

# Interface/

---

Volume 1, Issue 1: *Theorizing the Web 2014*

Article 14

---

8-4-2015

## TMI: Theorizing Big Data: Plenary Session at Theorizing the Web 2014

Mike Pepi

© 2015 Pepi

All rights are reserved by the author(s). For any use of this article beyond fair use (17 U.S.C. Sec. 107), contact the author(s) for permission.

---

### Recommended Citation

Pepi, Mike. (2015). TMI: Theorizing Big Data: Plenary Session at Theorizing the Web 2014. [Review] In Jenny Davis & Nathan Jurgenson (eds.) *Theorizing the Web 2014* [Special Issue]. *Interface* 1.1. 1 - 6. DOI: <http://dx.doi.org/10.7710/2373-4914.1014>

*Interface* (2373-4914) is published annually by The Berglund Center (Pacific University) in collaboration with the Pacific University Libraries.

## *TMI: Theorizing Big Data*

*Plenary Session at Theorizing the Web 2014*

Reviewed by Mike Pepi

What are big data, exactly? Definitions vary. For engineers it may refer to the new parallel computing architecture required to handle terabytes of constantly-updating data from ubiquitous devices and sensors. For others, big data is as a catch-all term for surveillance, social media, or the untold quantifying reach of a society enmeshed in digital transmission. A consensus seems to have been reached. “Big” data must satisfy the “Three V’s”: *Voluminous*—organizations routinely collect, store, and analyze terabytes of data; *Velocity*—data are created and captured in real time, often by pre-programmed means; and perhaps the key element, these data are *Various*—involving the integration of unstructured data sources from previously-unconnected phenomena in order to predict outcomes or detect patterns.

Big data is of interest to theorists, however, due to its double life as a practice, a sociality, and a discourse that promises to upend

assumptions central to epistemology, scholarship, and the political economy. [1] Yet in spite of implicating a wide range of scholarly fields, the narratives surrounding big data often derive from so called “Big Data vendors”, the private, often massive technology companies such as IBM, Intel, or Google. Descriptions, metaphors, and—most troubling—the ethics of big data are wrapped up in the marketing effort surrounding its “buzzword” status. *Theorizing the Web* gathered Zeynep Tufekci, Kate Crawford, Janet Vertesi, and Winter Mason for “TMI: Theorizing Big Data,” a keynote panel tasked with the prescient work of sorting through big data’s implications for politics, ethics, and the consumer outside of the complicity of these status quo narratives.

The panel’s mix of both scholars and practitioners operating in the field provided for a grounding in theory and practice. Janet Vertesi, Assistant Professor of Sociology and Faculty Fellow at the Center for Information Technology Policy, Princeton, focused her talk on a personal experiment that investigated the invisible layer of beacons, bots, and cookies that feed marketing databases. Vertesi decided to use the very marketable news of her pregnancy as an opportunity. By diligently avoiding any indication of her pregnancy that could be tracked by servers, she revealed the ever-expanding regime of online and offline tracking.

Due to the valuable purchasing decisions new parents are about to make, online marketers can pay up to 1,000% more for pregnant women’s browsing data. Vertesi’s tactics involved Tor (the untraceable browser better known for illicit trading) as well as strict avoidance of any text-based communication on social media, email, or any platform

that might disclose her pregnancy. This extensive attempt to “opt out” revealed what Vertesi calls our often overlooked “relationship with the server.” Many studies look at how people interact with each other, but few take into account the economic impact of the data we routinely hand over to servers, or how the accumulation of these data inscribes new categories. This server layer increasingly determines what counts as moral behavior. This was illustrated when her husband received a suspicious behavior notification after attempting to repeatedly purchase anonymous prepaid gift cards—or in Vertesi’s words “How purchasing a stroller can make you look like a criminal.” “To be moral is to participate in that regime” and to step outside of that is to appear to appear illicit. Vertesi called for greater options for describing the network, or for new technologies that allow a “transaction to be just a transaction.” Vertesi offered a biting refutation to the libertarian-leaning, free-market ideology of Silicon Valley types that claim that a benevolent and autonomous market will just correct itself: “if opting out makes you look like a criminal, then we are no longer talking about a free market, then we are talking about coercion.”

Kate Crawford’s presentation turned to the question of affect, searching for the “lived reality” of big data, one that she argues is “suffused with a kind of surveillant anxiety.” Crawford decoded the leaked powerpoint presentations from a GCHQ data collection program, code-named “Squeaky Dolphin”, which “outlines an expansionist program” of what they call their “human science operations cell.”

Big data is itself is an emerging Weltanschauung, Crawford asserts. This is a “tale of anxiety” that is captured in big data’s very infrastructures.

An epistemological limit point where “the data itself is not enough.” The lived reality of big data then carries with it a twin anxiety: The anxiety of the surveilled is matched, if not outdone, by that of the ones doing the surveillance. “The fear that there can never be enough data” is mutual with the fear “that one is standing out in the data.”

For Crawford, the anxiety of the surveilled is further evidenced by recent phenomena of “acting basic,” the trend also referred to as Normcore as launched into the lexicon by K-Hole, a collective of artists and designers that toes the line between style forecasters and social commentary. Rather than standing out in a crowd, forward thinking “art kids” have taken to donning mainstream brands, which Crawford sees as one of the many cultural manifestations of mass surveillance. Hard core encryption and acting basic are both ways to blend in, yet both only serve to heighten the thirst for data collection. The resulting “affective residue” is a neo-positivist mode that claims that data speaks for itself alongside a consumer tactic of avoiding detection through a sort of sameness, both which “shed their own subjectivity.”

Winter Mason, a Data Scientist at Facebook, spoke last, implicitly advocating for the benefits of data collection. He made a distinction among informative targeted advertising that acts as a service and traditional brand advertising that simply aims at generating positive emotions around products. Mason addressed the general suspicion of data collection methods by online platforms, including his employer’s, by noting several instances where it has been used for “social good.” Mason cited the “I Voted” experiment Facebook ran during the 2010

US congressional elections that resulted in a “statistically significant” increase in voter participation. This provoked reactions from several people, including panel moderator, Zeynep Tufekci, who pointed out how such behavioral modification carries a potential for abuses if run at tremendous scale.

The panel was welcome inquiry into the ideology at work among advocates of big data operations. If the original use of the term “data” arose alongside “modernity and the growth and evolution of science,” then today “Big” data can be tied to Silicon Valley solutionism and technological determinism whose true impact is still under development and scrutiny. As nearly every panelist pointed out, separating behavior into “online” and “offline” not only fundamentally mis-recognizes the foundations of big data regimes, but also serves to abstract ethical questions, doing harm to critical perspectives of massive analytical endeavors that are often invisible to its subjects.

# Notes

[1] Patricia Clough has written of a “Datalogical Turn” in the field of Sociology. See: Clough, P., K. Gregory, B. Haber, & J. Scannell. (Forthcoming). Nonrepresentation at the Scale of Big Data: A Challenge to the Sociological Imagination.” In *Nonrepresentational Methodologies: Re-Envisioning Research*, ed. Phillip Vannini. Oxford: Taylor & Francis.

# References

Rosenberg, D. (2013) ‘Data before the fact’, in L. Gitelman (ed.), *‘Raw Data’ is an Oxymoron*. MIT Press, Cambridge, MA, pp. 15–40, quoted in Kitchin, R. (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications Ltd, London, pp. 3