After Big Data Failed:
The Enduring Allure of Numbers in the Wake of the 2016 US Election

YANNI LOUKISSAS[1]

GEORGIA TECH

ANNE POLLOCK[2]

GEORGIA TECH

Abstract

When widespread polling failed to accurately predict the 2016 US presidential election, producers and consumers of data didn't abandon faith in numbers. Instead, they have reconfigured their relationships with big data. Producers are formulating redemption narratives, blaming specific datasets or poor interpretation, and the broader reception looks similar. Seeking an explanation for Trump's unexpected victory, news audiences are calling out failed pre-election polling numbers, while at the same time embracing empirically dubious exit polls. This Critical Engagement piece argues that Science and Technology Studies scholarship has prepared us to see that polling errors would not undo the prestige and power of quantitative methods, but rather reveal the intensity of our attachment to data as a readily available arbiter. We show that data's ambivalent qualities make it a durable ground for claims-making, with the capacity to be mobilized to do different kinds of work: blame, exoneration, and broader sense-making.

Keywords

Leading up to the 2016 US presidential election, waves of polling data were brought together from every state in the country. The flood of data seemed to indicate, overwhelmingly, that

[1] Yanni Loukissas, Email: yanni.loukissas@lmc.gatech.edu

[2] Anne Pollock, Email: apollock@gatech.edu

Hillary Clinton would triumph over Donald Trump. However, voting on November 8th produced the opposite outcome. This event might have fostered increased skepticism over the explanatory power of big data, a vague technological and cultural phenomenon defined chiefly in terms of magnitude. And yet, in the immediate wake of the election, we have observed widespread efforts to redeem both data and its manifestation in big data. Pollsters and journalists —the producers and disseminators of election data—have responded by reflecting on and reworking their approach to data instead of rejecting its relevance. In parallel, news audiences— seeking an explanation for Trump's unanticipated victory—are calling out failed pre-election polling numbers while at the same time exalting empirically dubious exit polls.

For our part, we are agnostic about the truth-value of election data. Instead, we want to understand how producers and consumers of data are reconfiguring their relationships with the numbers after November 8th. Moreover, what are the implications of that day for our understanding of data, more generally, as a cultural artifact?

We argue, perhaps counter-intuitively, that data's ambivalent qualities make it a durable ground for claims-making. Consider the way we use the term as both a mass singular (i.e. water) and a plural noun. We routinely say "data is" as well as "data are." The both/and quality of data gives the term a pragmatic flexibility. As a result, we can say data "is" transparent—the ground truth of political discourse—even though we know that, individually, data "are" opaque; we rarely hear the story behind any particular data point. We can also say that data "is" accessible everywhere—thanks to mobile technology—even though we understand that data "are" different in each place they are made and made use of. To put this another way, consider a popular political aphorism attributed to former speaker of the House Tip O'Neill: "all politics are local." We might say the same for data. All data "are" local: made by people and their obedient machines in the technological moment. And yet, we treat data as universal. Indeed, data "is" the basis upon which political claims are widely shared. Because the term data has two faces—one abstract, the other concrete—it can fail us one day and save us the next.

Almost as soon as their predictions were shown to be wrong, producers and analysts began to formulate redemption narratives for their data. These narratives have generally taken one of two forms. Some blame particular data, explaining that individual state polls were off. Others claim that we don't have enough data on elections to reveal a clear pattern. However, even while criticizing specific data sets, such narratives reinforce the belief in big data as an approach. For they implicitly suggest that given enough—and good enough—data, we should be able to create comprehensive views into the past and the future. Drawing on a widely used social media grammar, we characterize this mode of redemption as #NotAllData. Meanwhile, other redemption narratives involved in the production of big data seek to explain away widespread

shock at the election results as a problem of <u>data literacy, rather than data</u>. It's okay, data; it's not you, it's us. Indeed, we agree with these analysts: polling offers probabilities, not results; there are margins of error; results need to be presented and interpreted in nuanced ways; and, more generally, election forecasts require contextual information to be meaningful. As producers and analysts of big data reflect on their work leading up to November 8th, they are finding a variety of believable ways to argue that the data are redeemable. Although it was a "<u>rough night for number crunchers,</u>" writers at the *New York Times* tell us, we shouldn't lose our "faith" in the power of data—for politics, and for sports, science, and business. It's not necessarily surprising that professional data analysts cling to faith in data: their professional roles are at stake. But this attachment is not merely occupational; it is also authentic. Occupations and preoccupations are inescapably intertwined as they connect arguments to the power and prestige of numbers.

Meanwhile, the broader reception of events looked similar. In the search for someone to blame in the immediate wake of the election, the blogosphere and *Op Ed* pages combed over exit polls. In some ways, exit polls are more reliable than pre-election polls, because they ask people to describe what they've just done rather than what they anticipate doing. And yet, they have many of the same problems as other polls: many people decline to participate, and respondents might not answer questions about sensitive matters in a straightforward way. The exit polls also had a problem—their relative smallness— that was different from that which attended the huge numbers of pre-election polling: with fewer than 25,000 respondents, <u>CNN drew from exit polls of just .008% of voters</u>. This is not particularly big data. And yet it has had remarkable symbolic power.

The most prominently mobilized findings have been those that showed a margin for Trump among <u>white women</u> and <u>those with incomes over $50,000 a year</u>. This data has traveled widely without caveats, even though we know little about who was asked, where, and how. Consider: in the US almost everyone claims to be middle class, and so voters might well have overstated their incomes whether inadvertently or consciously to avoid stigma. But commentators took the numbers at face value, and mobilized them to <u>exonerate the poor and working class from blame with respect to Trump's victory</u>. The white women statistic itself is slippery. Although widely publicized at the 53% earlier reported, it currently shows as 52% on the CNN site. We'll never know how (in)accurate the number is. But 53% has become emblematic of broader issues, especially for black feminists who are critical of the shortcomings of <u>non-intersectional white feminism</u> and disappointed with <u>the consistent historical failure of white women to show up in coalition with oppressed people</u>. Debunking the 53% number is beside the point, though, both because it is not technically possible and because the righteous anger does not rest on whether the number accurately conveys what happened on November 8th. Like all

data about race, class, and gender, these exit polls cannot be extracted from broader [durable preoccupations](#) with difference. Data's power does not rest on its underlying truth or falsity, but on its capacity to be mobilized to do many and different kinds of work: exoneration, blame, or broader sense-making.

In this post-election period, scholars in Science and Technology Studies (STS) have a role to play in tracking and unpacking the implications of persistent attachments to data, in its many manifestations: numbers, maps, tables, and data-based claims. Indeed, many scholars have already contributed to this effort, in foundational literature on epistemic [cultures](#) and [communities](#) as well as in contemporary work on [critical data studies](#). Generations of scholarship have prepared us to see that polling errors would not undo the prestige and power of quantitative methods, but rather reveal the intensity of our attachment to data as a readily available arbiter. [Trust in numbers](#) is re/newed.

How will the concepts of data, in general, and big data, in particular, continue to develop in relation to one another? For even as the phenomenon of big data [proliferates](#) into every conceivable domain, it lacks the multi-valence of its progenitor. We don't say "big data are…" Indeed, big data is predominantly seen as a [homogenous category](#), a [universal](#), and an [independent arbiter](#). Meanwhile, data retains its ambiguity, malleability and contestability. To us, the current investment in big data looks like a bubble, unstable over the long term. But the concept of data in varied forms will continue to be workable, not in spite of the vagueness and opacity of the term, but because of it.

Author Biography

Yanni Loukissas is an Assistant Professor of Digital Media in the School of Literature, Media, and Communication at Georgia Tech. His research draws together practices from data studies and data visualization. Before coming to Georgia Tech, he was a lecturer at the Harvard Graduate School of Design, where he co-coordinated the Program in Art, Design and the Public Domain.

He was also a principal at metaLAB, a research project of the Harvard Berkman Center for Internet and Society. He has taught at Cornell, MIT, and the School of the Museum of Fine Arts. He is the author of *Co-Designers: Cultures of Computer Simulation in Architecture* (Routledge, 2012).

Author Biography

Anne Pollock is an Associate Professor of Science, Technology & Society in the School of Literature, Media, and Communication at Georgia Tech. An STS scholar by training, her research explores biomedicine and culture and theories of race and gender. She is particularly interested in the ways in which science and technology are mobilized in social justice projects. She is the author of *Medicating Race: Heart Disease and Durable Preoccupations with Difference* (Duke, 2012), and is currently completing her second book manuscript, which has the working title *Synthesizing Hope: Matter, Knowledge, and Place in South African Drug Discovery*.