Surplus Data: An Introduction

Orit Halpern, Patrick Jagoda, Jeffrey West Kirkwood, and Leif Weatherby

It is no longer enough to say that data is big. Data is now in a state of surplus. As we have progressed from the megabyte to the terabyte, the petabyte, and now in 2022 debatably to the zettabyte era—all within the span of a mere two decades—we have witnessed a quantitative increase manifesting itself as a qualitative change. In a well-known 2008 provocation, the editor in chief of *Wired*, Chris Anderson, announced that the ability to produce and analyze enormous data sets using artificial intelligence (AI) was rendering the bedrock of human knowledge systems—the scientific method itself—obsolete. For the first time in history, correlation began to supersede causation and science advanced "without coherent models, unified theories, or really any mechanistic explanation at all."[1] It supposedly spelled the end of theory. But theory has faked its own death many times.[2]

We call attention to this much-cited end to highlight that such statements harbor traces of a historical change in governmentality, epistemology, and political economy. They gesture toward evolutionary changes in the relationships between the nature of data and the governance of populations in our present.

1. Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, 23 June 2008, www.wired.com/2008/06/pb-theory/

2. The charge that theory is over has been made frequently, often with the objective of replacing it with some putative science, as in the recent case of the digital humanities. But this cycle has a much longer history. Starting with Friedrich Nietzsche, philosophy has repeatedly asserted its own death, perhaps most emphatically in Martin Heidegger's own reading of Nietzsche in the 1930s; see Martin Heidegger, *The End of Philosophy*, trans. Joan Stambaugh (Chicago, 2003). Deconstruction figured itself in the margins of that death and then also appeared to die; see Jeffrey Williams, "The Death of Deconstruction, the End of Theory, and Other Ominous Rumors," *Narrative* 4 (Jan. 1996): 17–35. Beyond critical theory, there are also other kinds of announcements of theory's end; see, for example, Victor Burgin, *The End of Art Theory: Criticism and Postmodernity* (London, 1986). As these examples suggest, theory is as unlikely to end as history.

Anderson's provocation implies, whether he intends it or not, that we live in a world of noncausal relations, inductive reasoning, incoherence, and multiple truths (or fakes). We call this condition, in which data no longer serves just as an abstraction from the "real" world but rather also as description and material agent, not big but surplus data.

The passage from *big* to *surplus* has produced surprising effects. Our contemporary moment marks a new episteme in the histories of data collection and the management of populations that rely on "assemblages of data layers and potentials, densities and probabilities that are never linear, causal, or inevitable."[3] While there are many histories of big data, we highlight for our purposes a genealogy that begins in the late nineteenth century when data became the preserve of entities that included the US Census Bureau and the Hollerith tabulating machine. By the end of the twentieth century, this largely state-controlled world expanded into privatized big data in the wake of the personal computing revolution and the emergence of the World Wide Web. Popularized by computer scientists such as John Mashey in the 1990s, "'big data'" became pervasive as the first decade of the twenty-first century drew to a close.[4] Personal computing and the ubiquity of networked

3. Orit Halpern, *Beautiful Data: A History of Vision and Reason since 1945* (Durham, N.C., 2014), p. 34.
4. Steve Lohr, "The Origins of 'Big Data': An Etymological Detective Story," *New York Times*, 1 Feb. 2013, bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/

Orit Halpern is an associate professor in sociology and anthropology at Concordia University. She is also the director of the Speculative Life Research Cluster and D⁴: The Disrupting Design Research Group, laboratories bridging the arts, environmental sciences, media, and the social sciences. She is the author of *Beautiful Data: A History of Vision and Reason since 1945* (2015). Patrick Jagoda is a professor in the Departments of English Language and Literature, Cinema and Media Studies, Obstetrics and Gynecology, and the College at the University of Chicago. He is executive editor of *Critical Inquiry* and director of the Weston Game Lab. He is the author of *Network Aesthetics* (2016), *The Game Worlds of Jason Rohrer* (2016, cowritten with Michael Maizels), and *Experimental Games: Critique, Play, and Design in the Age of Gamification* (2020). He is also a recipient of a 2020 Guggenheim Fellowship. Jeffrey West Kirkwood is an assistant professor in the Department of Art History at Binghamton University, State University of New York. He is the author of *Endless Intervals: Cinema, Psychology, and Semiotechnics around 1900* (2022). Leif Weatherby is associate professor of German and director of the Digital Theory Lab at New York University. He is the author of *Transplanting the Metaphysical Organ: German Romanticism between Leibniz and Marx* (2016).

connections made the promise of the extension of data seem palpable. But it was the processors of that data that shifted *big* into *surplus*. In the US, this included massive networked platforms (for example, Facebook, 2004;, and Twitter, 2006) and neural networks (for example, Fei-Fei Li's ImageNet, 2009, and her team's breakthrough image-recognition algorithm, 2012) that rely on big data and led to Google's conversion from search engine to all-purpose data behemoth. Perhaps a *terminus ante quem* could be marked with Amazon's sudden realization of large, seemingly sustainable profits with the meteoric rise of Amazon Web Services, solidified in the fourth fiscal quarter of 2017.[5] These changes mark a shift from the gathering to the redeployment of data, from accumulation to feedback. Data has become a source of capital dynamics, a means of governance and control of populations, a mode of the administration of territory, in short, a new structural condition.

Data was once a stable, recorded point for static reference; it became, with always-on computing and the new Internet of Things, a structural condition of capital itself. The capacity for these systems to use population-scale data not merely to describe and represent worlds but to generate new ones has evolved in a holistic infrastructure made up of both technologies and epistemologies.[6] Haunted by the histories of colonialism, race, and population in which statistics had its modern origins, surplus data is both continuation and departure.[7] This special issue seeks to outline what is new without losing the evolutionary picture, which extends into the future.

The paradigm *big* is a quantitative designation of data itself; surplus is the *quality* of the social after the quantitative surge. This transformation from quantity to quality demands a new conception of the relationship between what is being measured and represented by data and its efficacy and impact on the world.[8] The Google research scientists Alon Halevy, Peter Norvig,

5. See Jason Del Rey, "Amazon Has Posted a Profit for 11 Straight Quarters—Including a Record $1.9 Billion during the Holidays," *Vox*, 1 Feb. 2018, www.vox.com/2018/2/1/16961598/amazon-jeff -bezos-record-profit-11-quarter-q4-2017-earnings

6. Just as Charles Darwin's concept of evolution changed the epistemologies of time and population in the nineteenth century, extending a scientific theory into a general social apparatus, so data does the same today. More recently, Melinda Cooper has described emergent practices of contemporary "neoliberal biopolitics" that outstrip older reproductive population-stabilizing forms of governmentality to produce an exploitable "surplus of life" as a resource itself (Melinda Cooper, *Life as Surplus: Biotechnology and Capitalism in the Neoliberal Era* [Seattle, 2008], pp. 8, 49).

7. For more on the connection between colonialism and data extraction, see for example Nick Couldry and Ulises A. Mejias, "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject," *Television & New Media* 20 (May 2019): 336–49. For more about how surveillance, policing, biometrics, and data reify categories of race and blackness, see Simone Browne, *Dark Matters: On the Surveillance of Blackness* (Durham, N.C., 2015).

8. Louise Amoore, for instance, embraces Tim Estes's phrase, "'aperture of observation,'" to describe how, in processing the enormous data enabled by cloud computing, a "machine learning algorithm condenses the output of multiple layers to a single output," uncovering "attributes and

and Fernando Pereira have noted that there is an "unreasonable effectiveness" that big data introduces.[9] Data does more than it was intended to and often produces greater effects than mere measurement of a world that stands externally to it. Put simply, in our era, data is not simply descriptive or analytical but actively constructive. The assumption that complex data sets yield the most comprehensive truth returns us to the problem of theory and a question of history. This situation poses fundamental questions: Do we have the terms and concepts to account for such "unreasonable effectiveness" of large data sets that allegedly yields better results than the "elegant theories" pursued in earlier periods?[10] Are there ways to complicate the widespread belief that data is valuable (just like minerals or energy resources) and generative of a more robust reality? Or has data become so complex in its constantly growing and changing surplus that conventional critique is now untenable, unable to keep up as data claims to produce more and more insight?

While well intentioned, discourses of data extractionism affirm rather than critique data's status as sovereign and representative of the world. Shoshana Zuboff, typifying such talk of data as "extraction," characterizes data as a tool of "surveillance capitalism" that exploits a "behavioral surplus" in recording the activities of consumers.[11] Google, in the story she tells, suddenly realizes that it possesses patterns of that behavior, a gold mine collected unwittingly under the early corporate slogan: "Don't be evil."[12] But it is not behavior that is in surplus, even if behavior recorded bolsters surplus value. What Google has achieved is the transformation of a finite, if extremely large, resource into a seemingly endless source of value through the recombination and discovery of new relations and patterns in the same data set. Finitude becomes a flexible frontier through new techniques of data analytics and machine learning. These new social "facts" that emerge and produce new types of interaction thus create social relations from *within* the system. Google

---

relations that would not otherwise be perceptible," giving the purely quantitative a sensible shape that is as much a reflection of the myriad computational parameters as it is of the objects it intends to define (Louise Amoore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* [Durham, N.C., 2020], p. 43).

9. See Alon Halevy, Peter Norvig, and Fernando Pereira, "The Unreasonable Effectiveness of Data," *IEEE Intelligent Systems* 24 (Mar.–Apr. 2009): 8. The phrase is borrowed from Eugene Wigner's famous notion that mathematics is unreasonably effective in the physical sciences; see Eugene P. Wigner, "The Unreasonable Effectiveness of Mathematics in the Natural Sciences," *Communications on Pure and Applied Mathematics* 13 (Feb. 1960): 1–14.

10. Halevy, Norvig, and Pereira, "The Unreasonable Effectiveness of Data," p. 8.

11. Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York, 2019), pp. 63–97.

12. "Google Code of Conduct," Alphabet Investor Relations, 25 Apr. 2012, available as a web page capture on *Internet Archive*, web.archive.org/web/20160202030302/https://abc.xyz/investor /other/google-code-of-conduct.html

is constantly deriving more value by optimizing the functional relations between data sets. There is always a little bit more to obtain from an existing dataset.

This form of derivation and optimization is based on the extension of data's life beyond the use envisioned in its initial gathering. The blue dot that marks your place on Google Maps serves both to locate you and to operate as a source of insight into ongoing and future traffic patterns. In this respect, data is in an analogous position to what Karl Marx called "surplus labor."[13] As Marx argued, in capitalism value is in perpetual, constitutive surplus—self-generating, an "animated monster" or a goose laying golden eggs—value that is always *more*, or *Mehrwert*, as Marx translated the English term (*C*, p. 302). The surface phenomenon of growth harbors, for Marx, the secret of surplus labor. What Marx calls the valorization process is nothing more than an extension of the working day beyond the time required to replace the value put into production as constant capital (materials, overhead, and so forth)—labor in surplus. Labor confers value on the materials it transforms, but "valorization [*Verwertung*]" is the secret of the "value-form," the internally contradictory dual values of use and exchange in the commodity (*C*, pp. 252, 138). This process of valorization is "nothing but the continuation of the [process of creating value] beyond a definite point" (*C*, p. 302). Just when labor operates to build capital rather than replace it, it becomes surplus, excess labor from the standpoint of the worker yet requisite labor for the growth of capital. The quality of labor, extended beyond this point, becomes itself "quantitative" (*C*, p. 305). The dialectical conversion gives us the character of modern labor: done for quantity (growth of capital) by quantity (extension of the working day), labor itself becomes quantity, measured as it is in hours or, increasingly by software, in quarters of an hour or minutes. That granular measurement is, of course, an operation of data today. And data, collected to pinpoint and describe some event, action, or identity, is now tasked with *more*; joining a sea of other data points, it becomes a source for the constant derivation for "better" insight, "more efficient" systems. *Big* does not capture this extension, much less its precipitate in the social. Always immanently providing *more*, data is in surplus.

When data is accumulated, warehoused, reused—extended beyond its initial recording of some individual point—it tends towards surplus. Global logistics, finance, and even the mundane details of what we see on everyday web pages as we navigate from point A to point B—all of this is generated by this internal *more* that data affords. This *more* is not necessarily more value. As the

---

13. See Karl Marx, *Capital: A Critique of Political Economy*, trans. Ben Fowkes, vol. 1 of 3 vols., trans. Fowkes and David Fernbach (London, 1990); hereafter abbreviated *C*.

economist Robert Solow once remarked dryly, one sees the digital revolution "everywhere but in the productivity statistics."[14] Perhaps surplus data merely shuffles a stagnating capital back and forth, concentrating it in patient forms of monopoly capital around the largest data stakeholders, such as Amazon and Google.[15] But the question of the health of capital is not the point. Data begins its life as quantification, and *surplus* designates its conversion into a quality. The buzzword *data driven* captures this reverse dialectical motion. *Big* was a promise; *surplus* is the condition. This issue explores and critiques the state of data in surplus. In the remainder of the introduction, we highlight the logics of derivation and optimization that surplus data has realized across a wide swath of social phenomena.

Surplus Epistemology

Concerns about big data have hinged on an enduring faith in the relationship between data and reality and ongoing concerns about false inferences from group to individual (commonly known as the ecological fallacy). The computer scientists Christian Calude and Giuseppe Longo, for example, have argued that the "bigger the database which one mines for correlations, the higher is the chance to find recurrent regularities and the higher is the risk of committing" fallacies amid the "deluge of data."[16] This fear of "spurious correlations," however, is still vested in the idea that data's essential vocation is to help reveal a testable theory of the world that comes increasingly closer to that world's actual form.[17] From this perspective, the function of data is descriptive. The notion of surplus, by distinction, designates a function beyond description. Even spurious correlations, once understood as the flotsam of mass datafication, get reinserted into the world as the data mounts and the patterns proliferate. Luciana Parisi, in this issue, suggests that we might label such actions "recursive"; data is reassembled and recombined to produce new

14. Robert M. Solow, "We'd Better Watch Out," *New York Times Book Review*, 12 July 1987, p. 36.

15. See K. Sabeel Rahman and Kathleen Thelen, "The Rise of the Platform Business Model and the Transformation of Twenty-First-Century Capitalism," *Politics & Society* 47 (June 2019): 177–204.

16. Cristian S. Calude and Giuseppe Longo, "The Deluge of Spurious Correlations in Big Data," *Foundations of Science* 22 (Sept. 2017): 609, 596. A wave of critique has pointed to the high stakes of the algorithmic perpetuation of these "correlations," which tend to reproduce the worst economic and racial inequalities our society has on offer; see ibid. See, for example, Ruha Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code* (Medford, Mass., 2019); Ramon Amaro, *The Black Technical Object: On Machine Learning and the Aspiration of Black Being* (forthcoming); and Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York, 2017).

17. See Calude and Longo, "The Deluge of Spurious Correlations in Big Data."

"truths" from within the system.[18] Data extends beyond description, creating the world it would describe.

The new condition of data has led to a transformation of statistical thought. As Antoinette Rouvroy and Thomas Berns note, statistical data once consisted of pieces of information that were assumed to reveal a reality but were simultaneously always contestable because of limited "realities" and singular truths.[19] In fact, the debate over such finite facts, they argue, was the underlying protocol of the Euro-American vision of the public sphere. Even if the facts were debatable, the rules governing the debate were not. And this constitution or protocol for debating data comprised the public commons. In our present, they maintain, this space and these rules no longer exist. Rather than capturing an individual, algorithms processing large data loads create "dividuals," in Gilles Deleuze's usage.[20] Dividuals generate countless data points that correspond with tastes, desires, behaviors, and affects, instead of self-consistent individuals. Thus, contemporary understandings of data undermine previous assumptions about reality, the public sphere, and truth.

The social and political consequences of the drift into a sea of predictive and diagnostic correlations that shape the indeterminate future with possible futures have been immense. Consider that police departments now frequently employ experts in big data who help to define a terrain of possible criminality according to algorithmically determined patterns in personal and criminal records and social media. For instance, the Chicago Police Department used an algorithm developed at the Illinois Institute of Technology to create a "'heat list'" that ranked potential shooters and victims that was used to investigate violent crimes.[21] The list "worked": "80 percent of the 51 people shot over two

18. See Luciana Parisi, "Recursive Philosophy and Negative Machines," *Critical Inquiry* 48 (Winter 2022): 360–80.

19. Antoinette Rouvroy and Thomas Berns, "Algorithmic Governmentality and Prospects of Emancipation: Disparateness as a Precondition for Individuation through Relationships?" trans. Elizabeth Libbrecht, *Réseaux* 177, no. 1 (2013): www.cairn-int.info/article-E_RES_177_0163–algorithmic -governmentality-and-prospect.htm. Also, this point contrasts with more familiar questions about insensitivity to sample sizes that have been taken up in behavioral economics; see, for example, Amos Tversky and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185 (Sept. 1974): 1124–31. In particular, Amos Tversky and Daniel Kahneman raise the opposite problem of overly small sample sizes. Small samples can produce faulty judgments and errors regarding the representativeness of data and the availability of scenarios that support a particular theory. Of course, behavioral economics draws more from microeconomic theory and psychology, which focus on issues such as individual-bounded rationality and processes by which people make choices. The kinds of problems and opportunities introduced by the type of aggregation proper to big data (and the condition of surplus data that interests us here) operate more regularly at the level of macroeconomics.

20. Gilles Deleuze, "Postscript on Control Societies," in *Negotiations, 1972–1990*, trans. Martin Joughin (New York, 1995), p. 180.

21. Andrew Guthrie Ferguson, "Beyond Data-Driven Policing," *American Scientist* 105 (Nov.–Dec. 2017), www.americanscientist.org/article/beyond-data-driven-policing

days" during Mother's Day weekend in 2016 were "correctly identified" on the list and "78 percent of the 64 people shot" over Memorial Day of that same year were also on the list.[22] But the discourse of data "correctness" in such situations veils the self-justifying nature of the enterprise, which can always find further correlations to affirm previous correlations. The artificial intelligence researcher Kate Crawford has similarly noted that the "drive to accumulate and circulate is the powerful underlying ideology of data" and "machine learning models require ongoing flows of data to become more accurate."[23] Accuracy is an operation that requires more data, yet more data undermines the very paradigm of accuracy. As the artist Hito Steyerl writes, under conditions of big data "veracity is no longer produced by verifying facts."[24] The plausibility furnished by correlation abandons accuracy for a conspiratorial paranoia that finds every future in a past assembled from empirical patterns. Fake news, conspiracy thinking and practice, police brutality, and speculative market bubbles are all the logical products of surplus data.

Surplus Derivation

"Data is the new capital asset of the 21st century," announces Tom Wheeler, former chairman of the US Federal Communications Commission, commenting on the rise of Amazon over companies like Walmart.[25] We can further extend this line of thinking to consider Facebook, Alphabet, and Twitter's role in the Capitol riots of January 2021. Democratic members of Congress have suggested that the mayhem that day was driven by informational excesses, whose exploitation was responsible for simultaneously destabilizing the American political system and generating a huge windfall for the largest tech companies. According to Wheeler, such situations lay bare the inadequacy of old regulatory concepts for capturing new technological, social, and commercial realities. The regulation that Wheeler and others are accostomed to is based on "industrial antitrust, anti-centralization kinds of concepts."[26] What Wheeler suggests is that our contemporary situation in both politics and economy no longer functions according to the ideals of efficiency, energy, and scarcity that preoccupied industrial economies. Surplus data is the condition that Wheeler

---

22. Ibid.

23. Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven, Conn., 2021), p. 114.

24. Hito Steyerl, "A Sea of Data: Pattern Recognition and Corporate Animism (Forked Version)," in Clemens Apprich et al., *Pattern Discrimination*, ed. Götz Bachman et al. (Minneapolis, 2018), p. 5.

25. Quoted in Klint Finley, "Former FCC Chair Tom Wheeler Says the Internet Needs Regulation," *Wired*, 27 Feb. 2019, www.wired.com/story/former-fcc-chair-tom-wheeler-says-internet-needs-regulation/

26. Ibid.

places beyond the industrial, and its paradigm is *derivation*. It was once the imagined limits to resources and energy that shaped industrial conceptions of efficiency, energy, and labor power.[27] In the early twenty-first century, data capitalism changes this formula by putting the derivative before the source. Derivation takes the place of extraction, and where there was efficiency, there is now *optimization*.[28]

We glimpse the centrality of such inefficiency and derivation in the high-profile case of the r/wallstreetbets subreddit, whose members in January 2021 (and again in February and again in June) strategically bought up shares of dying brick-and-mortar companies, such as GameStop and AMC Theatres, which had high levels of short interest. These actions triggered a massive short squeeze that nearly drove some hedge funds, like Melvin Capital, out of business. The improbably parabolic price movement was made possible by ferreting out the unhedged positions of (ironically) hedge funds in the share interest data and mobilizing a vast army of traders invisibly in plain sight. What had come to feel like a guarantee of endless surplus to mega-money investment firms was, in a matter of days, undone by a data overload in the form of digital buy orders sent by retail traders on desktop and smartphone trading apps. The amount of trading data was so great that it created liquidity problems for brokerages, who decided to block buying of some popular meme stocks at various times. Conspiracy speculation took root on the Reddit boards, which then passed to mainstream attention and finally to hearings in Congress.

As this case demonstrates, the actions of the masses are now a resource for capital. Robinhood, a trading app launched in 2015 that advertises a dark utopian mission to "democratize finance for all," offers commission-free trading and became the popular vehicle for the retail traders who joined the GameStop mania.[29] But, as Richard Serra and Carlota Fay Schoolman contended in their 1973 piece, *Television Delivers People*, producing a statement that has since become a foundational principle of media studies: when something is free, you are not the consumer, "you are consumed."[30] And sure enough, Robinhood makes much of its money from selling traders' order flow data to market makers like Citadel, whose CEO had invested $2 billion in Melvin Capital, the very hedge fund that was caught in the short squeeze. Beneath the David and Goliath story of Main Street investors sticking it to Wall Street villains was a

27. See Jeffrey West Kirkwood, "From Work to Proof of Work: Meaning and Value after Blockchain," *Critical Inquiry* 48 (Winter 2022): 313–33. For industrial-era concepts, see Anson Rabinbach, *The Human Motor: Energy, Fatigue, and the Origins of Modernity* (Berkeley, 1992).

28. These notions of derivation and optimization are partially based on Halpern and Robert Mitchell, *The Smartness Mandate* (forthcoming).

29. "About Us," Robinhood, robinhood.com/us/en/about-us/

30. Richard Serra and Carlota Fay Schoolman, *Television Delivers People* (1973), YouTube, www.youtube.com/watch?v=LvZYwaQlJsg&ab_channel=KunstSpektrum

more nefarious revelation that the real surplus at work in the meme stock affair was reaped as data that helped shore up the more traditional forms of surplus among big institutional firms that control the very contours of a supposedly free market. Moreover, the qualitative, affective response to such market dynamics, as recorded on Reddit and Twitter, have now become a tactical resource of hedge funds, who have learned to profit from even the best attacks against them. Quantitative trading algorithms analyzing massive amounts of social media data using advanced natural language processing are deployed to perform *sentiment analysis* and *opinion mining*. And so the cycle of surplus continues from data to affect to data, ad infinitum—each derived from the last with the derivative more fundamental than the putative source of derivation.

Surplus Politics

During the COVID-19 pandemic, an unprecedented portion of the population was confined to their homes, producing and consuming data in a state of hermetic globalism, straining the already overloaded bandwidth of global data transfer.[31] On 6 January 2021, a group of right-wing supporters of Donald Trump attacked the Capitol building in Washington, D.C., fueled by the conspiracy theory of the group QAnon, a widespread online network surrounding a putative source high up in the "deep state" (the figure known as Q) and propagating racist, anti-Semitic, and xenophobic propaganda. As we see in Cullen Hoback's documentary about the movement, *Q: Into the Storm* (2021), Q operates on the suspicion that the truth is in hidden byways of digital data, sometimes yielding deadly consequences. To witness Hoback accompany Jim Watkins—a businessman and the operator of 8Chan, the main platform on which Q, an alleged intelligence officer, posted his "drops"—laughing as the crowd breaks into the Capitol building is to see the conflation of the digital and the social all too directly.[32] Q has created a semiotic world of clues that severs itself and its followers from the fabric of social reality altogether, gamifying it as Hoback suggests in a comparison to *Cicada 3301*, alternately characterized as an actual secretive organization or a fictional alternate reality game that has run complex digital scavenger hunts since 2012.[33] Q's game indeed has

31. This state of safe confinement did not extend, of course, to service workers, especially those whose labor runs along the channels of surplus data.

32. See "The Storm," 4 Apr. 2021, *Q: Into the Storm*.

33. Beginning in 2020, the comparisons between QAnon and alternate reality games began to proliferate. See, for example, Reed Berkowitz, "A Game Designer's Analysis of QAnon: Playing with Reality," *Medium*, 30 Sept. 2020, medium.com/curiouserinstitute/a-game-designers-analysis-of -qanon-580972548be5; Clive Thompson, "QAnon Is like a Game—A Most Dangerous Game," *Wired*, 22 Sept. 2020, www.wired.com/story/qanon-most-dangerous-multiplatform-game/; and Patrick Jagoda and Kristen Schilt, "How Alternate Reality Games Are Changing the Real World," *Big Brains* podcast, 10 Dec. 2020, www.youtube.com/watch?v=J0CBJ4vC9bU&ab_channel=TheUniversity ofChicago

rules, a perverse affective sense of fun, and easter eggs that provide domesticated surprise. QAnon's slogan "'do your own research'" might be taken as a command to surf your own surplus data channels.[34] And the Q movement has one thing right: data *is* worldly; digital channels *do* shape the world and are in excess of any heuristic intent. Events like the Capitol riot reify the data surround, among other things giving rationale to the increasingly datafied police to expand their quantitative vision.[35] The events themselves are shocking and somehow predictable all at once: it is as though image boards (4Chan, 8Chan, 8kun) premeditate events by sniffing them out of the back alleys of data and insinuating them into reality.

This eruption of conspiratorial violence reminds us that data has inherited the legacy of biopolitics, particularizing its manipulation of society as a mass. As Rob Kitchin has argued, it is not just size that makes data *big*. Even speed of transfer and variety of format make up necessary but insufficient conditions for the revolution we were promised. Data deserving the name *big* also has to be "flexible" and "relational"—open to the inclusion of new fields—and, crucially, both "exhaustive" and "fine-grained."[36] The usefulness of data was traditionally attached to the precision with which it was gathered and defined. Sparse data, very exact, could create predictions to guide action by means of averages. The resulting categories, like those in an actuarial table, did not apply to individuals directly but at the level of the mass. This type of data was a crucial technique of what Michel Foucault called *biopolitics*, governance not of the individual body but at the level of generality. However, if biopolitics still relied on the assumed reality of demographic data, surplus data is something entirely novel. What was once a disjunction between individual and mean has become a partly automated loop between machine vision (or more generally, categorization) and its application to singular states of affairs. This logic stretches from FICO scores to healthcare data, from global logistics to finance capital.

Data has indeed become big and granular, and it has gained the ability to move from particulars to generalities and back again. Ecological fallacies emerging from large data sets now simply become new sources of value in both markets and politics. Without norms or quantifiable risks, we enter endless loops of uncertainty. David Bering-Porter, in his contribution to this issue, juxtaposes W. E. B. Du Bois's data visualizations and speculative fictions with

---

34. Berkowitz, "A Game Designer's Analysis of QAnon."
35. See Hannah Harris Green, "How US Capitol Attack Surveillance Methods Could Be Used against Protesters," *Guardian*, 7 Apr. 2021, www.theguardian.com/us-news/2021/apr/07/us-capitol-attack-surveillance-methods-protesters-first-amendment
36. Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences* (Los Angeles, 2014), p. 68.

the famous case of Judge Schreber's paranoid fantasies. Extrapolating into our present, we might imagine the paranoid conspiratorial politics of QAnon as occupying the space of paranoic dreams, ones of absolute counting, datafication, and control of the future, aspirations whose impossibility always drives violent forms of speculation and politics. But, Bering-Porter suggests, there are other pathways available. In the quantitative countermyths put forth by Du Bois to document racism in America, there was also an alternative aspiration "to reconcile the aims of visuality and data in two senses: as sight and apparition, evidence and aspiration."[37] Perhaps there is a future in which data stories offer evidence of a reality surplus data seems to foreclose in the present, the reality of the Black lives that Du Bois highlights and that have taken center stage in US politics today. It is the new task of a progressive politics to turn the endless extendable and colonizing frontiers of machine learning systems into something other than conspiratorial derivative instruments. In the surplus of data, any faith in the singularity of the real has been shattered—but these systems might harbor another way to encounter the world, one grounded in the experiences and data of the diverse multitudes. Our machines make technically visible what perhaps has always been there—the social nature of our technical lives. They need only be turned toward that future.

Surplus Data

The articles in this issue seek to situate and conceptualize surplus data as the start of just such a project of accounting and imagining surplus data otherwise. The issue begins with two contributions that interrogate the nature of data itself. Alexander Galloway's "Golden Age of Analog" questions the theoretical regime that has come to view digital phenomena from the standpoint of continuity, arguing that the notion of number, beginning with Euclid, offers resources that expand beyond the analog turn. He advocates a frank return to a symbolic analysis in our heavily symbol-mediated world. In "On the Digital Ocean," Sarah Pourciau tracks the history of this divide in modern mathematics, arguing that Alan Turing's implicit model of reality overturns the "continuous" reality that was assumed even in the work of Georg Cantor and underlies a world of data that is always feminized.[38] Data is the name for the symbolic ocean that we find ourselves in, in which *discrete* and *continuous*, *difference*, and *identity* cannot be easily distinguished.

---

37. David Bering-Porter, "Data as Symbolic Form: Datafication and the Imaginary Media of W. E. B. Du Bois," *Critical Inquiry* 48 (Winter 2022): 285.

38. See Sarah Pourciau, "On the Digital Ocean," *Critical Inquiry* 48 (Winter 2022): 233–61.

The symbols of the digital ocean are the carriers of ideology as the next articles show. Bering-Porter's "Data as Symbolic Form: Datafication and the Imaginary Media of W. E. B. Du Bois" argues that data harbor an ideology of reference, seeking to unequivocally name and alter the world. Data visualization for all its obvious contingency carries the force of seeming ontological necessity, leading, in the famous case of Schreber, to madness. Bering-Porter excavates an alternate history of data use in both visualizations and fictional narrative by Du Bois, who attempted to inject quality into the quantity of images of data while critiquing its incipient ideology. Matthew Handelman also investigates the ideology of data but shifts the focus to language. In "Artificial Antisemitism: Critical Theory in the Age of Datafication," he shows how the neural network chatbot Tay learned to deny the Holocaust within hours of its online release. For Handelman, Tay's history, which he tracks through message and image boards, requires a return to and update of the Frankfurt School's method of ideology critique. When quantity becomes qualitative hate speech, the dominance of media in the shape of politics becomes freshly evident, and imitation—which Adorno located in the heart of modern anti-Semitism—is algorithmized, producing more hate. This meditation on algorithmic hatred leads into Luciana Parisi's "Recursive Philosophy and Negative Machines," which explores transcendental instrumentality via the case of Jordan Peele's *Get Out* (2017). Parisi argues in a close reading of the Black horror film that it is in the very heart of learning systems and data-heavy implementations that we should seek an emancipatory "cosmotechnics," in media philosopher Yuk Hui's phrase.[39]

From the heights of number theory to the thickets of datafied identity and representation, surplus data overruns mathematics and computer science but escapes back from the misery of the social into the arid climes of advanced mathematics once again. The final articles turn to the systemic aspects of data in the logistics of capitalism, where quality and quantity have always been mixed. Orit Halpern's "The Future Will Not Be Calculated: Neural Nets, Neoliberalism, and Reactionary Politics" develops a theory of conspiratorial media. She argues that neoliberalism and machine learning share a genealogy through the history of neural networks. Her article traces how concepts of reason, decision-making, and freedom were reformulated in relation to the state and technology from the 1950s through the 1970s. The article outlines how the relationship between neoliberal economic thought, artificial intelligence, and reactionary politics is historically specific and technologically reinforced. In "From Work to Proof of Work: Meaning and Value after Blockchain," Jeffrey West Kirkwood argues that computational capitalism has inverted core

---

39. Quoted in Parisi, "Recursive Philosophy and Negative Machines," p. 327.

value structures of the industrial era. In the designed inefficiency of block-chain technologies Kirkwood locates a new engine of surplus, showing how industrial production's focus on control and optimization has given way to a value born of profligate energy waste and a cryptographic destruction of meaning. Cryptocurrencies, as he contends, offer a paradigm for understanding the transition from industry to computation, from productive surplus value to surplus data. Finally, Leif Weatherby and Brian Justie argue in "Indexical AI" that neural nets, the dominant data users in contemporary global capitalism, are semiotic engines relying on the index to make and manipulate meaning. Platform capitalism, they argue, relies on the function of "pointing," undermining both belief in the intelligence of AI and some too-easy critiques of this very powerful tool.[40] Inside the black box is a series of signs, linked by indexes.

The core concept of surplus data covers the distance from complex mathematics to political ideology to systems of capital. Surplus is the quantity that underlies the social, the affect generated by the sheer volume of circulating and acting data converted into the quality of quantity itself. This special issue seeks to sketch this terrain, by definition a territory that is impossible to exhaust. The articles draw from different disciplines, work with disparate tools, put forward varied levels of hope. Our goal in what follows is to take notice of a contemporary and still emergent but also major and paradigmatic change.

40. See Leif Weatherby and Brian Justie, "Indexical AI," *Critical Inquiry* 48 (Winter 2022): 381–415.