
Databases and Doppelgänger: New Articulations of Power

Sandra Robinson
Carleton University

ABSTRACT: This article explores the power and efficacy of databases as part of the turn to intensive datafication in contemporary life in which we are all enfolded within a datafied milieu: a catalogued and curated data assemblage comprising aspects of our life. This assemblage of people, processes, and things is described as having a generative power—data power—producing not the “one” profiled individual, but the many multiple proxies out of the data assemblage. I reimagine the profiled individual through the literary figure of the doppelgänger: a *data doppelgänger* that gestures to difference and repetition and all that is ambiguously changeable within digital culture. Thinking with the data doppelgänger to interrogate the profiling apparatus enables a more complex understanding of data power beyond panoptic metaphors that ground the subject in descriptive and spectral terms rather than performative simulations. To illustrate aspects of the contemporary profiling apparatus, I briefly explore the troubling relationship between Facebook and the data analytics firm Cambridge Analytica to demonstrate how effective profiling systems are when enriched with large, social datasets.

In the early and mid-1990s, critical theorist and media scholar Mark Poster turned his analytical focus to the role of databases in contemporary culture, suggesting that the database operates as discourse because it is implicated in the construction of new subjectivities generated by profiling technologies. Poster’s work is an important and early critique of profiling technologies enhanced by the expansion

of the Internet and digital technologies, and in particular, the advanced ability to monitor consumers and practices of consumption. Information solicited from consumers in exchange for products and services is being stored in greater amounts in databases and subject to analysis with increasing finesse, rendering consumer subjects transparent to the scrutiny of the market. Poster suggests that the database operates as a “superpanopticon”: a “perfect writing machine [that] constitutes subjects as decentered from their ideologically determined unity.”¹ For Poster, the idea of a “perfect” constitutive machine points to the precision in contemporary profiling systems operating as part of networked communication.² The postmodern subject is thus perpetually reconstituted by the grinding repetition of a profiling apparatus that draws on a vast repository of personal information held in consumer databases to continuously recraft different consumer identities.

My analysis re-engages with Poster’s salient critique expressing concerns over the power and efficacy of databases as part of contemporary information and media infrastructure. This infrastructure is the scaffold over which “big data” flows from the entanglement of practice, processes, and things that make up the contemporary profiling apparatus and that enables the tracking, collection, and analysis of people. The profiling apparatus is generative and dynamic, producing not only “one” profiled individual, but multiple proxies. Our lifeworld is a “datafied” milieu: a catalogued and curated mélange of information produced from every aspect of our lives. Recent scholarship has engaged the notion of power in and through data, settling on the idea of *data power*,³ which hinges on the connection between the expanding role and influence of “big data” on contemporary life and its authoritative resonance as power.⁴ In the last two years,

1. Mark Poster, “Databases as Discourse; or, Electronic Interpellations,” in *Computers, Surveillance, and Privacy*, ed. David Lyon and Elia Zureik (Minneapolis: University of Minnesota Press, 1996), pp. 175–192, at p. 184.

2. Poster explores databases and contemporary networked communication and media across much of his work in the 1990s in Mark Poster, *The Mode of Information: Poststructuralism and Social Context* (Chicago: The University of Chicago Press, 1990), pp. 90–111; *The Second Media Age* (Cambridge: Polity Press, 1995), pp. 87–93.

3. The idea of data power appears in a range of scholarship examining data, algorithms, and databases including from Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (London: SAGE Publications, 2014); Helen Kennedy and G. Moss, “Known or Knowing Publics? Social Media Data Mining and the Question of Public Agency,” *Big Data and Society* 2:2 (2015): 1–11; Jo Bates, Yu-Wei Lin, and Paula Goodale, “Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows,” *Big Data and Society* 3:2 (2016): 1–12.

4. Antoinette Rouvroy, “The End(s) of Critique: Data Behaviourism Versus Due Pro-

the troubling relationship between Facebook and the data analytics firm Cambridge Analytica has demonstrated how powerful profiling systems can become when enriched with large, social datasets. Details are only just emerging in mid-2018 about how Cambridge Analytica obtained and used data linked to Facebook accounts to generate political advertising, but as I outline below, this case shows how profiling systems are designed and deployed.⁵

In this analysis, I reimagine the profiled, the proxy, and the double through the figure of the *doppelgänger* as an apt metaphor for the sometimes conflicted and fraught idea of our multiple selves circulating in the flows of information. From Roger Clarke's "digital persona" to the data doubles of the "surveillant assemblage," the idea of our similar, but fleeting other gathered from all our online traces is as unsettling as the literary motif of the *doppelgänger* as an "inveterate performer of identity."⁶ In this context, I critique data power in part through Poster's notion of the database as a "perfect writing machine" that amplifies the power and control of corporate database owners by generating our other selves.⁷ Poster's "perfect writing machine," however, is only a starting point: there are now a plethora of data collection sites from personal communication devices and applications to traffic cameras and smart utility monitors, which have expanded the data cycle. Features such as the capacity for natural language query in Internet search services, functions enabling participation and collaboration on and through social platforms, and the intensive connections between consumer activity and tracking systems in general, means the implied unidirectional operation of Poster's "perfect writing machine" does not adequately capture the performative aspect of the database.⁸ The complexity of contemporary digital media and communication technology has

cess," in *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, ed. Mireille Hildebrandt and Katja de Vries (New York: Routledge, 2013), pp. 148–168.

5. Carole Cadwalldr, "'I Made Steve Bannon's Psychological Warfare Tool': Meet the Data War Whistleblower," *Guardian*, March 18, 2018, <https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-facebook-nix-bannon-trump>.

6. Roger Clarke, "The Digital Persona and Its Application to Data Surveillance," *The Information Society* 10:2 (1994): 77–92, at p.78; Kevin D. Haggerty and Richard Ericson, "The Surveillant Assemblage," *British Journal of Sociology* 51:4 (2000): 605–622, at p. 606; Andrew J. Webber, *The Doppelgänger: Double Visions in German Literature* (Oxford: Clarendon Press, 1996), p. 3.

7. Poster, "Databases as Discourse" (above, n. 1), p. 184.

8. Bart Simon, "The Return of Panopticism: Supervision, Subjection and the New Surveillance," *Surveillance and Society* 3:1 (2005): 1–20.

shifted consumer practice as well as the data collection and retention strategies of governments and corporations. The rich scholarship in consumer surveillance studies and consumer profiling in the last couple of decades through the work of Greg Elmer, Detlev Zwick and Niklas Dholakia, Zwick and Janice Knott, and code studies from geography to sociology to media studies, including work from Martin Dodge and Rob Kitchin, Adrian Mackenzie, and Lev Manovich, respectively, trace the burgeoning assemblage of people, processes, and things expanding our datafied milieu.⁹ More recently, scholars have worked to delve more deeply into the troubling aspects of the data assemblage and its capacity to govern individuals, from building on Michel Foucault's governmentality thesis to how algorithms control our datafied selves and algorithmic identities and do the work of culture, sorting and classifying people, ideas, and things.¹⁰ Popular texts have also explored the issue of consumer profiling and "doppelgänger searches" directed at finding "people like us" such as the sort Amazon uses, for example, in its recommendation system to offer us products linked to our own past search and purchase behaviors in combination with others deemed similar to us. Cathy O'Neil's popular book, *Weapons of Math Destruction*, details the power and privilege in secretive financial algorithms that reflect the bias and assumptions of their creators and the discriminatory potential for consumer-citizens in banking, housing, and education.¹¹ As I discuss

9. Greg Elmer, *Profiling Machines: Mapping the Personal Information Economy* (Cambridge, MA: The MIT Press, 2004); Detlev Zwick and Niklas Dholakia, "Whose Identity Is It Anyway? Consumer Representation in the Age of Database Marketing," *Journal of Macromarketing* 24:1 (2005): 1–42; Detlev Zwick and Janice Knott, "Manufacturing Customers: The Database as New Means of Production," *Journal of Consumer Culture* 9:2 (2009): 221–247; Martin Dodge and Rob Kitchin, "Codes of Life: Identification Codes and the Machine-Readable World," *Environment and Planning D: Society and Space* 23:6 (2005): 851–881; Adrian Mackenzie, "The Performativity of Code: Software and Cultures of Circulation," *Theory, Culture & Society* 22:1 (2005): 71–92; Lev Manovich, *The Language of New Media* (Cambridge, MA: The MIT Press, 2001).

10. Rouvroy, "End(s) of Critique" (above, n. 4); John Cheney-Lippold, *We Are Data: Algorithms and The Making of Our Digital Selves* (New York: New York University Press, 2017); Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015); Tarleton Gillespie, "The Relevance of Algorithms," in *Media Technologies: Essays on Communication, Materiality, and Society*, ed. T. Gillespie, P. J. Boczkowski, and K. A. Foot (Cambridge, MA: The MIT Press, 2014), pp. 167–193; Ted Striphas, "Algorithmic Culture," *European Journal of Cultural Studies* 18:4–5 (2015): 395–412.

11. Seth Stephens-Davidowitz, *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us about Who We Really Are* (London: Bloomsbury Publishing, 2017); Cathy O'Neil, *Weapons of Math Destruction* (New York: Crown, 2016); Michele Willson, "Algorithms (and the) Everyday," *Information, Communication & Society* 20:1 (2017): 137–150.

in more detail below, the expansive access to social networks by third parties such as application developers, advertisers, and data analysts is what enabled Cambridge Analytica to harvest data from millions of Facebook users, which was then used to target individuals with intensively personalized and politically charged advertising during the 2016 presidential election in the United States.¹²

The intensive turn to “datafication” as part of making the world intelligible through data is altering fundamental conceptions about what it is to be an individual subject understood in part through information used to identify, categorize, predict, preempt, and control aspects of life. In what follows, I offer a short history of the literary figure of the *doppelgänger* and discuss its connection to themes of disruption, disorder, difference, and repetition, arguing that the *data doppelgänger* is a productive constituent through which to critique data power and the role of databases within the data assemblage. I then examine databases as part of data infrastructure, exploring new avenues for thinking with Poster’s original “database as discourse” as a means to moving beyond earlier work linking profiling with surveillance. Thinking with the data *doppelgänger* to interrogate the profiling apparatus enables a more complex understanding of data power beyond panoptic metaphors that ground the subject in descriptive and spectral terms rather than performative simulations. Claims abound with regard to what data can *do*—from what data may reveal to how data may transform the human subject through the objectifying processes of datafication with the promise to *know* human wants, needs, and risks and thus arrange and control future life chances, actions, and opportunities. Where once a politics of privacy was the focus of a critical engagement with the expansion of communication networks, a politics of databases may force us to consider other forms of agency and strategies of resistance to forms of intelligibility actualized through the database and profiling apparatus.

The Doppelgänger

As an imagined figure, a soul, a shadow, a ghost or a mirror reflection . . . the psychological power of the double lies in its ambiguity, in the fact that it can stand for contrast or opposition, but likeness as well.¹³

The notion of a double or *doppelgänger* has, over the last five centuries, become a well-worked literary motif that troubles identity and

12. Cadwalldr, “Steve Bannon’s Psychological Warfare Tool” (above, n. 5).

13. Milica Živković, “The Double as the ‘Unseen of Culture’: Toward a Definition of Doppelgänger,” *Facta Universitatis* 2:7 (2000): 121–128, at p. 122.

difference, certainty and ambiguity, therefore defying any “stable origin of subjectivity.”¹⁴ Doubles, twins, self-division, and duplication inhabit the rich landscape of traditional storytelling and mythology in many different cultures. In the West, the double appears in early works by the theologian Ludvig Lavater in 1572 and the folklorist John Aubrey in 1696 in relation to Teutonic and Celtic folkloric superstition about ghosts and spirits. The word “doppelgänger” itself, which can be translated from German as the “double-goer,” first appeared in a work by the German Romantic author Jean Paul Richter in 1796. In Richter’s novel, *Siebenkäs*, the doppelgänger is formally introduced into the Western cultural imagination. Sir Walter Scott, the British adventurer, and Percy Bysshe Shelley, the Romantic poet and novelist, both wrote of doubles, and Shelley is said to have seen his doppelgänger just before he drowned in 1822.¹⁵

In the nineteenth century, the doppelgänger assumed a Gothic quality reflecting the fascination authors such as Shelley held for exploring mimetic themes in Romantic literature. This fascination can be traced through E. T. A. Hoffmann’s *The Devil’s Elixir* in 1815, Mary Wollstonecraft Shelley’s *Frankenstein; or, The Modern Prometheus* in 1818, Edgar Allan Poe’s short story “William Wilson” in 1839, Fyodor Dostoyevsky’s novella *The Double* in 1846, and to Robert Louis Stevenson’s *The Strange Case of Dr. Jekyll and Mr. Hyde* in 1886, each with their fictional explorations of doubles and doppelgängers.¹⁶ These stories parallel a growing preoccupation with life and vitalism in the nineteenth century in fictional and scientific writing concerning the vivified, the reanimated, and the alienated.¹⁷ The

14. Dimitris Vardoulakis, *The Doppelgänger: Literature’s Philosophy* (New York: Fordham University Press, 2010), p. 136.

15. Ralph Randal and Margaret Randal, “Shelley’s Magus Zoroaster and the Image of the Doppelgänger,” *Modern Language Notes* 6:1 (1961): 7–12; Hillel Schwartz, *The Culture of the Copy: Striking Likenesses, Unreasonable Facsimiles* (New York: Zone Books, 2014); Živković, “Double as the ‘Unseen’” (above, n. 13); Jean-Paul Richter, *Flower, Fruit, and Thorn Pieces . . . Siebenkas* (Boston, MA: Ticknor and Fields, 1863), <https://archive.org/details/flowerfruitthorn00jeanuoft>.

16. E. T. A. Hoffmann, *The Devil’s Elixir* (Edinburgh: William Blackwood; T. Cadell, London, 1815), <https://archive.org/details/devilselixir01hoffgoog>; Mary Wollstonecraft Shelley, *Frankenstein; or, The Modern Prometheus* (London: Oxford University Press, 1969); Edgar Allan Poe, “William Wilson” [electronic resource] (Charlottesville: University of Virginia Library; Boulder, CO: NetLibrary, 1993); Fyodor Dostoyevsky, *The Double*, trans. George Bird (Bloomington: Indiana University Press, 1958); Robert Louis Stevenson, *The Strange Case of Dr. Jekyll and Mr. Hyde, and Other Famous Tales* (New York: Dodd, 1961).

17. Robert Mitchell, *Experimental Life: Vitalism in Romantic Science and Literature* (Baltimore: Johns Hopkins University Press, 2013). The theme of the double is also influential

idea of mistaken identity, twins, clones, alter egos, doubles, and doppelgängers continues to permeate culture from themes in books such as J. R. R. Tolkien's *Lord of the Rings* trilogy and Frank Herbert's *Dune* series, to movies such as Hitchcock's *Vertigo*, Terry Gilliam's *Brazil*, and recently Netflix's *Stranger Things* TV series and its "upside down" in which dwells a sort of shadow self, alongside a very scary monster.¹⁸ Across these fictional accounts, from old to new, the double or doppelgänger is often portrayed as a negative or evil entity through which "[h]ostile actions . . . ascribed to [a] foreign self [are] performed by proxy."¹⁹

Milica Živković suggests that the double resists and opposes the established order of society because of its potentiality, the possibility of "innumerable other selves" that dilute the dominant system where it is reproduced in the individual.²⁰ In Romantic-era fiction in particular, the doppelgänger always has some kind of flaw that renders the double at cross-purposes to the individual: the doppelgänger is always at a distance from the human, and the "replication is always unbalanced . . . at least one significant trait nullified [and] seeming congruence frayed."²¹ For example, in Poe's "William Wilson," the doppelgänger appears without warning to intervene on Wilson's mischief-making on more than one occasion, with the doppelgänger disrupting the actions of the morally questionable protagonist. The doppelgänger is a "mirror-twisted twin" that profoundly troubles the continuity and control of one's life with "a life contravening yours, but its fate your fate."²² Fast-forward almost 150

in art, for example, in Dante Gabriel Rossetti's painting *How They Met Themselves* (1851–1860) discussed in Sophia Andres, *The Pre-Raphaelite Art of the Victorian Novel* (Columbus: Ohio State University Press, 2005); and in René Magritte's paintings in the early twentieth century including *Portrait of Paul Nougé* (1927) and *An End to Contemplation* (1927) included in the exhibition Magritte: The Mystery of the Ordinary, 1826–1930, New York, The Museum of Modern Art, <https://www.moma.org/calendar/exhibitions/1298>.

18. J. R. R. Tolkien, *The Fellowship of the Ring: Being the First Part of The Lord of the Rings* (London: HarperCollins, 2007); Frank Herbert, *Dune* (New York: Ace Books, 2005); *Vertigo*, directed by Alfred Hitchcock (USA: Alfred Hitchcock Productions, 1958); *Brazil*, directed by Terry Gilliam (UK: Embassy International Pictures, 1984); and *Stranger Things* (USA: 21 Laps Entertainment, 2016).

19. Karl Miller, *Doubles: Studies in Literary History* (Oxford: Oxford University Press, 1985), p. 25.

20. Živković, "Double as the 'Unseen'" (above, n. 13), p. 7.

21. Bryan Alexander, "Dialectical Nightmares: The Historicity of the Romantic-Era Doppelgänger in the Works of Godwin, Hogg, Blake, Burney, and the Shelleys" (PhD. diss., The University of Michigan, 1997), p. 16.

22. Schwartz, *The Culture of the Copy* (above, n. 15), p. 54.

years from Poe's work, and this is echoed in the dystopian society envisioned by director Terry Gilliam in the movie *Brazil*.²³ In the movie, a totalitarian state collects information about its citizens with bureaucratic (in)efficiency and subjects them to intensive scrutiny enabled through a complex steam-punk-like apparatus of mechanical and outsized analog infrastructure of tubes, wires, levers, gears, and cathode ray-tubed screens. When a bureaucrat in one of the information departments squashes an oversized fly on the ceiling just above a machine generating a list of profiled citizens, it falls down and briefly clogs the mechanically keyed output replacing the "T" in Tuttle with a "B," creating Buttle. This mistake amid such socio-technical complexity proves impossible to disentangle, and the Ministry of Information finds great confusion and difficulty in welding together an explanation between differing government departments as to the identity and status of the data doppelgänger, Buttle. So, as the character Mr. Kurtzman notes in the movie, Buttle has become something of a puzzle: "Population Census have got him down as dormant, the Central Collective Storehouse computer has got him down as deleted, and the Information Retrieval have got him down as inoperative. . . . Security has him down as excised, [and] Admin have him down as completed."²⁴

While many of us may encounter less dramatic data doppelgängers than in the *Brazil* example, such doppelgängers work in ways that are consequential and full of potentiality. For example, the profiled bank customer is denied a mortgage without understanding why, or tracked and analyzed consumers are offered a deal on their favorite brands based on their recent purchases. The digital doppelgänger is a proximal data object that can *do* something in the context of a decision-making apparatus such as Amazon's recommendation system; it can *act* to trigger particular kinds of suggestions for our consumer selves by merging our past actions and choices with our many similes within an aggregated corpus of data to which we are compared. These actions and consequences reflect what Andrew Webber calls the "performative character" of the doppelgänger engaged in "enactments of identity,"²⁵ whereby each "new" self has consequential outcomes for the individual. Mr. Tuttle's muddled informational proxy, his performative data doppelgänger, Buttle, is determinative and substantial within the orthodoxy of the governing apparatus in the Ministry of Information. The data doppelgänger triggers a cascade

23. Gilliam, *Brazil* (above, n. 18).

24. *Ibid.*

25. Webber, *The Doppelgänger* (above, n. 6), p. 3.

of consequences that appear impossible to correct: not so dissimilar to the difficulty human subjects have in our contemporary world in addressing their data doppelgängers.

The performative aspect of the doppelgänger has been explored recently in the context of self-trackers such as the FitBit and quantified self-movement. The doppelgänger in this dimension is understood through a “set of distinct practices” that occur in what Matthias Bode and Dorthe Kristensen call “digital doppelgängering,” binding it to processes of self-construction, self-management, and the production of facts about oneself, publicized in niche media circuits such as a friends list or circle whether in the FitBit app or on Facebook.²⁶ The digital doppelgänger is also in kinship with the data double of the “surveillant assemblage” that “operates by abstracting human bodies from their territorial settings and separating them into a discrete series of flows,” which are reassembled into “data doubles” for the purposes of profiling.²⁷ The surveillant assemblage conveys the force and flow of a convergent stream of data about individuals and things that grows ever larger today. The data double of fifteen years ago is now the bloated doppelgänger of consumer surveillance and digital culture—a constellation of performative affect, communicative practice, and digital tracking that doubles down on consumer identity and activity from purchases to searches to tweets and deletes.

The doppelgänger in fictional narratives is never an identical copy of the self, and yet the idea of a data double is a persistent and frequent construct in surveillance studies and data studies wherein the word “double” suggests our *datafied self* is an identical proxy.²⁸ However, information about us is neither a copy nor an exact double of the flesh and blood person. Rather, the idea of a data *double* reflects a semblance of familiarity best ascribed to the performative data doppelgänger: a digital self that looks and feels familiar but can perform quite differently depending on what data is used to construct it and the context of its application within a particular analytical process. The doppelgänger, therefore, does not spring from a grounded origin; it propagates from a continuously variable data set from any number of databases. Data—archived, catalogued,

26. Matthias Bode and Dorthe Kristensen, “The Digital Doppelgänger within: A Study on Self-Tracking and the Quantified Self Movement,” in *Assembling Consumption: Researching Actors, Networks and Markets*, ed. Robin Canniford and Domen Bajde (New York: Routledge, 2015), pp. 119–134, at p. 119.

27. Haggerty and Ericson, “The Surveillant Assemblage” (above, n. 6), p. 606.

28. Amit Marcus, “Recycling of Doubles in Narrative Fiction of the Twentieth and Early Twenty-First Centuries,” *Partial Answers: Journal of Literature and the History of Ideas* 11:2 (2013): 187–217, at p. 191.

curated, and “databased”—stands in for our vital, lively qualities. In this way, any sort of profile, any constructed data doppelgänger summoned through/by the processes of profiling, is a liminal construct imbued with a performative capacity to act for us in particular decision-making scenarios. The data assemblage produces a vital and changeable data doppelgänger: it has an assortment of data that is extracted from the database and assembled in a profile but that never represents an identical copy of the person. In its myriad instantiations, the data doppelgänger is always reinscribing difference and resists any reconciliation between versions of the datafied self and the individual.²⁹ The data doppelgänger is thus a mercurial figure that gestures to difference and repetition and all that is ambiguously changeable within digital culture, a culture evermore constituted by information about us generated through the incessant production, collection, and analysis of data. There are no doubles in cyberspace; there are only multiples and mutations, which can appear in infinite recursion.

It is for these reasons that I summon the doppelgänger as the constituent of the powerful apparatus of data collection, analysis, and circulation that contemporary networked communication enables. Multiplying identities produced as a matter of contemporary transactional life destabilize the liberal concept of the coherent, rational self. The data doppelgänger is continuously in a process of formation and transformation; it is ontologically linked to the “past but also laden with a future.”³⁰ From the literary motif to a metaphor for the multiples of cyberspace, the doppelgänger is a modulating form and always in a process of construction, a liminal being occupying a transverse flow across our lifeworld, appearing here and there unannounced, at once familiar yet at the same time a source of perplexity in its alterity.

The Database: A Generative Assemblage

The data doppelgänger fixes the digital traces of our communicative and transactional lives in a fleeting moment of analytical creation, “making explicit what is implicit, present what is latent” out of the shards of our lives captured in databases.³¹ Gilles Deleuze theorized “societies of control” following what Foucault called the disciplinary society by describing a shift from spaces of confinement and encl-

29. Vardoulakis, *The Doppelgänger* (above, n. 14).

30. *Ibid.*, p. 7.

31. Alexander, “Dialectical Nightmares” (above, n. 21), p. 39.

sure to continuous, modulating control by code or software.³² At the present moment, it is no longer the value *of* the reflexive individual, the person; rather it is the value *in* the data produced by and about the individual—what Deleuze called a multiplicity of “dividuals.”³³ These “dividuals,” or shards and measurable pieces of our lives, are the myriad data points that circulate as part of “information flows” within systems that may include social networks, financial services, or transactional processes all connected by networks.³⁴ All of these systems collect and store data points in what Deleuze called data “banks” and what we now call databases.³⁵ In the context of a society of control in which software has become integral to networked communication and the coordination and control of information flows, the whole of the profiling apparatus operates as “an assemblage, in its multiplicity, [and] necessarily acts on semiotic flows, material flows, and social flows simultaneously.”³⁶ The database is an often forgotten piece of this assemblage.

The database has been crucial to the collection and storage of data since the 1960s, but long before that, the list form clearly served as the organization and structure of information in antiquity preceding the database by several thousands of years.³⁷ Databases are “carefully arranged lists” separated into tables with “grids of specification,” which is in part Poster’s argument to consider them as discourse following Foucault’s “rules of formation” for discourse.³⁸ Any type of data on any person, place, or thing—including information,

32. Gilles Deleuze, “Postscript on the Societies of Control,” *October* 59:4 (1992): 3–7; Michel Foucault, *Discipline and Punish: The Birth of the Prison*, trans. Alan Sheridan (New York: Pantheon Books, 1977).

33. Deleuze, “Postscript” (above, n. 32), p. 5; and see for related discussion, Rouvroy, “End(s) of Critique” (above, n. 4), pp. 147 and 157.

34. Manuel Castells has written extensively on what he sees as a shift to a network society governed by the logic of “information flows” in *The Rise of the Network Society* (Cambridge, MA: Blackwell Publishers, 1996), p. 469.

35. Deleuze, “Postscript” (above, n. 32), p.5.

36. Gilles Deleuze and Felix Guattari, *A Thousand Plateaus: Capitalism and Schizophrenia*, trans. Brian Massumi (Minneapolis: University of Minnesota Press, 1987), p. 23. This idea is later explored in Scott Lash, *Critique of Information* (London: Sage, 2002), p. 112, where he argues the “logic of flows is the logic of communications” and socio-technical assemblages, which aligns his perspective more closely with Deleuze than Castells.

37. Liam Cole Young, “On Lists and Networks,” *AMODERN* 2 (2013): <http://amodern.net/issues/amodern-2-network-archaeology/>.

38. Poster, “Databases as Discourse” (above, n. 1), pp. 184, 185; Michel Foucault, *The Archaeology of Knowledge*, trans. A. M. Sheridan Smith (London: Tavistock Publications, 1972), p. 42.

sound, text, images, and transactions—can be stored in a database and interconnect with databases containing data on completely separate objects for the purposes of analysis. Databases have a specific structure and in turn organize a collection of data in particular ways according to the data model in use—a fact made very clear in the data model designed by Cambridge Analytica discussed below. In the past, hierarchical databases grouped similar kinds of data objects in a cascading hierarchy, making the interconnection of other databases unwieldy, whereas contemporary relational databases organize data by grouping them in tables of rows and columns out of which relational links among data objects and features can be built.³⁹ The relational model is more flexible, yet still highly organized with descriptions of things and the properties of things clearly categorized according to an established data ontology.⁴⁰

Paul Dourish argues databases “make the world” as material assemblages that combine software, hardware, and data and are a prolific part of communication and media infrastructure:

The world is increasingly made up of databases, as digital technologies continue to supplement, surround, or displace other forms of record keeping [and] this very spread of digital forms means that we increasingly understand, talk about, think about, and describe the world as the sort of thing that can be encoded and represented in a database.⁴¹

39. Martin Dodge and Rob Kitchin, “Code and the Transduction of Space,” *Annals of the Association of American Geographers* 95:1 (2005): 162–180; Rahul Mukherjee, “Interfacing Data Destinations and Visualizations: A History of Database Literacy,” *New Media & Society* 16:1 (2013): 110–128; Ray M. Chang, Robert J. Kauffman, and YoungOk Kwon, “Understanding the Paradigm Shift to Computational Social Science in the Presence of Big Data,” *Decision Support Systems* 63 (2014): 67–80.

40. Relational databases are much more complex than detailed here and require a formal structured query language (SQL) to easily enter, store, query, and analyze the data. The challenge in the era of “big data” is that relation databases cannot directly handle unstructured data; that is, the unstructured data generated by, and enabled through, Web 2.0 (and beyond) network applications including graphic content such as photos and videos, streaming measurement data (for example from sensors), webpages, PDF files, blogs, as detailed in Chang, Kauffman, and Kwon, “Understanding the Paradigm Shift” (above, n. 39). Unstructured data that does not easily fit a clear classification scheme can be used with systems such as Hadoop’s noSQL framework, which enables ETL processes (extract, transform, load) to pull data out of an unstructured dataset and organize it. I use “database” to denote the data structure and cataloguing of data, but databases require a database management system. For example, a relational database management system (RDBMS) contains the rules that structure the database, but also extend to the profiling function facilitating the input/output regime crucial to the reproduction of data doppelgängers.

41. Paul Dourish, “NO SQL: The Shifting Materialities of Database Technology,” *Com-*

Lev Manovich is equally effusive about the database, suggesting it is a “cultural form of its own” and the “symbolic form” of the computer age.⁴² He argues that the “world is reduced to two kinds of software objects which are complementary to each other: data structures and algorithms [which] are two halves of the ontology of the world according to a computer.”⁴³ Manovich defines a data structure as “a particular way of storing and organizing data in a computer so that it can be used efficiently,” and make up the “structured collection of data” we know as a database.⁴⁴ World-making powers are ascribed to databases in both Manovich and Dourish’s conceptions, and, together with algorithms—the rule sets that parse and analyze data—they establish the computational form as an authentic way of producing and understanding reality. However, it is important to interrogate this computational array—the database within an algorithmic milieu—as what Foucault called a “régime of truth,” that is, as “mechanisms and instances which enable one to distinguish true and false statements, the means by which each is sanctioned; the techniques and procedures accorded value in the acquisition of truth.”⁴⁵ The world-producing vision of Manovich and Dourish swerves a little too closely to an incontrovertible determinism enabled through and by computation as the key motivator to progress. Databases, along with algorithms, are a structured and structuring system that powerfully intervene in and often replace human decision-making through prediction, preemption, and control, in which “‘Truth’ is to be understood as a system of ordered procedures for the production, regulation, distribution, circulation and operation of statements.”⁴⁶ These processes, or “ordered procedures,” are neither neutral nor purely objective, and classification, clustering, and representation embed particular kinds of categories such as race, class, and gender, with particular sorts of meaning and potential bias

putational Culture (November 2014): <http://computationalculture.net/no-sql-the-shifting-materialities-of-database-technology/>.

42. Manovich, *Language of New Media* (above, n. 9), p. 194.

43. *Ibid.*, p. 198.

44. Lev Manovich, “Database as Symbolic Form,” *Convergence* 5:2 (1999): 80–99, at p. 81.

45. Michel Foucault, *Power/Knowledge: Selected Interviews and Other Writings*, trans. Colin Gordon (New York: Pantheon Books, 1980), at p. 131. See also Sandra Robinson, “The Vital Network: An Algorithmic Milieu of Communication and Control,” *communication + 1* 5:1 (2016): doi:10.7275/R5416V0R.

46. Foucault, *Power/Knowledge* (above, n. 45), p. 133.

within databases.⁴⁷ The “ordered procedures” designed as part of the data assemblage—data and algorithms—emerge from the complex social and technical milieu of practice in which such system design unfolds, and software development processes have long struggled with systemic bias.⁴⁸ Objects in the contemporary relational database ontology are thus assigned properties and can be separated into classes, clustered together, and through analysis generate a diagram or map of relations between entities, effectively producing information. Databases are a framework for object relations, but “[k]nowledge in this context is a database of *representations* which can be translated into language” to describe a person, place, or thing.⁴⁹

This, in many ways, is the point of Poster’s work through the 1990s to argue that the mode of information, and its incorporation of databases, has a powerful decentering effect on the human, reflexive individual. Databases, according to Poster, are “writing at the border of subject and object,” and *representations* of the individual are neither solid nor stable, but rather an “unsettling simulation of unity” because they draw on disparate data to manufacture the profiled.⁵⁰ Poster extends Foucault’s panoptic model of surveillance by an order of magnitude by coupling “circuits of communication” to the databases now required for intensive information gathering.⁵¹ This constitutes what Poster calls a “superpanopticon” as a means of controlling people at a distance.⁵²

Greg Elmer’s critical study of consumer surveillance underscores this point. In his detailed study of consumer profiling, Elmer delves deeper into the behavior of the database systems within technologies of commercial surveillance, exposing the binary push/pull between consumer solicitation and response. His work offers an explanation for the operative features of data acquisition, and he maps the “everyday data economy in which habits, routines, rhythms, and flows are digitized, coded, and diagnosed for the purposes of control,” sug-

47. See, for related discussion, Rena Bivens, “The Gender Binary Will Not Be Deprogrammed: Ten Years of Coding Gender on Facebook,” *new media & society* 19:6 (2015): 880–898; O’Neil, *Weapons of Math Destruction* (above, n. 11).

48. Tara MacPherson, “Designing for Difference.” *differences* 25:1 (2014): 177–188; Alison Adam, *Artificial Knowing: Gender and the Thinking Machine* (London: Routledge, 1980).

49. June Power, “The Object in Perspective,” *ACM SIGPLAN OOPS Messenger* 4:2 (1993): 28–32, at p. 2 (emphasis added).

50. Poster, *The Mode of Information* (above, n. 2), p. 111.

51. *Ibid.*, citing Foucault (above, n. 2), p. 93.

52. *Ibid.*

gesting consumers accept monitoring as a part of ordinary life transactions.⁵³ This is behavior referred to by Poster as the “interpellation by database,” in which the observed become willing participants in such surveillance by providing the information requested or solicited during the consumer (or other) transaction.⁵⁴ Foucault’s panoptic model, which required surveillance to function, also relied on a “network of relations” between hierarchical figures in a disciplinary power structure, but Poster suggests the superpanopticon leverages the power of databases to operate productively, and *horizontally*, across and between multiple networks.⁵⁵ No body need be present; as in Deleuze’s societies of control, “the surveillance apparatus does not act on bodies or minds, but on information about bodies and minds.”⁵⁶ For Poster, the superpanopticon is a new “discourse/practice” that “reconfigures the constitution of the subject” in absentia.⁵⁷ Under this model, the subject is capable of being acted upon by code and inscribed by facts in the database, which operates as a “perfect writing machine” at some distance from the individual. Poster takes Foucault’s idea that power today is a more “perfect” power because it is expressed with greater precision through a multitude of capillary-like extensions throughout society, which Poster understands as “a sleek operation . . . occluded in the willing participation” of individuals who give up or leave behind a trail of digital information.⁵⁸

Poster understands the database as discourse—a “perfect writing machine.” This is the discursive logic embedded in databases that have the ability to write out identity redux from discrete personal facts captured and stored as data. Individuals can imagine how they might be profiled, but they cannot know precisely how personal facts are mapped within a profile generated by databases. The process is more often incomprehensible to the individual, and yet in spite of the lack of transparency, individuals are, in effect, “recruited into an understanding of themselves as data patterns.”⁵⁹ A set of clustered

53. Elmer, *Profiling Machines* (above, n. 9), p. 42.

54. Poster, “Databases as Discourse” (above, n. 1), p. 187; and see, for original discussion on interpellation, Louis Althusser, “Ideology and Ideological State Apparatuses (Notes towards an Investigation),” trans. Ben Brewster, in *Lenin and Philosophy and Other Essays* (New York: Monthly Review Press, 2001), pp. 170–177.

55. Foucault, *Discipline and Punish* (above, n. 32), pp. 176–177; Poster, “Databases as Discourse” (above, n. 1).

56. Simon, “Return of Panopticism” (above, n. 8), p. 15.

57. Poster, “Databases as Discourse” (above, n. 1), p. 182.

58. *Ibid.*, p. 184; Poster, *Second Media Age* (above, n. 2), p. 87.

59. Robert Cluley and Steven Brown, “The Dividualised Consumer: Sketching the New

facts, a light patterning of similarity, stands in for them during important life calculations and decisions from whether they get a bank loan to an assessment for insurance. This is the *doppelgänger* that emerges from the database, often fleetingly, in a moment of record that serves to abstract individuals from their material surroundings. It points to Poster's construction of databases as "nothing but performative machines, engines for producing retrievable identities."⁶⁰ This idea is in dialogue with Mackenzie's framework for contemporary software as performative and operationalized, whereby code is described and enacted "within a discursive formation associated with information and communication processes" that focuses attention on the "cultural life of code in circulation" through which processes of circulation generate "performative effects."⁶¹ These performative effects are the expression of power through computational systems as part of our datafied milieu.

Poster's Foucaultian approach considers the "performative aspect of language [and] what language *does* rather than what it denotes or connotes" to specifically link what databases *do* through their practice as discourse.⁶² Bart Simon suggests that Poster's argument is flawed because there is not necessarily a human subject directly interacting with the database and no moment when interpellation actually occurs whereby a person "recognizes themselves as subjects of the call of another."⁶³ In many scenarios, a database is automatically queried by another system, for example during a credit check, when a set of facts about a person are assembled according to the requirements of the financial product and credit scoring process without direct contact with the applicant. However, Poster does not suggest that the database as discourse, with its powerful decentering effect on the subject, operates through the same process of subject formation as in Foucault's relation of knowledge/power understood through subjectivation. The *databled* self, our data *doppelgänger*, is a "simulation of identity" and is generated through processes of objectification "producing individuals with dispersed identities, identities of which the individuals might not even be aware."⁶⁴

Mask of the Consumer," *Journal of Marketing Management* 31:1-2 (2015): 107-122, at p. 108.

60. Poster, "Databases as Discourse" (above, n. 1), p. 186.

61. Mackenzie, "The Performativity of Code" (above, n. 9), pp. 75, 77.

62. Poster, "Databases as Discourse" (above, n. 1), p. 186 (emphasis added).

63. Simon, "Return of Panopticism" (above, n. 8), p. 17.

64. Poster, *The Mode of Information* (above, n. 2), p. 111; "Databases as Discourse" (above, n. 1), p. 190. And for a discussion of credit scoring through economic categor-

Poster first points to this argument in *The Mode of Information*, noting that Foucault originally wrote of the way discourse organized practice into “structures of domination” that he termed “technologies of power,” which had a way of “objectivizing” the subject.⁶⁵ Subjects do not disappear because databases are at work as part of the data assemblage; rather, the assemblage incessantly generates identities, profiles, and proxies, and the data doppelgänger emerges from a cloud of facts as the silicon body on which decisions, predictions, preemptions, and control are exercised. Each of those exercises is an expression of power with consequential outcomes for the individual. This is closer to a manoeuvre John Cheney-Lippold makes in his examination of “digital selves”—yet with more specificity than Simon’s critique of Poster’s thesis noted above. Cheney-Lippold suggests that the statistical and computational basis of profiling produces an “interpolated subject,” which is “only talked about, not to.”⁶⁶ There is no warm-bodied person being called to or hailed in this exercise; rather, such systems ensure the interpolated subject is an array of data points that serve to communicate something about a person through his or her datafied proxy. As networked communication and datafication have expanded in the last few decades, a more complex understanding of data power and of the quantification of life is required to see beyond the construct of Poster’s “superpanopticon” and other panoptic metaphors that ground the subject in descriptive and spectral terms rather than performative simulations.

Facebook, Cambridge Analytica, and Our Data Doppelgängers

Where are some of our data doppelgängers? Are they hiding in plain sight? For the most part, they are very near indeed. Our data doppelgängers are digital containers, which can be filled and refilled, mixed and modified, with many different data sources to constitute us as data subjects. Data that has been or is being collected funnels into a database to enable decisions, whether in real-time, such as when Amazon offers suggestions for products as you search for specific items on its website, or when financial services companies such as

ization and stratification, see Marion Fourcade and Kieran Healy, “Markets and Classifications: Categorizations and Valuations as Social Processes Structuring Markets,” *Historical Social Research* 42:1 (2017): 23–51; Andrew Leyshon and Nigel Thrift, “Lists Come Alive: Electronic Systems of Knowledge and the Rise of Credit-Scoring in Retail Banking,” *Economy and Society* 28:3 (1999): 434–466.

65. Poster, *The Mode of Information* (above, n. 2), p. 90.

66. Cheney-Lippold, *We Are Data* (above, n. 10), p. 170 (emphasis added).

banks gather historical data from a range of sources to construct a risk profile in the credit scoring process when you apply for a bank loan or credit card. In these common examples, most of us have a basic understanding of what is going on and can catch a glimpse of our *doppelgänger*, but the calculative procedure through which the algorithm is coded to process and act on information as part of the decision process is not transparent. This lack of transparency about what bits of data, or data points about us, are drawn out of the database(s) and how an algorithm may make sense of the data remain a fundamentally mysterious aspect of the data cycle.⁶⁷ Companies such as Facebook, for whom privacy and data control are a work in progress driven reactively by changes in global privacy regulation, give us a suite of tools seemingly to control how much information is collected and retained and to what extent we can opt in to how that information is used and shared with third parties.⁶⁸ However, nowhere in their privacy self-audits and settings do they tell us *how* the data assemblage works; we know very little about how their database is structured and almost nothing about how they have designed their sense-making apparatus—the algorithms that parse and interpret personal data. More powerful forces of control are at work than individuals can bring under their own control simply by adjusting the parameters of their account profiles.

Very recently, Facebook has been the focus of intense scrutiny for its past relationship with a data analytics firm called Cambridge Analytica and its role in microtargeting segments of the US electorate during the 2016 presidential election. This case has provided a glimpse into the messy world of data brokering and what can happen when a third-party company goes rogue and exploits access to data beyond what—in this particular incident—Facebook claims Cambridge Analytica was authorized to do.⁶⁹ Before turning to the details of this incident, it is important to understand how our proximal selves emerge out of the data cycle in general terms. People active online are profiled at the singular, individual level when using a social network like Facebook, but of course all online activity generates personal information connected to our search history, transaction details, financial payment information, and interactions with other

67. Pasquale, *Black Box Society* (above, n. 10).

68. See Facebook.com, Data Policy, last revised April 19, 2018, https://www.facebook.com/policy.php?CAT_VISITOR_SESSION=c7b73ebc78d1681ade25473632eae199.

69. Alex Hern, "Cambridge Analytica: How Did It Turn Clicks into Votes?," *Guardian*, May 6, 2018, <https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie>.

web services. For every interaction through which we consciously give information about ourselves, including personal information that may be required to fulfill a request or complete a transaction or interact on our favorite online social platform, an array of data points flow alongside our main activity, including location data, IP address, device type and platform (mobile, laptop, or desktop), Internet browser, operating system, recently visited or referral websites, and where we go online after we complete our activity.⁷⁰ In addition, ad trackers or cookies, bits of code used by marketing companies to assemble our data points into meaningful personalization data, also enter our data stream stored by browsers, and even ad blockers and tracking obfuscation tools cannot block all online tracking. The surveillance scholar Roger Clarke, writing in 1994, noted that our digital persona performed as a “model of an individual’s public personality based on data and maintained by transactions and intended for use as a proxy for the individual.”⁷¹ These proximal selves can make things of consequence happen for the flesh and blood person; they are powerful performative simulations. Past actions and choices when analyzed alongside data from “people like us,” our digital similes, are used to make determinations about our future tendencies and actions.⁷² Not only are there more data doppelgängers proliferating at different points along the network and generated in different corporate and governmental locales, but these proximal selves are dynamic and changeable depending on the algorithmic milieu in which they were generated.

While we are tracked individually, our data is then also anonymized and aggregated with other people’s activity on the same sites using the same services and even combined with other data such as keyword searches and social network data. This process of data aggregation is significant to the process of social profiling: it requires sophisticated algorithms designed to search for patterns and similarities within large datasets and identify previously unseen relationships between data points.⁷³ Data aggregation enables a much deeper analysis by algorithms coded to search for patterns in the data to glean new insights and generate finer-grained categorizations out of this large dataset to enable microtargeting; that is, to be able to focus on a niche audience or market generated by analyzing behavioral

70. Cheney-Lippold, *We Are Data* (above, n. 10).

71. Clarke, “The Digital Persona” (above, n. 6), p. 79.

72. Rouvroy, “The End(s) of Critique” (above, n. 4).

73. O’Neil, *Weapons of Math Destruction* (above, n. 11); Pasquale, *Black Box Society* (above, n. 10); Hern, “Cambridge Analytica” (above, n. 69).

data across a large dataset. Microtargeting has emerged as a powerful technique in recent years to circulate political advertising, socially divisive false news stories, and “click-bait” on social networks to sow dissension and to influence people identified by an algorithm as susceptible to persuasion around socially and politically charged issues, including race, immigration, sexuality, and nationalist sentiment. This process of microtargeting was made visible in a spectacular way in recent months when the data access and sharing policies of Facebook meant that Cambridge Analytica and its academic partner gained access to up to 87 million Facebook accounts beginning in 2014.⁷⁴

While the story surrounding Facebook and Cambridge Analytica is significant to broad concerns around informational privacy, data security, and the question of who has access to personal information and for what purposes, it also provides a powerful example of the propagation and circulation of our data doppelgängers. Detailed information about Cambridge Analytica operations has come from the “whistleblower” Christopher Wylie, a former employee. Wylie makes clear how Cambridge Analytica created their microtargeting system from individual profiles and activities they had access to from Facebook, to the aggregation of that data within a larger dataset, and then the analytical process required to make sense of all of that data to generate specific microaudiences who would receive tailored political messaging.⁷⁵ In simple terms, the flow is from the *individual profile*, to the *aggregate dataset*, and then back out to *microtarget individuals* based on the refined knowledge gained through the power of the algorithm to make sense of the aggregated data. For the purposes of microtargeting, the data doppelgängers that emerge from the analysis require a certain specificity. This specificity is generated from *behavioral data*, or data that signals our likes and dislikes, feelings and opinions, and our choices and actions, captured in our myriad digital activities.

Wylie’s exposé details how Cambridge Analytica worked with the

74. Cadwalldr, “Steve Bannon’s Psychological Warfare Tool” (above, n. 5). Since this story was first reported in detail by Cadwalldr, Facebook has said their data policy has been tightened up to ensure third party access to Facebook account data is more closely regulated; see Facebook.com, Data Policy (above, n. 68). See also Kari Paul, “Facebook Reveals the 87 Million Accounts Affected by Privacy Violation—What to Do If You’re One of Them,” MarketWatch, April 10, 2018, <https://www.marketwatch.com/story/facebook-prepares-to-reveal-the-87-million-accounts-affected-by-privacy-violation-what-to-do-if-youre-one-of-them-2018-04-09>.

75. See Christopher Wylie cited in Cadwalldr, “Steve Bannon’s Psychological Warfare Tool” (above, n. 5).

data they harvested from Facebook profiles, but importantly, his narrative also serves to highlight the sociotechnical entanglement of people, processes, and things at work to create the targeting system. Part of the story, therefore, is in the individual choices software developers and coders made, from the training data used to design and test the targeting system, to the prejudices and preferences of the system developers themselves, to the form and function of the design process, all influencing how the system would operate once deployed. Wylie's disclosure traces how personal information and aspects of personality can be linked to political behavior through psychographic profiling because, as he divulged to Carole Cadwalladr in her exposé on the company, "Cambridge Analytica had its data. This was the foundation of everything it did next—how it extracted psychological insights . . . and then built an algorithm to profile millions more."⁷⁶ The story emerging around Facebook and Cambridge Analytica, however, is only partially exposed through investigative reporting and public hearings. It will require deeper study to unravel the decisions and actions of people involved at Cambridge Analytica to fully understand the technical development and deployment of a platform they designed to leverage personal information in often discriminatory and unethical ways. In the case of Cambridge Analytica, their use of Facebook profiles augmented by (allegedly) unauthorized use of users' Friends list gave them a very large dataset to begin their targeting campaign for clients on the right of the political spectrum in the 2016 US election. As Wylie has disclosed in media interviews, Cambridge Analytica created ads that could be targeted at Facebook users whose behavioral profile produced a data doppelgänger susceptible to influence on issues of race, immigration, gun control, and so on.⁷⁷

The performative aspect of the database in the hands of a company such as Cambridge Analytica on behalf of its political clients is, as William Bogard observes, an expression of control that is "now an inclusive, continuous, and virtual function, traversing every level . . . simultaneously molecular and planetary, no longer limited by walls or schedules,"⁷⁸ or, arguably, corporeal bodies. As we trace the flow of data from person to database, through the algorithm and back out to our generative doppelgänger, the body is literally left

76. Ibid.

77. Wylie cited in *ibid.*; and also see Hern, "Clicks into Votes?" (above, n. 69).

78. William Bogard, "Welcome to the Society of Control: The Simulation of Surveillance Revisited," in *The New Politics of Surveillance and Visibility*, ed. K. Haggerty and R. Ericson (Toronto: University of Toronto Press, 2006), pp. 55–78, at p. 59.

behind. When Haggerty and Ericson considered the data double of the surveillant assemblage, they saw it as the “formation and coalescence of a new type of body, a form of becoming which transcends human corporeality and reduces flesh to pure information.”⁷⁹ Arguably, there is no such thing as “pure” information, but the point for Haggerty and Ericson is surely to draw our attention to the incessant pull of profiling systems in which, as Cheney-Lippold offers in more recent work, the data subject is “understood in the context of statistical estimation that fills in the holes of existing data with new algorithmic approximation [as a] composite algorithmic identity.”⁸⁰ This performative simulation emerges in varying contexts as our data doppelgänger, and Cambridge Analytica has shown us just how accurate contextualized statistical estimations can be based on our proximal selves. Who needs a body when data will do?

It is important to think about how fleeting the data doppelgänger is: it gains form as an object of analysis and then decomposes; its data points drift back into the flows of information to be recompiled in another proxy. The data doppelgänger has simultaneity; there can be many propagating at once across networked systems prompting particular kinds of targeted ads or recommending people to follow on social media. Our proximal selves are often not finite; they morph and modulate, and this very instability shatters any suggestion of control the individual has over its substantive form and function merely by tweaking one’s privacy settings in online applications such as Facebook. Thinking with the doppelgänger helps to make real the sheer velocity in our datafied world in which the data cycle operates and wherein bits of personal information recombine to stand in for us amid our similar others. The performative aspect of the database, with its productive power to inscribe new identities on the subject, gives substance to the data doppelgänger and, as the Facebook and Cambridge Analytica example demonstrates, produced consequential outcomes at the individual and societal level once the targeting campaign was unleashed.

New Articulations of Power: Algorithmic Vitality

Poster suggests “[p]ostmodern culture configures multiple dispersed subject positions whose domination no longer is effected by alienated power but by entirely *new articulations* of technologies of power.”⁸¹ The collection, analysis, and dissemination of data have increased

79. Haggerty and Ericson, “Surveillant Assemblage” (above, n. 6), p. 613.

80. Cheney-Lippold, *We Are Data* (above, n. 10), p. 59.

81. Poster, “Databases as Discourse” (above, n. 1), p. 190 (emphasis added).

exponentially across every facet of life, bringing a new kind of power in which “[k]nowledge is not produced *about* the world anymore, but *from* the digital world.”⁸² Folds, connections, articulations, and attenuations are all part of an incessant modulation producing codified belonging; that is, of belonging in a fleeting instant to a category or a class to form a “compact structure” as a profiled subject to “ensure and control the identity of each agency, including personal identity.”⁸³ The database is new and improved in today’s accretive data assemblage that grows ever larger, and processes of segmentation operate to continuously divide the human subject into ever-smaller slices or data points just as the programmers at Cambridge Analytica were able to do with the Facebook dataset.

All the data being generated and the complex forces in the sense-making apparatus arrayed around it—databases and algorithms—mean that the profiling process itself becomes an augmented “perfect writing machine” able to generate infinite futures to guide and persuade us toward certain, select choices, or to preempt action to control opportunities, actions, and outcomes. However, it is difficult to think of the profiling apparatus as a “perfect” machinic enterprise, given the potentiality of the data doppelgänger to perform its “virtuoso act of imitation” across virtually every aspect of contemporary life.⁸⁴ New articulations of power are combinatorial: the sense-making apparatus is comprised of the database plus various and variable analytic techniques exercised through the power of the algorithm that altogether generates a more “perfect” regime of visibility and of intelligibility, which Antoinette Rouvroy calls “algorithmic governmentality.”⁸⁵ The regime of visibility is algorithmic and specialized, reflecting what Rouvroy notes is the computational turn in governmentality: it is not visible necessarily to those in its

82. Rouvroy, “The End(s) of Critique” (above, n. 4), p. 147.

83. Deleuze and Guattari, *A Thousand Plateaus* (above, n. 36), p. 195.

84. Webber, *The Doppelgänger* (above, n. 6), p. 9.

85. Antoinette Rouvroy, “Epilogue: Technological Mediation, and Human Agency as Recalcitrance,” in *Law, Human Agency and Autonomic Computing: The Philosophy of Law Meets the Philosophy of Technology*, ed. Mireille Hildebrandt and Annette Rouvroy (New York: Routledge, 2011), pp. 217–222, at p. 121; Rouvroy, “The End(s) of Critique” (above, n. 4). See for other discussions on visibility through social media in Tania Bucher, “Want To Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook,” *New Media and Society* 14 (2012): 1164–1180; and for aspects of visibility and power, see Samuel Mateus, “Visibility Regimes in Mediatized Publicness,” *MATRIZES* 8:2 (2014): 259–278; Evelyn Ruppert takes a related Deleuzian approach to exploring the use of distributed data in “The Governmental Topologies of Database Devices,” *Theory, Culture & Society* 29:4–5 (2012): 116–136.

sights, but rather its spectral and purposeful utility is apprehended by corporations and governments as a way to predict future action based on past behavior. The intelligibility resides in advanced computation using proprietary algorithms to parse and make sensible a broad swath of personal data, which is precisely what Cambridge Analytica did so successfully. In this new regime of power, as Scott Lash reminds us, the “structures of discourse have been displaced by structures of information” in the knowledge/power relation, and the data *doppelgänger* is a powerful actor in this algorithmic milieu.⁸⁶

Rouvroy argues that “algorithmic governmentality” is really a kind of “data behaviourism” because as a process it collects, selects, and analyzes data *about people* to produce statistical bodies that can be acted on by simulating future behavior and exercising preemptive power over the profiled.⁸⁷ This is a significant idea that keeps the analytical focus on aspects of *intelligibility* in algorithmic systems as opposed to panoptic models of visibility in the contemporary datafied milieu. The continuously active and generative apparatus of the data cycle, that is, collection, analysis, and profiling, therefore conveys what I would call an *algorithmic vitality*. This is an emergent constellation of forces within the data assemblage that constructs a new truth regime around algorithmic reason and makes the notion of data power more complex and effective. Taken together, these ideas, in keeping with a recent turn in communication studies and other disciplines toward vital materialism, ensure that the “agentic contributions of nonhuman forces” are considered,⁸⁸ particularly as powerful algorithms are convened across decision-making systems that have a capacity to act on personal information with material consequences for the reflexive human subject.

Manovich claims that algorithms are complementary to databases and a crucial part of the “computerization of culture,” as noted in the preceding section.⁸⁹ While Manovich’s focus is primarily on media systems, it is a totalizing view, leaving little room for a more critically inflected analysis. Algorithms can be thought of as an “abstract, formalized description of a computational procedure,” but it is their intervention in the social field to sort and profile people

86. Lash, *Critique of Information* (above, n. 35), p. 191; Robinson, “The Vital Network” (above, n. 45).

87. Rouvroy, “The End(s) of Critique” (above, n. 4), p. 146.

88. Jane Bennett, *Vibrant Matter: A Political Ecology of Things* (Durham, NC: Duke University Press, 2010), p. xvi.

89. Manovich, “Database as Symbolic Form” (above, n. 44), p. 84; see for related discussion Manovich, *Language of New Media* (above, n. 9), p. 198.

that many critical scholars have focused on.⁹⁰ Algorithms are part of a wide sociotechnical ensemble including people, processes, and things arrayed in communication networks, and, more specifically, in the context of their technical embedding with databases, they are part of what Rob Kitchin calls the data assemblage.⁹¹ The notion of an assemblage long predates our fascination with “big data” and datafication, and can be linked with Deleuze and Guattari’s philosophical program and also with Manuel DeLanda, who has worked to develop assemblage theory across a range of philosophical projects. DeLanda’s assemblage theory incorporates the Deleuzian notion that the dynamic and emergent properties of an assemblage arise out of heterogeneous parts assembled in relation, which “retain their autonomy, so that they can be detached from one whole and plugged into another one, entering into new interactions.”⁹² So while we may understand the data assemblage as a “thing,” its component parts are variable: data collection intensifies, data flows are mobilized, databases store and organize, and algorithms sort and output in a dynamically responsive and generative cycle.

This dynamic quality is important for two reasons. First, while algorithms are part of the data assemblage, they convey vital and lively action by assembling information into data doppelgängers for the purposes of control. The data doppelgänger is the performative feature act of the combination of a set of rules—the algorithms—applied to a particular selection of data. Its sheer changeability destabilizes any notion of fixed identity and, as noted in the foregoing sections, the profiled rarely catch a glimpse of the factors enrolled to generate this constitutive assembly. Second, the doppelgänger is the nonhuman constituent around which to trouble new aspects of the “compact structure” of the profiled subject because it operates at the boundary between the reflexive flesh and blood person and the data assemblage out of which the informational proxy arises. In particular, it is through the data doppelgänger that we can interrogate data power as a constitutive process producing rationally performative digital dossiers.

The profiled, or data doppelgängers, are distant and discordant

90. Paul Dourish, “Algorithms and Their Others: Algorithmic Culture in Context,” *Big Data & Society* (2016): 1–11, at p. 3; Bivens, “The Gender Binary” (above, n. 47); Danah Boyd and Kate Crawford, “Critical Questions for Big Data,” *Information, Communication & Society* 15:5 (2012): 662–679.

91. Kitchin, *The Data Revolution* (above, n. 3), p. 25.

92. Manuel DeLanda, *Assemblage Theory* (Edinburgh: Edinburgh University Press, 2016), p. 10; Deleuze and Guattari, *A Thousand Plateaus* (above, n. 36).

machinic selves assembled to fit within a digital cosmology that understands code as rational, information as ordered, and communication as controlled. This understanding is a persistent remnant of twentieth-century cybernetics in which Norbert Wiener “dreamed of a world where there is no ‘unknown’ left to discover, only an accumulation of records that must be recombined, analyzed, and processed” to determine the future.⁹³ It is this cybernetic logic that seeps into the contemporary profiling apparatus and ultimately “depends on the proceduralized choices of a machine, designed by human operators to automate some proxy of human judgment or unearth patterns across collected social traces.”⁹⁴ Even the terms “cyber” and “cyborg,” as Orit Halpern notes, permeate our “imaginings of digital technology, information networks, and human-machine interaction.”⁹⁵ In a similar vein, Lash interrogates the shift to a “communication order” as the register of control embedded deeply into our lives as a fundamental expression of power through, and in, the algorithm where “we swim in its ether.”⁹⁶ This shift embodies a cybernetic logic such that “[i]n the communication order, power is not just in the flows: it is in the emergent non-linear socio-technical systems that channel, block and connect the flows. Hence, literally, power through control. Cybernetic power works through command, control, communications and intelligence.”⁹⁷

Where once the cyborg embodied our human-machine couplings, the evolving complexity of the database and the algorithm demands a more evocative object through which to trace these two computational dimensions and their generative world-making powers.⁹⁸ We have such an object in the data doppelgänger, which enables us to “see” the multiplicity of figurations and fictions spun out of databases that stitch together various strands of personal data into informational proxies. The profiling apparatus, as part of the wider data assemblage, today demands that we deconstruct the parts in order to see the whole of its power and efficacy. The database and the algorithm produce our many “other selves” in an infinite regress, and

93. Orit Halpern, “Dreams for Our Perceptual Present: Temporality, Storage, and Interactivity in Cybernetics,” *Configurations* 13:2 (2005): 283–319, at p. 284.

94. Gillespie, “Relevance of Algorithms” (above, n. 10), p. 192.

95. Halpern, “Dreams for Our Perceptual Present” (above, n. 93), p. 289.

96. Lash, *Critique of Information* (above, n. 35), p. 66.

97. Scott Lash, “Power after Hegemony Cultural Studies in Mutation?,” *Theory, Culture & Society* 24:3 (2007): 55–78, at p. 67.

98. See, for related, Sherry Turkle, *Evocative Objects: Things We Think With* (Cambridge, MA: The MIT Press, 2007).

if we interrogate those databased creations, we can see the “effects of *mise en abyme*, whereby figures or structures are reflected within each other”—meaning, the data doppelgänger performs its “virtuoso act of imitation” from a repertoire that is never-ending.⁹⁹ Data doppelgängers are performative simulations; they are constructs that emerge dynamically out of the processes of profiling each time data is mined and molded for specific purposes to identify, infer, predict, preempt, and control the individual through the information gleaned from an ever-widening stream of collection points.

The Politics of Databases

As I have argued in this essay, big and small data are nothing without a sense-making apparatus generated through and by an algorithmic system that can provide a deep analysis to make data *about* individuals comprehensible in such a way as to extract valuable insights to shape actions, choices, and potential futures. For individuals, however, it is difficult to see how data’s power shapes us in ways beyond our explicit consumer experiences. When consumers search for products and services, select streaming media content, or shop online (or off), there is a sense or feeling for the data cycle because we have begun to understand in broad terms how algorithms are at work in online platforms and shape search activities, advertising, media streams, newsfeeds, and all manner of digital interaction. Even when we peel back the layers and poke around in our Google search history, examine our Facebook newsfeed, ponder the unregulated meddling by third party data analysts such as Cambridge Analytica, or explore our Netflix choices and recommendations, it is not all that revealing as to how our activities are explicitly feeding into our data doppelgänger(s), which are continuously and effortlessly mutating and multiplying.¹⁰⁰ Wylie’s explanation of the ways in which training data were used to perfect Cambridge Analytica’s targeting system lends some transparency to how the power of the database and algorithms operating as an assemblage can be mobilized as a form of social control. If we think *with* the data doppelgänger to help us understand consumer and political profiling as dynamic and changeable, it brings into focus an uncomfortable reality: contemporary efforts to regulate the collection of personal information through the rubric of informational privacy conceived as data control, even with recent efforts in the European Union’s General Data Privacy

99. Webber, *The Doppelgänger* (above, n. 6), p. 6 (emphasis added).

100. See for related discussion, Tania Bucher, “Want To Be On the Top?” (above, n. 65); Cheney-Lippold, *We Are Data* (above, n. 10).

Regulation (GDPR), have historically been unsuccessful, and may not constrain the data cycle and its incessant propagation of data doppelgängers in the future.¹⁰¹ The privacy approach will never reveal in detail how we are profiled and how the individual is reconstituted in the digital domain as the informational proxy around which consequential decisions and actions may be taken. Poster argued that we need a “politics of databases” around which to organize resistance as opposed to a politics of privacy. Might a politics of databases activate our critique *and* our resistance to data power? A politics of databases must address new “forms of agency appropriate to a dispersed, multiple subject and to generate strategies of resistance.”¹⁰² It is no longer the politics of privacy that can subtend or delimit privacy and visibility, because it is *intelligibility* that provides data with its power to produce knowledge about human subjects through information about them. Making vast repositories of data intelligible and thus actionable is the means to deciding things about people by incessantly generating profiles used in analytical procedures with consequential outcomes for the individual as witnessed in the work of the analysts at Cambridge Analytica; this is a persistent versioning process that produces data doppelgängers. Undoing or interrogating that process through our encounters with our data doppelgängers is hardly easy or trivial, and opting out, which privacy regulators encourage as one way to resist data collection, slots us into another category as a “refusnik.”

In Gilliam’s film *Brazil*, the ineffable lists, the databases, are methodically output without end, churned out of the machinery at the Ministry of Information, in a mundane, regularized way. Yet such lists are clearly constitutive and generate subjectivities in a factory-like setting. Power lies within this convoluted sociotechnical ensemble; yet even here, ghosts in the machine, or in this case a fly, can serve to inflict maximum damage with one slight key shift. This is perhaps an outlandish example, but it serves to highlight the turn to data and

101. See, for example, Canada’s approach to governing private sector privacy through *The Personal Information Protection and Electronic Documents Act* (PIPEDA) in which the first line of the Act states, “An Act to support and promote electronic commerce by protecting personal information that is collected, used or disclosed in certain circumstances, by providing for the use of electronic means to communicate or record information or transactions” (Office of the Privacy Commissioner of Canada, last updated January 23, 2018, <https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/>). While data collection, use, and retention are outlined in PIPEDA, it serves to entrench data standards and frame privacy in terms of information and data control, but does not dramatically curtail the incessant data cycle.

102. Poster, *Second Media Age* (above, n. 2), p. 93.

databases in which the sheer churn of the system erases distinctiveness and individuality. In the contemporary data assemblage, the coupling of databases and algorithms ensures a new “truth regime” as an expression of power with the “widest zone of indistinction between reality and the world” in which “algorithmic governmentality” is the standard of rational order and governance.¹⁰³ Where once neoliberalism produced the subject it needed, algorithmic governmentality does not need to produce *a* subject; rather, it needs only to function as an affective regime of power.¹⁰⁴ The algorithmic vitality of this space of potentiality means that on the Internet where we all interact as “Users,” our quantified digital traces act as feeder lines into the intelligible objects Users have become, “held together by platforms for deep biographical comparability.”¹⁰⁵ This is an aspect Cambridge Analytica potentially understood and exploited.

Resisting the “ethos of simplification” through which the big digital technology companies such as Facebook have abstracted away the “complex and messy details”¹⁰⁶ required to ensure our contemporary communication networks and interactive transactional spaces function is a daunting task—and one that I, like many people, have difficulty in navigating other than through abstention. Altogether, this make a politics of databases difficult to assemble against the well-developed profiling apparatus and the data assemblage, which produce data doppelgängers that are always already resistant interlocutors in subdividing and repackaging “deep biographical” data. The cultural imaginary of the doppelgänger troubles any notion of the unified individual and disrupts the “definition of the self as a coherent, indivisible and continuous whole which has dominated western thought for centuries.”¹⁰⁷ Data doppelgängers are proximal objects—a proxy on whose silicon body of data are inscribed facts, indicia, details, habits, records of past actions and choices arrayed as part of an anticipatory apparatus. Making our doppelgängers visible and the means of their assembly clearer is important if we want to know how these entities are generated and what they can do. Perhaps resistance can be marshaled through data transparency and public measures to democratize access to data repositories ensuring

103. Rouvroy, “The End(s) of Critique” (above, n. 4), pp. 146, 152–153.

104. *Ibid.*, p. 153.

105. Benjamin Bratton, *The Stack: On Software and Sovereignty* (Cambridge, MA: MIT Press, 2016), p. 260.

106. Ed Finn, *What Algorithms Want: Imagination in the Age of Computing* (Cambridge, MA: The MIT Press, 2017), p. 17.

107. Živković, “The Double as the ‘Unseen of Culture’” (above, n. 13), p. 126.

citizen-consumers have data agency,¹⁰⁸ but this is also increasingly complicated as our doppelgängers continue to proliferate unabated. As a start, what we may require is an awakening to the sheer disruptive force of the multiple selves of contemporary life, because whether we look into the Cloud or into the wires, it's doppelgängers all the way down.

108. Kennedy and Moss, "Known or Knowing Publics?" (above, n. 3).