# 18 Error

**Nanna Bonde Thylstrup**

The emergence of big data has displaced questions of error and truth with a strong desire for efficiency and experimentation. Rather than looking to solve errors, big data develops a tactic that works through an acceptance of them. As Viktor Mayer-Schönberger and Kenneth Cukier (2013) point out in their now classic book on big data, "Looking at vastly more data also permits us to loosen up our desire for exactitude. . . . It's a trade-off: with less error from sampling we can accept more measurement error." Furthermore, many commercial big data ventures operate with a temporality that favors speed over patience: a medium-quality answer obtained quickly is thus often preferred over a high-quality answer obtained slowly.

The displacement and reconfiguration of the notion of error in big data regimes indicate a more fundamental clash of scientific paradigms regarding what constitutes knowledge and how best to achieve it. These clashes are again nested within a deeper politics of how we understand error, who has the power to determine something as an error, and who is made to endure the consequences of these errors. One now classic illustration, which has become eerily prescient in the onslaught of COVID-19, is the scientific dispute between data science and applied statistics in 2014 around the (now defunct) Google Flu Trends and the political implications that followed. In an article from 2009 published in *Nature*, researchers from Google claimed that they would be able to "nowcast" flu epidemics based on Internet searches (Ginsberg et al. 2009). The underlying logic was that people's search results—for example, for symptoms of "fever," "cough," and "sore throat"—would reflect whether they were falling ill with the flu and thus provide Google with real-time signals of imminent flu outbreaks. Google would be able not only to detect flu prevalence but also to do so two weeks earlier than the flu-tracking information from the US Centers for Disease Control and Prevention.

Five years later, however, Google Flu Trends missed the 2013 flu season by 140 percent. Statisticians David Lazer and colleagues (2014) offered their explanation of why Google Flu Trends had failed. They noted that Google Flu Trends had originally been developed as a

machine-learning algorithm to predict the number of flu cases on the basis of Google search terms. While the underlying data management and machine-learning algorithms had been correct, a misunderstanding about the uncertainties involved in the data collection and modeling process had led to highly inaccurate estimates over time. As several statisticians noted, applied statistics would have thought carefully about the sampling process, identified time-series components in the spatial trend, investigated why the search terms were predictive, and tried to understand the likely reasons why Google Flu Trends was working. In other words, classically trained statisticians would have devoted more attention to detecting and identifying errors. While at first this might not have seemed to be a problem, Lazer et al. (2014) point out that it became one over time as the machine-learning algorithms continued to learn from faulty assumptions.

Google Flu Trends' erroneous results caused critics to conclude that the technology was better at finding patterns in past data than it was at predicting future epidemics. And soon after, Google shut down Google Flu Trends as an active operation.

While it would be easy to recount Google Flu Trends through a narrative of failure, however, the phenomenon is in fact a symptomatic example of how major tech companies reconfigure archival glitches into what Orit Halpern, in chapter 12, calls demos: "Experiments that prove which forms of research and technology need to be invoked next; that *should* exist and must be built." Indeed, as sociologist Noortje Marres points out in relation to the ongoing experimental implementation of self-driving cars, such approaches are exemplary of a new, "experimental" mode of industrial innovation, in which experiments and beta testing that would previously have occurred in a lab are today located in everyday societal and intimate settings like streets, personal computers, and smartphones.

This experimental mode of industrial innovation allows tech companies to reframe errors from failed events to progress. Indeed, within the experimental epistemology of tech companies, each new technology that appears on the back of a failed experiment becomes yet another stage in an ever-unfolding tech metamorphosis in which each new demo sheds its old skin to unveil a newer, larger, and more successful one.

Coming back to Google Flu Trends, we see this logic in action: while the operation itself failed to detect the actual unfolding of the 2013 flu epidemic, the more generalized Google Trends has today become a central site of knowledge production for the COVID-19 epidemic. Moreover, Google and Apple have assumed leading positions within the field of digital epidemic detection with their new contact-tracing technology. Leading researchers are already contesting the validity and security of this technology, pointing out its flawed data collection practices (it will not include children and older individuals without new smartphones) and its discriminatory and potentially lethal implications. Ali Alkhatib's (2020) insightful blog post on the problematic aspects of Google and Apple's digital contact tracing thus notes:

"We'll have to constantly follow up on the errors these systems make, struggling to fix the damage it does in its false positives and false negatives, all in this hopeless chase to automate what essentially needs humans."

The central role and experimental approach of Google and its like in the wake of COVID-19 demand that we urgently discuss errors in big data archives, not only as technical occurrences that can be fixed with a tweak to the algorithms but also as political problems rooted in cultural imaginaries of what errors are, who gets to define them, and what their implications might be.

## Vocabularies and Imaginaries of Failure, Error and Fault

How might we understand errors in big data archives as cultural and political, as much as technical, problems? As the Institute of Electrical and Electronics Engineers' *Standard Classification for Software Anomalies* (2010, vi) shows, the semantic and epistemological complexity of the concept of error haunts the computational sciences as much as the humanities and social sciences. The *Standard Classification* thus points out that while *error* has been used interchangeably with other related terms such as *anomaly*, *fault*, *failure*, *flaw*, *glitch*, *defect*, and *bug* in computer science, it would be beneficial for the computational sciences if distinctions could be upheld better between these definitions.

Algirdas Avizienis et al. (2004) offer one effort to define what an error should mean in computing. They provide a linear and causal chain between the concepts: ". . . fault → error → failure → fault . . . ," creating an epistemology in which an *error* is a minor part of the total state of the system, a discrete deviation that will not necessarily lead to complete system failure if the system is resilient. We can experience an error in a resilient program and still work within it, albeit with limited functionality. The system may even anticipate such errors. We see this anticipation in basic terms when we try to divide by zero on a calculator, or in the 404 error responses we get when the server is unable to find the website we are looking for. But the anticipation of error is also built into more fundamental infrastructures of big data—for instance, the web—whose architects and ideologists emphasize resilience and survivability, including the capability to withstand or reroute around local errors and network losses (Hu 2015).

The semantics of error in today's data landscapes indicate that what counts as an error in big data operations is an unresolved and complex question that pertains as much to the interdisciplinary realm of big data as to the deeper conceptual and political landscape of error itself. Big data thus reproduces historical ambivalent notions of error as, on one hand, an anxiety-inducing problem and, on the other hand, a productive practice that can lead to new discoveries and innovations. In this way, big data archives repurpose and extend the

fundamental challenge of historical notions of error but also imbue it with new political and epistemological meanings.

Etymologically, the term *error* incorporates the mobile notion of erring, producing a spatial understanding of error as a deviation from a route—a departure from principles. An error, then, is literally an act of errantry, a navigation along a devious course, implying rambling, roaming, and even going astray (Bruno 2011). As David William Bates (2002) points out in *Enlightenment Aberrations: Error and Revolution in France*, this spatial imaginary gives the word *error* a complex and ambivalent connotation that suspends it between "merely aimless wandering" and "a more specific aberration from some path" (20). This ambivalence also means that error has never been wholly determined by an epistemological structure of truth but has always enjoyed a certain conceptual independence that offers it a much richer terrain.

This complexity, as Bates points out, has often been captured by the figure of the labyrinth, where error is an act of wandering as well as a form of knowledge production. *Erreur* could thus be taken to mean the "action of erring this way and that," "an excursion, a voyage involving adventures," or a "vagabondage of the imagination, of the mind which is not subject to any rules" (Bates 2002). These journeys produced a spatial framework of knowledge in which error and knowledge were productively linked but also frustrating to experience. Just listen to Enlightenment philosopher Jean-Louis Castilhon lamenting the feeling of being lost in the labyrinth of knowledge: "How cruel and painful is the situation of a Traveller who has imprudently wandered into a forest where he knows neither the winding paths, nor the detours, nor the exits!" (Castilhon, quoted in Bates 2002, 132). Many big data operations are haunted by similar ambivalent frustrations: How do they distinguish unproductive or aimless wanderings in big data from productive adventures that lead to new discoveries and inventions?

At the same time, datafication also resurfaces troubling echoes from history, when the detection of error was designed to both produce and contain deviant subjectivities. In his lecture on the concept of error, Georges Canguilhem outlined how the rise of the human sciences in the mid-nineteenth century, especially psychology and early biology, transformed the notion of error from subjective transformation to an external human problem (Talcott 2014). This subtle shift, he argued, produced a new worldview in which living beings did not make errors but rather were subjected to them so that error appeared "as a malformation or a failure" that no longer implied "conversion" but rather "remedy" (Canguilhem, cited in Talcott 2014). Canguilhem ultimately linked this new worldview to the rise of eugenics. Foucault, famously, later developed Canguilhem's theory into his own account of biopower. These perspectives allow us to abandon the technological hunt for an algorithmic "fix" and instead frame the question of error in big data as one of power and politics that has as much to do with racialized and gendered structures as with innovation.

## The Politics of Error

At the end of the essay "Life: Experience and Science,"[1] Michel Foucault (1998) concludes, "At the most basic level of life, the processes of coding and decoding give way to a chance occurrence that, before becoming a disease, a deficiency, or a monstrosity, is something like a disturbance in the informative system, something like a 'mistake.' In this sense, life—and this is its radical feature—is that which is capable of error" (476). Foucault's analysis points to the ambivalence of error as both a creative event and a moment of power. This understanding of error can help us move out of simplified ideas of error as a purely productive process or as technical glitches that can be "corrected" to instead *repoliticize* error as a fundamentally human question of power that has always also been tied up with the human body.

Catherine D'Ignazio shows in chapter 40 on *outlier* that the historical positioning of certain bodies as more anomalous than others also means there is often uncertainty as to whether an outlier is an error in the recording of data or represents a true variation in the population. D'Ignazio thus reminds the reader that rejecting outliers as errors in data sets has serious implications for data subjects and that these implications also tend to reproduce gendered and racialized discriminations. Moreover, in chapter 35 on *(mis)gendering*, Os Keyes shows how these lines of oppression also remain lodged within the binary imaginary of data science, which at once excludes trans experience from its organization of information and at the same time continually reinserts trans people into static gender narratives drawn from archival material from pretransition lives.

The idea of certain genders and sexualities as outliers and errenous has deep historical roots (Agostinho & Thylstrup 2019). As Mary Russo (1995) reminds us in *The Female Grotesque*, female bodies have historically been constituted as "in error" (10). And Yetta Howard (2014) points out that the notion of error today is still "bound up with diagnosing and understanding trans identifications in terms of wrong embodiment" so that "trans-as-error functions in tandem with rubrics of identificatory, mental, and bodily disorders that have historically included nonheterosexual identities and intersexed bodies" (82).

Jacqueline Wernimont (2018) offers one damning indictment how history haunts contemporary best practices of data collection, showing that women are literally dying because of persistent errors due to data collection biases that favor men over women in everything from clinical trials to car safety. These disparities are further exacerbated when one factors in the question of race. As Safiya Noble has shown (2018, 10), marginalized groups are particularly vulnerable to misrepresentation in search engines, where they often appear in erroneous, stereotypical, or even pornographic ways. Such errors should not be perceived merely as technical flaws that can be fixed by improving algorithms but rather as symptoms of deeper societal problems of "white supremacy" (Benjamin 2019). This pervasive cultural and

institutional racism is in turn linked to a sociotechnical system that treats minority voices as less capable of telling the truth, and more liable to error, than the voices of those in power (Ahmed 2015; Crenshaw 2018). Hence, as Amelia Acker points out in chapter 33 on *metadata*, we must ask not only what it means to classify and misclassify something but also who gets to determine and uphold structures of knowledge.

The fundamentally political question of error has also given rise to a reengagement with error as an inroad for critical and subversive engagement with power. This entails regarding errors as holding alternative potential, which can be used against hegemony by showing that alternatives are embedded in dominant cultures: ultimately, power is never total, consistent, or omnipotent. As Rebecca Schneider notes in chapter 26 on *glitch*, "Error bears the potential of ruin, but ruin bears the promise of alternatives." Such alternative approaches to error recast the potential for errors in the archive as avenues of possibilities, inhabitation, and even escape from calculation and predictability. As Jack Halberstam (2011) suggests in *The Queer Art of Failure*, failure and error can offer different kinds of rewards: "Under certain circumstances, failing, losing, forgetting, unmaking, undoing, unbecoming, not knowing may in fact offer more creative, more cooperative, more surprising ways of being in the world" (3). And failure and glitch as subversion has also been a generative strain of thought in digital media theory (Menkman 2011; Nunes 2012).

Rebecca Schneider offers a meditation on the role of error and glitch in relation to feminist artist Carolee Schneemann, who sought to inhabit as much as challenge the category: "I think it is right to say that unlike Enlightenment man and his exploratory jaunts among the irrational or underdeveloped, Schneemann was not fetishizing error as other. With *Interior Scroll* or *More Wrong Things*, she was not getting off on the errors of others so much as residing in the error she was given to be a female and artist simultaneously." Schneider locates the subversive potential of performance in its wholly unpredictable, and often failing, mode of existence; a mode of existence, she argues, that might also make it able to resist co-option.

Queer and critical race scholar Kara Keeling (2014, 157) also foregrounds the potential of queer and critical race theory in unsettling racialized and gendered power structures, this time in relation to data environments, noting, for instance, how "Queer OS can be grasped as a malfunction within technologies . . . with a capacity to reorder things that can, perhaps, 'make race do different things,' tell 'Mama to let herself go free,' and make what was legible soar into unpredictable relations." In her later work, *Queer Times, Black Futures*, Keeling (2019) expands her analysis to compellingly argue that the metaphor of the black swan, a concept rooted in colonial history and usually used today to explain outlier events, in fact points to a much more political question about horizons of knowledge and failures of political imagination. Using the Haitian Revolution as an example, Keeling argues that while the revolution

appeared as a black swan event to European colonialists, a successful revolt unthinkable both before and during the uprising, the event was in fact long anticipated and prepared by those who organized and sympathized with the revolt. In both Schneider and Keeling, we see how error depends on perspective and event horizons. And in both of them, we can also detect the ambivalence of error. Schneider, writing against a political backdrop that increasingly seems to revel in error, thus raises crucial questions for feminists and antiracists in today's political landscape: "How do we deploy failed performatives, or mobilize more wrong things, as modes of interruption into the 'herdlike' stream of consumable data production that is the neoliberal everyday? And how do we do so without refetishizing error in the service of the billionaires who deploy 'alternative facts' to profit from financialization?"

If error now finds itself in an ambivalent alliance with political economies where performances of critique are praised as ventures whose pursuit promotes and strengthens those economies, then big data archives also call for a reengagement with error as a fundamental, and fundamentally political, part of knowledge production. Who gets to determine when something is an error? Which errors are reconfigured as productive errings? What subversive potentials remain for failure and error in the face of resilient and ubiquitous systems that have turned the entire globe into one big site of experimentation? As history shows us, error has always been an ambivalent concept. Yet big data archives confront us with significant new political questions about the role of errors and the agency we are afforded in relating to them.

## Notes

This chapter partly draws from the introduction to the error workshop at the Royal Danish Library in 2017, coauthored by Kristin Veel, Annie Ring, Daniela Agostinho, and myself. I would like to thank Kristin Veel and Daniela Agostinho for their brilliant and helpful comments on earlier iterations of this chapter. I am so grateful for their intellectual and social generosity. I am also grateful for my fellow coeditors' inspiration and care and for colleagues who have actively engaged with and/or contributed to Uncertain Archives over the years.

1. Itself a modified version of Foucault's introduction to the English translation of George Canguilhem's *The Normal and the Pathological* (1989).

## References

Agostinho, Daniela, and Nanna Thylstrup. 2019. "'If truth was a woman': Leaky infrastructures and the gender politics of truth-telling." *Ephemera: Theory & Politics in Organization* 19 (4): 745–775. http://www.ephemerajournal.org/contribution/if-truth-was-woman-leaky-infrastructures-and-gender-politics-truth-telling.

Ahmed, Sara. 2015. "Against students." *Feminist Killjoy*, June 25. https://feministkilljoys.com/2015/06/25/against-students/.

Alkhatib, Ali. 2020. "We need to talk about digital contact tracing." *Ali Alkhatib*, May 1. https://ali-alkhatib.com/blog/digital-contact-tracing.

Avizienis, Algirdas, Jean-Claude Laprie, Brian Randell, and Carl Landwehr. 2004. "Basic concepts and taxonomy of dependable and secure computing." *IEEE Transactions on Dependable and Secure Computing*.

Bates, David William. 2002. *Enlightenment Aberrations: Error and Revolution in France*. Ithaca, NY: Cornell University Press.

Benjamin, Ruha. 2019. *Race after Technology: Abolitionist tools for the new Jim code*. Medford, MA: Polity.

Bruno, Giuliana. 2011. *Atlas of Emotion: Journeys in Art, Architecture, and Film*. New York: Verso.

Canguilhem, Georges. 1998. *The Normal and the Pathological*. New York: Zone Books.

Crenshaw, Kimberlé. 2018. "We still haven't learned from Anita Hill's testimony." *New York Times*, September 27. https://www.nytimes.com/2018/09/27/opinion/anita-hill-clarence-thomas-brett-kavanaugh-christine-ford.html.

Foucault, Michel. 1998. "Life: Experience and science." In *Aesthetics, Method, and Epistemology*, edited by James D. Faubion 465–478. London: Allen Lane.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. "Detecting influenza epidemics using search engine query data." *Nature* 457 (7232): 1012–1014.

Halberstam, Judith. 2011. *The Queer Art of Failure*. Durham, NC: Duke University Press.

Halpern, Orit, and Gökçe Günel. 2017. "Demoing unto death: Smart cities, environment, and preemptive hope." *Fibreculture Journal* 29: 1–23. http://twentynine.fibreculturejournal.org/fcj-215-demoing-unto-death-smart-cities-environment-and-preemptive-hope/.

Howard, Yetta. 2014. "Error." *Transgender Studies Quarterly* 1 (1–2): 82–83.

Hu, Tung-Hui. 2015. *Cloud: A Pre-history of the Network*. Cambridge, MA: MIT Press.

Institute of Electrical and Electronics Engineers. 2010. *Standard Classification for Software Anomalies*. New York: Institute of Electrical and Electronics Engineers. https://standards.ieee.org/standard/1044-2009.html.

Keeling, Kara. 2014. "Queer OS." *Cinema Journal* 53 (2): 152–157.

Keeling, Kara. 2019. *Queer Times, Black Futures*. New York: New York University Press.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The parable of Google Flu: Traps in big data analysis." *Science* 343 (March 14): 1203–1205.

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

Marres N. (2020) What if nothing happens? Street trials of intelligent cars as experiments in participation. In: S. Maasen, S. Dickel, and C. Schneider, eds. *TechnoScienceSociety. Sociology of the Sciences Yearbook*, 30. Cham: Springer. https://doi.org/10.1007/978-3-030-43965-1_7.

Menkman, Rosa. 2011. *The Glitch Moment(um)*. Amsterdam: Institute of Network Cultures.

Noble, Safiya U. 2018. *Algorithms of Oppression: Data Discrimination in the Age of Google*. New York: New York University Press.

Nunes, Mark. 2012. *Error: Glitch, Noise, and Jam in New Media Cultures*. New York: Bloomsbury.

Prasad, Divya, John McDermid, and Ian Wand. 1996. "Dependability terminology: Similarities and differences." *IEEE Aerospace and Electronic Systems Magazine* 11 (1): 14–21.

Russo, Mary. 1995. *The Female Grotesque: Risk, Excess and Modernity*. London: Routledge.

Talcott, Samuel. 2014. "Errant life, molecular biology, and biopower: Canguilhem, Jacob, and Foucault." *History and Philosophy of the Life Sciences* 36 (2): 254–279.

Wernimont, Jacqueline. 2018. *Numbered Lives: Life and Death in Quantum Media*. Cambridge, MA: MIT Press.