necessary to identify communication as moral communication. It is specific and universal at the same time because, once invented, there cannot be an uncoded moral communication . . . But this evolutionary universal is void of content. It does not give any information about what is good and what is bad . . . As a complement to its code, the moral needs criteria to decide which behaviour is good and which behaviour is bad. Since there are no good versus bad criteria, the criteria or programmes of the moral cannot be identified with the values of the code."[29] Where the moral couples with the system of computation is not in some kind of value judgment of the system's binary. As Luhmann explains, moral communication is framed *within*, not *with*, the binary code. Therefore, it is at the level of operation that moral value can couple with programmes, for example, to create "differentiation." In the system of the law, moral conflict concerning segregations laws in the American South, particularly in the 1960s, offers a clear instance of just how the moral can "irritate" operations. Individuals stoke this differentiation as they "choose the programmes that favor their own interests and opinions."[30] The resulting operation of moral communication that occurs from the choice between supporting or condemning segregation laws, codes individual action in what Luhmann identifies as an "esteem/disesteem" binary, like that achieved through epideictic rhetoric. Thus, while the binary of legal/illegal exists in what might be identified as the autopoietic system of law through which the system's function is to essentially beget more law, the coupling of moral communication within the legal system irritates so that in some ways that operational closure has meaningful rhetorical openness too.

However, the system of computation's coupling with moral communication is perhaps more difficult to identify than that of the law's. Take the example of protocol. Alexander Galloway identifies *protocol* as "that machine, that massive control apparatus that guides distributed networks, create cultural objects, and engenders life forms."[31] In order for these transactions to occur, they must be constructed in such a way as to conform not to the protocols of human norms—whether in language or behavior—but to the hierarchical layers of network architecture.[32] For Galloway, however, there is no coupling of moral communication with protocol: "People often ask me if I think protocol is good or bad. But I'm not sure this is the best question to ask. It is important to remember first that technical is always political, that *network architecture is politics*. So protocol necessarily involves a complex interrelation of political question, some progressive, some reactionary."[33] While it may seem that coupling the system of computation with the system of politics at the operation level of protocol achieves something essentially similar to what a coupling with moral communication might do (i.e., cast value), this is not the case. Putting aside the intricacies of the system of politics,[34] political communication does not involve differentiation rooted in esteem/disesteem, except when the system of politics itself is coupled with moral communication. As its core, according to Luhmann, moral communication's autopoiesis only occurs as a result of a certain kind of problem, one that political consensus would eliminate. If this were to take place with the moral, however, "This would bring moral communication to an end."[35]

Yet, the promise to remove human consideration of goodness beyond that of operation or (if one must) consequence ought to be viewed suspiciously as the kind of hype that can result from new science. In 2009, for instance, Google Flu Trends' Jeremy Ginsberg et al. noted that, in using search engine query data to detect influenza epidemics, "We can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day."[36] Unlike the CDC methods that relied on empirical evidence, including "both virologic and clinical data, including influenza-like illness (ILI) physician visits," Google Flu Trends methods included generating a model by aggregating "hundreds of billions of individual" search queries submitted between 2003 and 2008 and then current web queries starting in the 2008–2009 flu season. What is more, Google Flu Trends, by "harnessing the collective intelligence of millions of users," cut the one to two week lag time of ILI reports used by the CDC to provide same-day results. In short, Google Flu Trend purported to offer increased goodness both technically and methodologically.[37] Among the praise offered for Google Flu Trends included that of Viktor Mayer-Schönberger and Kenneth Cukier, who in *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, remark on the ways in which "sophisticated computational analysis can now identify the optimal proxy—as it did for Google Flu Trends, after plowing through almost half a billion mathematical models."[38] In doing so, Google Flu Trends evidenced strong correlation, meaning that "when one of the data values changes, the other is highly likely to change as well."[39] The result, according to Mayer-Schönberger and Cukier, is this: "We don't have to develop a notion about what terms people search for when and where the flu spreads . . . Instead we can subject big data to correlation analysis and let it tell us what search queries are the best proxies for the flu . . . In place of hypothesis-driven approach, we can use a data-driven one. Our results may be less biased and more accurate, and we will almost certainly get them much faster."[40]

Given the number of instances of illness and death caused by seasonal flu each year, to say nothing of thirty to fifty million estimated to have been killed in the 1918–1919 pandemic, the "optimal proxy" created through computational analysis purported to offer the best means by which to provide the greatest degree of statistical reliability and validity, not to mention the increased awareness of a health threat in real time. Through Google Flu Trends the promise of big data's inherent goodness, beyond that otherwise achieved through traditional, empirical methods, was optimized, seemingly to the point of certainty rather than probability.

In 1977, the International Association for Statistical Computing described its mission as linking "traditional statistical methodology, modern computer technology, and the knowledge of domain experts in order to convert data into information and knowledge."[41] As later described by Ian Hacking, there are two aspects to statistical probability: "It is connected with the degree of belief warranted by evidence, and it is connected with the tendency, displayed by some change devices, to produce stable relative frequencies."[42] The stability generated through the result of strong correlation in big data is often misrecognized as certainty. As Ginsberg et al. warned, "Despite strong historical correlations, our system remains susceptible to false alerts caused by a sudden increase in ILI-related queries. An unusual event, such as a drug recall for a popular cold or flu remedy, could cause such a false alert."[43] Even so, what is easily overlooked in computation is the tendency to confuse statistical probability with certainty. As Google Flu Trends came under critique for its less-than-stellar results for flu season 2012–2013[44]—and this in spite of the fact that these findings were based on twice as many ILI doctor visits as evidenced by data collected by the CDC[45]—Google simply characterized these critiques as useful feedback in honing the new model for the new flu season. Yet, in 2014, Google ended its flu trend analytics project.

For data scientist Michael Jordan, the promise that big data seemingly holds and the hype that surrounds it too often encourage overlooking the "false positives" that are frequently generated through such computation. As he explains, "I think data analysis can deliver inferences at certain levels of quality. But we have to be clear about *what* levels of quality. We have to have error bars around all our predictions."[46] Without this kind of qualification on the results generated through data analytics, Jordan predicts a "big-data winter." Yet, for those such as Justin Washtell, who asks—"So, what *is* the single best predictive modeling technique available, imho?"—it is the minimization and possibly even the eventual elimination of ambiguity induced by humans, who are not by their nature computational beings, that could be the ultimate promise of big data analysis. Or, as Washtell succinctly puts it, "Take human judgment out of the equation where it is not required."[47] In doing so, the goodness made possible by big data stands to enjoy greater stability, reliability, and validity, possibly to the point of certainty. But, while removing the human element may lead to the achievement of *calculemus* or, what might more accurately be referred to now in an English translation as "let them compute," statistical probability is not certainty. Consequently, big data is nothing less than a rhetorical enterprise, one that necessitates judgments that go beyond the limitations induced at the level of system or protocol. To understand this necessitates going beyond protocol or system to acknowledge the varying rhetorical and situated constructions of goodness that always occur in data analytics.

Rhetorical Stacking

To argue that "Big Data" is a rhetorical enterprise is to invoke the concept of "Big Rhetoric." As Aristotle explains, rhetoric, itself an art, is special because it is made use of by all other *techne*. Because rhetoric presents "us with alternative possibilities,"[48] we might easily understand how discussions of best methods can be limited to rhetorical aspects of data science. However, to acknowledge that science itself is also rhetorical, a position not supported by the Aristotelian configuration of *episteme*, means to reject the idea that anything can be inherently "free of rhetoric."[49] In Edward Schiappa's definition, "big rhetoric" refers to the position "that everything, or virtually everything, can be described as 'rhetorical.'"[50] And because everything is rhetorical, there is a necessary human calculation as to what is good. To explore the different ways in which goodness can be constituted in big data, we turn to Benjamin Bratton's formulation of "the Stack" in his 2015 book of the same name.[51] The Stack rests on the intuition that "an accidental megastructure" is being formed as we move to an era of "planetary computing" constituted by the complex interweavings of cloud computing technologies, big data repositories, trans-sovereign communications networks, the coexisting material economies of rare earth mineral acquisition, outsourced labor, and the global flows of manufacturing capital.[52] Bratton conceptualizes the Stack to describe the geopolitical framework formed by all these technologies: a kind of theoretical geological stratification of human and nonhuman activity into layers that operate at different intensities and scales—of time, of space, of matter.

Crucially, though, the Stack is (unlike a strict geological strata model, or a layer cake) both multidirectional and communicational: information can pass up and down (or between different configurations of) its layers: "Even as any one layer's operations unfold in relation to those adjacent layers, and so may also affect events well outside the entire platform's borders, the movement of hard and soft information must always pass through the protocols that divide and bind that layer's work from the others."[53] Bratton draws this metaphor from the way technological stacks function: he uses the example of the TCP/IP and OSI network models that allow for Internet connectivity and communication, which are made up of layers of different hard and soft technologies and operations—data, fiber optics, transport protocols.[54] We could also consider the open-source LAMP stack (Linux, Apache, MySQL, PHP/Python/Perl) as an example of this formulation: a suite of hard and soft technologies that pass information back and forth, even as they are reliant on layered dependencies within the stack to function.

Building on Bratton's conceptual terminology, we make use of "stacking" as a way of identifying ways in which the system of open source communities and technologies form what we could call a "rhetorical/technological stack," wherein both data (raw and processed) and rhetorical practices are communicated up and down through multiple technological layers. We are accustomed to thinking that data work is a messy and technical task. A stack model helps to illuminate the complex, multistage and multitool, often highly bespoke process of obtaining, cleaning, and interrogating data. But as with the case of the binary of "legal/illegal" identified by Luhmann in the autopoietic system of the law, wherein the system function is to essentially beget more law, the coupling of moral communication with data science's binary system of (statistically) "significant/insignificant" also irritates, so that the operational closure of a system where data begets more data has rhetorical openness too. Placing emphasis on rhetorical operations within the stack also helps to lay bare many other narratives of goodness implicit in working with data: for example, the way scholars must negotiate complex presentation, review, and revision processes in order to communicate and legitimize their work.

Reading Gender Bias on GitHub: A Sample Stack Analysis

In the remainder of this essay, we discuss a recent study that provides an ideal example of how a rhetorical/technological stack might function as a useful metaphor for the complexities of working with big data. The study, titled "Gender Bias in Open Source: Pull Request Acceptance of Women Versus Men," was authored by a group of researchers from Cal Polytechnic San Luis Obispo and North Carolina State University and evaluated the evidence for gender bias in open source software (OSS) contributions using GitHub as its primary source of data.[55] On February 9, 2016, the paper resulting from the study was originally posted to PeerJ, an open-access repository for non-peer-reviewed papers that might be in the process of revision for later publication. Because PeerJ operates on open-access principles, papers are freely available and open for public peer review and comment, with the result that this study was quickly picked up by media sources, including the *Washington Post* and the *Guardian*, and thence made its way through the social media ecosystem—including the usual suspects Twitter and Facebook, along with various Red-

One line short

dit groups and a wide number of blogs. The paper was revised into a second version under the title “Gender Differences and Bias in Open Source: Pull Request Acceptance of Women Versus Men” on July 26, 2016, with the new addition of a statistician as a third author. The study contains some familiar elements of other investigations we may know similarly from our own reading—particularly, “hiring bias” studies showing that resumes with African-American and Latino-sounding names were more likely to be rejected than those with White-sounding names, as well as the widely cited “Matilda Effect” study of several peer-reviewed communications journals that found women were less likely to be cited by male authors than female.[56] The study we are looking at differed, however, in that it was a “retrospective field study” rather than experimental (i.e., it analyzed existing, real-world data rather than setting up an unique study); it made use of a large data set gleaned from GitHub, isolating user data for contributors to projects under specific OSS licenses.

The paper itself presents evidence that there is a small but significant bias against acceptance of women’s contributions to OSS projects, but only when their gender identity is evident on the site. Where it is not readily available, conversely, contributions from women are more likely to be accepted than those from men, again by a small margin. Contributions were counted in the form of successful “pull requests,” the process whereby a contributor makes a change to a piece of code and that change is approved and merged back into the project. The study further broke down contributors by “insider” and “outsider” status on a project (i.e., whether the contributor was a known owner or collaborator), showing that the gender bias effect was confined to outsider contributors, whereas insider contributors showed “little evidence of bias.”[57] The study moves systematically through a number of research questions, seeking to isolate alternative theories for evidence of bias: for example, do women contribute differently to projects than men in terms of scope or task? But ultimately the authors find these alternative explanations insignificant.

The paper is simultaneously circumspect in its specific conclusion about gender bias and more ambitious in its general claim that big data is a crucial element in any such study. In its penultimate section, the authors note, “We caution the reader from interpreting too much from statistical significance; for big data studies such as this one, even small differences can be statistically significant. Instead, we encourage the reader to examine size of the observed effects.” They go on to note that “the effect we have uncovered is smaller than in typical gender bias studies.”[58]

In the paper’s conclusion, the authors shift their focus from the particularities of the GitHub study to argue more generally for the value of big data: “In closing, as anecdotes about gender bias persist, it is imperative that we

One line short

use big data to better understand the interaction between genders."[59] In making this claim that big data is an essential tool for unpacking and critiquing gender bias, they are implicitly acknowledging that consequentialist forms of goodness might result, whether it be a response to the problem of gender bias using technical solutions, such as changing identification practices on GitHub, or rhetorical action, such as opening up dialogue about gender bias in technical cultures and advocating for change.

Responses to the first draft of the paper posted to PeerJ range from general feedback on terminology (e.g., the conflation of "open source" and "GitHub") and discussion of the weakness of visualization (notably, the misleading effect of scale choice in bar charts) to more technical suggestions about using multiple regression models and releasing data, with some discussion about what it meant to "scrub" data given that it was already publicly available on GitHub. Such technical critiques resulted in some rewriting, new visualizations, addition of a new section on covariate analysis, and the addition of a statistician, presumably to address critiques of the goodness of the study on statistical and methodological grounds. One of the study's authors, Emerson Murphy-Hill, commented on the public's reception to the nuanced nature of the study in terms of its "significance," noting that "the difference is statistically significant, but whether the difference is substantial is another question that's open for interpretation."[60]

The first inklings of defensive responses by male coders, however, soon appeared in the commenting system. For example, referring to the opening anecdote, a self-described "actual developer" Francisco Villemaire argued, "Rachel claiming sexism is a cop-out move, it sounds like she can't handle being critiqued and instantly plays the victim card."[61] The commenter was called out quickly for his attempt to claim authority by the mere fact of being a web developer, but would prove fairly representative of responses once the paper made its way from the relatively rarified environment of PeerJ to spaces such as Reddit, where the conversation took a fairly predictable turn into misogyny, with one commenter suggesting that "the only conclusion that can actually be drawn from the paper is that there is a bias against *men* having their pull requested accepted."[62] These kinds of responses are, of course, easy pickings for rhetorical analysis and do not represent anything new for any feminist who has spent any time on Twitter or in a Reddit comment section.[63] Various arguments for technical goodness (or badness) are routinely used in the service of justifying or disputing the conclusion of a study, particularly when it involves a topic such as race or gender. But the kinds of public arguments we see in Reddit conversations and news site comment sections only represent one layer in a hypothetical rhetorical/technological stack where a binary cou-

One line short

pling irritates the otherwise seemingly closed operation of the data science system and therefore might usefully inform a more nuanced analysis of what good might actually come out of such a study and such a system.

Components in the Rhetorical-Technical Stack

Terrell et. al.'s study of gender bias in computing is notable for its complex interweaving of rhetorical and technological elements all the way through the process, from data collection to analysis to dissemination. In conducting a sample stack reading of the paper, we will concentrate on several layers:

1. *A rhetorical substrate*. First, it must be noted that any human activity is a complicated mix of actions, politics, technologies, and ideologies. In the case of the open source community generally, we have a well-documented system of software development dependent on multiple soft and hard elements—everything from coding bodies to machines to the open source stack of software that enables computation and communication. Let's consider this the "necessary substrate" of the study, the activity without which the study would have no research question.
2. *Data Layers and scrapers: GitHub, GHTorrent*. There are two specific services: GitHub, the repository for many software projects in development, both open licensed and closed, containing the actual data the study uses, and GHTorrent, a third-party service that offers "a scalable, queriable, offline mirror of data offered through the GitHub REST API" (The GHTorrent project). This service provides a means of accessing GitHub data that would otherwise be difficult to scrape, given the query limitations of the GitHub API on its own. The authors of the gender bias study made use of this service in order to extract a large corpus of user data, supplementing it with some scripts for scraping material that was not included in the torrent. GitHub is not the only repository of software development, but it provided a very useful binary metric for studying contributor activity: namely, acceptance of pull requests.
3. *Google Plus: identifying gender*. In the stack, we also have a novel response to the problem of identifying gender identity in GitHub user data, which is that users often do not make use of readily identifiable gender classes in their usernames. The authors, in an attempt to remedy this issue, created a method for linking GitHub user accounts with their associated Google Plus email accounts, thereby allowing a wider source of identifying information (e.g., "real" names, user uploaded images, Google's required "gender" dropdown during account setup, etc). In this layer of the stack, categorizing the gender performance of online users is cer-

tainly one rail upon which to potentially critique the study from many angles, both technical and social, and may provide fertile ground for a discussion of what it means to "identify" a gender and what privacy concerns result. The latter concern is laid out in the study, where the authors discuss their decision to scrub released data of identifiable gender/user information that might open up individual users to harassment.

4. *Analysis*. The next two layers in the stack are the ones on which most time is probably spent in academic discussions, and it is these we will tease out further. First, the data analysis and methodology layer: the authors made use of both standard tools (R, for example) and some bespoke scripts for interrogating data. As we have described, these choices of technologies (e.g., data analysis software, statistical tests) would form the basis of a discussion of goodness, both technical and methodological, by initial reviewers as they evaluated the appropriateness or fit of technical choices, resulting in a cycling between the analysis layer (layer 4) and dissemination layer (layer 5) that we describe, as comments were taken under consideration in the paper's next revision.
5. *Open access dissemination of scholarship*. PeerJ plays a crucial role in the stack. It is between the layer of data analysis and publication that a moment of transformation occurs: from the study as an activity that makes use of open data to the paper itself as an example of open-access scholarship. This openness not only facilitates the kind of information freedom valued by hackers[64] but also propagates the right to free speech at the level of raw data.
6. *Media layer*. The final layer in the stack is the "media layer," including venues legitimized by their status as news agencies (e.g., the *Washington Post*, the *Guardian*) and the social media commonly considered (from the perspective of academic discourse, in any case) to be frivolous or at least an unlikely source of legitimate discussion, although even this characterization is open to debate given the wide range of scholars who make use of social media to publicize or even conduct their research.

Here, then, are the components of a rhetorical/technological stack for the GitHub study:

- Substrate (programmer activity, associated languages and technologies, open source ideologies and licenses and identity construction, all constituting their own highly complex stacks)
- GitHub (the data itself)
- GHTorrent (and ancillaries: bespoke scrapers to capture missing information)

- Google Plus
- Analysis (looped back down through GitHub recursively)
- PeerJ upload; transition from open data to open access
- Community commentary in PeerJ
- News and social media (e.g., Reddit, Twitter, Facebook, etc.)

If there is one area in which Terrell et.al.'s study proves especially fruitful as a rhetorical figure for study, it is in the way it complicates our understanding of what constitutes *good* authorial practice in the academy, a practice that is both technical and social in its understanding of goodness. Just as data acquisition, cleaning, and analysis makes use of many technological tools and coauthors along the way, so too the work of authoring, peer review, and publication relies on many layers of negotiation with software tools, systems, ideologies of authorship, and hierarchies of knowledge valuation within and outside the academy. Consider the way in which the paper was disseminated: not via traditional peer review and publication in a journal, but rather via the PeerJ system. Such a move brings up issues of legitimacy on several fronts, in a highly stratified academic community that relies on peer review as its primary means of conveying status. In this respect, PeerJ can be characterized as a response to/critique of closed- publication methods in which "blindness" operates to both obfuscate (possibly subjective) editorial processes and provides a seeming rationale for scientific "objectivity." Institutionally, of course, the tension of scholarly legitimacy (peer review) versus soliciting response from multiple communities (open review, sharing) has long played out in every corner of the academy in the form of ardent discussions over tenure review cases, nearly almost exclusively to the detriment of the candidate. In part, this high-stakes value judgment has to do with the institutional performance of expertise, which shifts questions of legitimacy from author to audience (who is qualified to comment? This is the underlying principle of peer review). In the GitHub example, we see particularly stark levels of stratification in peer commentary as we move up and down the stack: everything from "peer" commentary by scholars on PeerJ's system, to subsequent (and varyingly accurate) reporting in the media, to the rough-and-tumble communities of Twitter and Reddit, themselves constituted of varying layers of expertise and opinion. Despite institutional protestations to the contrary, it is almost guaranteed that these layers of extra-expertise "review" will be entered into open debate during academic discussions of the work, by virtue of being open for anyone to read.

But beyond standing as a fairly typical exemplar of how academic expertise is negotiated, the study also lays bare the highly complex nature of interrogating and communicating knowledge in the age of the Stack. In particular, the creation of algorithms designed to automate and communicate portions of the analysis suggests that the way we work with our machines not only involves a notion of technical goodness but also a more expansive understanding of how our technical decisions might constitute a new kind of practice in which machines contribute *as rhetorical actors* to the expertise we ultimately disseminate. Two practices carried out by Terrell et. al. in the process of conducting and publishing the GitHub gender study allow us to understand how a rhetorical stack might function not merely as a means of stratification of data and knowledge practices but rather as a coauthorial communicative metamachine:

1. *Using GitHub as both source and site of practice.* In a neat act of scholarly recursion, the authors chose to make use of GitHub not only as the primary source of their data but also as a means of working through their own authoring (or rather, revision and reauthoring) process. Chris Parnin, one of the researchers on the paper, notes the value of isolating peer review comments into bite-sized chunks,[65] both for purposes of revision and more generally in order to deal with the psychological effects of having to respond to a deluge of commentary: an anxiety any academic can attest to who has received that "Reviewer 2" response. The authors of the study chose to treat peer responses as "Issues," rather than critiques, and made use of GitHub's "Issue Tracker" function in order to discuss and respond to each comment systematically. This had the benefit of providing a kind of authorial distance, transforming the usual academic response to critique into a software-facilitated list of tasks.
2. *Creating machine coauthors.* In addition to using GitHub as a kind of authorial task-tracker, the authors also created bespoke scripts that allowed them to automate parts of their analysis. For example, updates to the study data were communicated via scripts higher up the stack, where new analyses were run automatically in order to generate new conclusion data (e.g., number counts, variance, etc). This is a particularly interesting example of the way information work is transformed into authoring via an automated method, allowing the authors of the scripts and the scripts themselves to act as coauthors, as well as computational collaborators in the search for technically good results. Parnin describes this process as embodying the challenge of "getting from data to paper," characterizing the authors' process as the creation of what he calls a "living paper" in which, as the data changes, so too does the hypothesis, often resulting in the need to update the actual scripts to reflect new evidence.[66]

One line short

Conclusion

Combining rhetorical analysis with a stacking metaphor to discuss movements of data allows us to see more clearly several operations in our search for technical and ethical ideas of goodness:

- how data is necessarily collected, cleaned, and interrogated in stages, using multiple different technological tools;
- how each stage carries its own understanding of what *technical good* constitutes (e.g., good data, good cleaning practices, good processing practices, good analysis practices);
- how each stage carries many concomitant valuations of *ethical and ideological good* as we look at the potential effects of the data (e.g., Will using this data over that provide us with a social good? Is this analysis "good" enough to pass peer review?); and crucially
- how that understanding of what constitutes *good* might then be communicated up and back down the stack, forming a kind of rhetorical situation in which multiple understandings of goodness are negotiated between layers.

As Terrell et. al.'s study on gender bias so clearly shows, research in big data requires the assembly of a series of "stack layers" that are often deployed uniquely in each project configuration because of the particularities of the data, its context, its acquisition, and its operational system. Communicative practices coupled across/up/down the resulting assemblage similarly necessitate the bespoke creation of new additions or plugins to the stack (e.g., unique algorithms customized to specific data sets). But because these layers are nonstandard, it becomes hard to assess the technical goodness of the project as a whole. Instead, we are forced to fall back on assessment of individual layers (e.g., the choice of statistical test, the choice of technologies, the choice of platform) without being able to evaluate how they might complicate the entire stack, and how that might result in the propagation of errors through miscommunication or mismatch between technologies.

The underlying mediations and technical negotiations "under the hood" are then complicated by the addition of "social" layers to the stack. How the work is peer reviewed, published, critiqued, and subsequently reported on in technical/academic or nontechnical/nonacademic arenas influences our evaluations of the elements of *phronesis* that manifest in a project. As a result, competing narratives about what is good encompass such issues as blind versus open peer review, closed journal versus open archive publication, who is permitted to comment and in what forum, debates about the value of scientific communication in the age of click bait, and what *social justice* means and for whom.

Nevertheless, despite these complications, analyzing open access and open data projects using a "stack" model helps us see where the complexities lie in the search for goodness. Rather than limited to method or divorced from data science entirely, we instead see in the stack ongoing persuasive choices made in the constitution of particular views of goodness, as well as its complement, ill. For stacked within the layers, as well amongst the layers themselves, is rhetorical calculation that manifests multiple forms of goodness in this particular moment.[67] As a result, big data, instead of transcending the rhetorical through computation, actually reaffirms rhetoric's centrality. Rather than dealing in matters that "cannot now or in the future be, other than they are," Aristotle argues that rhetoric addresses questions regarding what is probable rather than true. In describing rhetoric as an "offshoot of dialectic and ethical studies," he makes clear that what is considered good is often a matter of persuasion. Although this conception of goodness raises the threat of relativism or, even worse, of a dark art that through fallacious reasoning makes what is ill appear good, Aristotle argues that through education and disposition, the *phronimos*—one who can see what is good both for herself and for others—ensures that human flourishing (*eudaimonia*) is always the ultimate end of action. But as Martha Nussbaum points out, goodness is fragile because humans are fragile: "What we find valuable depends essentially on what we need and how we are limited."[68] By realizing that data analytics encourages more rhetoric, not less, we come to realize the fragility of big data and the continued need to consider how best to achieve human flourishing.

Notes

1. G. W. Leibniz, *Logical Papers: A Selection* (Oxford: Oxford University Press, 1966), quoted in Bertrand Russell, *A Critical Exposition of the Philosophy of Leibniz* (Cambridge, UK: Cambridge University Press, 1900), 169–70.

2. Joseph M. Juran, *Juran on Quality by Design* (New York: Free Press, 1992), 9.

3. See, for example, Ahmad Ghazal et al., "BigBench: Towards an Industry Standard Benchmark for Big Data Analytics," in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data* (ACM, 2013): 1197–1208; Stephen Kaisler et al., "Big Data: Issues and Challenges Moving Forward," in *46th Hawaii International Conference on System Sciences* (IEEE, 2013), 995–1004; Joseph Krall, Tim Menzies, and Misty Davies, "Learning the Task Management Space of an Aircraft Approach Model," in *Modeling in Human-Machine Systems: Challenges for Formal Verification: Papers from the AAAI Spring Symposium* (The AAAI Press, 2014), 92–

97; Paul C. Zikopoulos et al., *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data* (New York: McGraw-Hill, 2011).

4. Peter Groves et al., "The 'Big Data' Revolution in Healthcare: Accelerating Value and Innovation" (Center for US Health System Reform Business Technology Office, 2013); Travis B. Murdoch and Allan S. Detsky, "The Inevitable Application of Big Data to Health Care," *JAMA* 309, no. 13 (April 3, 2013): 1351–52; Wullianallur Raghupathi and Viju Raghupathi, "Big Data Analytics in Healthcare: Promise and Potential," *Health Information Science and Systems* 2, no. 1 (December 2014): 1–10.

5. Debashis Aikat, "Big Data Dilemmas: The Theory and Practice of Ethical Big Data Mining for Socio-Economic Development," in *Ethical Data Mining Applications for Socio-Economic Development*, ed. Hakikur Rahman and Isabel Ramos, (Hershey: IGI Global, 2013, 106–30); Omer Tene and Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," *Northwestern Journal of Technology and Intellectual Property* 11, no. 5 (2013): 239–73; Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding, "Data Mining with Big Data," *IEEE Transactions on Knowledge and Data Engineering* 26, no. 1 (January 2014): 97–107.

6. Markus Christen et al., "Ethical Issues of 'Morality Mining': Moral Identity as a Focus of Data Mining," in *Human Rights and Ethics: Concepts, Methodologies, Tools, and Applications*, ed. Information Resources Management Association (Hershey: IGI Global, 2015), 1146–66.

7. World Health Organization, "Influenza (Seasonal)," last modified January 31, 2018, http://www.who.int/mediacentre/factsheets/fs211/en/.

8. Google, "About Flu Trends," accessed August 15, 2014, https://www.google.org/flutrends/about/.

9. Justin Kintz, "Driving Solutions to Build Smarter Cities," Uber Newsroom, January 13, 2015, https://newsroom.uber.com/us-massachusetts/driving-solutions-to-build-smarter-cities/.

10. Ibid.

11. Aristotle, "Nichomachean Ethics," in *The Complete Works of Aristotle*, vol. 2, ed. Jonathan Barnes (Princeton: Princeton University Press, 1984), 1138a32–33.

12. Andrew McAfee, "Big Data's Biggest Challenge? Convincing People NOT to Trust Their Judgment," *Harvard Business Review*, December 9, 2013, https://hbr.org/2013/12/big-datas-biggest-challenge-convincing-people-not-to-trust-their-judgment.

13. Aristotle, "Nichomachean Ethics," 1140a20–23.

14. Norbert Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, 2nd ed. (Cambridge: MIT Press, 1948), 28.

15. Ibid.

16. Wiener, *Cybernetics*, 29.

17. Grady Booch, "Morality and the Software Architect," *IEEE Software* 25, no. 1 (January–February 2008): 8.

18. Grady Booch, "All Watched Over by Machines of Loving Grace," *IEEE Software* 32, no. 2 (March–April 2015): 20.

19. Jacques Ellul, *The Technological System* (New York: Continuum, 1980); Langdon Winner, *Autonomous Technology* (Cambridge: MIT Press, 1977).

20. Niklas Luhmann, *Social Systems* (Stanford, CA: Stanford University Press, 1995), 136–39.

21. Ibid., 79.

22. Humberto R. Maturana and Francisco J. Varela, *Autopoiesis and Cognition: The Realization of the Living* (Dordrecht: D. Reidel Publishing Company, 1980), 9.

23. Luhmann, *Social Systems*, 3.

24. Niklas Luhmann, *Law as a Social System* (Oxford: Oxford University Press, 2004), 118.

25. Ibid., 94.

26. Niklas Luhmann, "Code of the Moral," *Cardozo Law Review* 14 (1993): 1005.

27. John Rawls, *Political Liberalism* (New York: Columbia University Press, 2003).

28. Luhmann, "Code of the Moral," 1005.

29. Niklas Luhmann, "The Sociology of the Moral and Ethics," *International Sociology* 11, no. 1 (March 1996): 31.

30. Ibid.

31. Alexander R. Galloway, *Protocol: How Control Exists after Decentralization* (Cambridge: MIT Press, 2004), 243.

32. Ibid., 40.

33. Ibid., 245.

34. Luhmann, *Law as a Social System*, 357–80.

35. Luhmann, "The Sociology of the Moral and Ethics," 33.

36. Jeremy Ginsberg et al., "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature* 457 (February 19, 2009): 1012.

37. Ibid.

38. Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Boston: Houghton Mifflin Harcourt, 2013), 55.

39. Ibid., 53.

40. Ibid., 55.

41. Gil Press, "A Very Short History of Data Science," *Forbes* Tech Blog, last updated May 28, 2013, https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#55ad9cc455cf.

42. Ian Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*, 2nd ed. (Cambridge: Cambridge University Press, 2006), 1.

43. Ginsberg et. al., "Detecting Influenza Epidemics," 1012.

44. Declan Butler, "When Google Got Flu Wrong," *Nature* 494 (February 13, 2013): 155–56.

45. David Lazer et al., "The Parable of Google Flu: Traps in Big Data Analysis," *Science* 343, no. 6176 (March 14, 2014): 1203–5.

46. Lee Gomes, "Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts," *IEEE Spectrum*, October 20, 2014, https://spectrum.ieee.org/robotics/artificial-intelligence/machinelearning-maestro-michael-jordan-on-the-delusions-of-big-data-and-other-huge-engineering-efforts.

47. Justin Washtell, "The Single Best Predictive Modeling Technique. Seriously," Analytic Bridge, last updated October 29, 2014, http://www.analyticbridge.com/profiles/blogs/the-single-best-predictive-modeling-technique-seriously.

48. Aristotle, "Rhetoric," *The Complete Works of Aristotle*, vol. 2, ed. Jonathan Barnes (Princeton: Princeton University Press, 1984), 1357a5.

49. See Dilip Parameshwar Goankar, "The Idea of Rhetoric in the Rhetoric of Science," *The Southern Communication Journal* 58, no. 4 (1993): 258–95.

50. Edward Schiappa, "Second Thoughts on the Critiques of Big Rhetoric," *Philosophy and Rhetoric* 34, no. 3 (2001): 260–74.

51. Benjamin H. Bratton, *The Stack: On Software and Sovereignty* (Cambridge: MIT Press, 2015).

52. Ibid., 8.

53. Ibid., 69.

54. Ibid., 61.

55. J. Terrell et al., "Gender Differences and Bias in Open Source: Pull Request Acceptance of Women Versus Men," *PeerJ Preprints* (2016), http://doi.org/10.7287/peerj.preprints.1733v2.

56. Marianne Bertrand and Sendhil Mullainatha, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review* 94, no. 4 (September 2004): 991–1013; Silvia Knobloch-Westerwick and Carroll J. Glynn, "The Matilda Effect—Role Congruity Effects on Scholarly Communication: A Citation Analysis of *Communication Research* and *Journal of Communication* Articles," *Communication Research* 40, no. 1 (February 2013): 3–26.

57. Terrell et al., "Gender Differences and Bias in Open Source," 18.

58. Ibid., 26.

59. Ibid., 27.

60. In Dalmeet Singh Chawla, "Researchers Debate Whether Female Computer Coders Face Bias," *Nature* 530, no. 7590 (February 2016): 257.

61. In Terrell et al., comments section.

62. Yourbasicgeek, "Women Considered Better Coders—But Only If They Hide Their Gender," *Reddit.com*, last updated February 12, 2016, https://redd.it/45f3mh.

63. See Esther Zukerman, "'Why Is Reddit So Anti-Woman?': An Epic Reddit Thread Counts the Ways," *The Atlantic*, July 26, 2012, http://www.theatlantic.com/entertainment/archive/2012/07/why-reddit-so-anti-women-epic-reddit-thread-counts-ways/325357/; Alex Cranz, "The Best Place to Find Stuff on Reddit Is Promoting Misogynistic Garbage," Gizmodo, April 26, 2016, http://gizmodo.com/the-best-place-to-find-stuff-on-reddit-is-promoting-mys-1772992114.

64. E. Gabriella Coleman, "The Political Agnosticism of Free and Open Source Software and the Inadvertent Politics of Contrast," *Anthropological Quarterly* 77, no. 3 (Summer 2004): 507–19; Steven Levy, *Hackers: Heroes of the Computer Revolution* (Sebastopol, CA: O'Reilly, 1994); Jennifer Helene Maher, *Software Evangelism and the Rhetoric of Morality: Coding Justice in a Digital Democracy* (New York: Routledge, 2016).

65. Chris Parnin, conversation with Helen J. Burgess, September 7, 2016.

66. Ibid.

67. Jennifer Maher, "The Artificial Rhetorical Agent and the Computing of Phronesis," *Computational Culture: A Journal of Software Studies* 5 (2016), http://computationalculture.net/article/artificial-rhetorical-agents-and-the-computing-of-phronesis.

68. Martha C. Nussbaum, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (New York: Cambridge University Press, 2001), 342.