

Technologies of Speculation

*The Limits of Knowledge in a
Data-Driven Society*

Sun-ha Hong



NEW YORK UNIVERSITY PRESS

New York

NEW YORK UNIVERSITY PRESS

New York

www.nyupress.org

© 2020 by New York University

All rights reserved

References to Internet websites (URLs) were accurate at the time of writing. Neither the author nor New York University Press is responsible for URLs that may have expired or changed since the manuscript was prepared.

Library of Congress Cataloging-in-Publication Data

Names: Hong, Sun-ha, author.

Title: Technologies of speculation : the limits of knowledge in a data-driven society / Sun-ha Hong.

Description: New York : New York University Press, 2020. |

Includes bibliographical references and index.

Identifiers: LCCN 2019041465 | ISBN 9781479860234 (cloth) |

ISBN 9781479883066 (paperback) | ISBN 9781479855759 (ebook) |

ISBN 9781479802104 (ebook)

Subjects: LCSH: Technology—Social aspects. | Artificial intelligence. | Algorithms.

Classification: LCC T14.5 .H636 2020 | DDC 303.48/3—dc23

LC record available at <https://lcn.loc.gov/2019041465>

New York University Press books are printed on acid-free paper, and their binding materials are chosen for strength and durability. We strive to use environmentally responsible suppliers and materials to the greatest extent possible in publishing our books.

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Also available as an ebook

CONTENTS

Introduction: Driven by Data	1
1. Honeymoon Objectivity	13
2. The Indefinite Archive	30
3. Recessive Objects	53
4. Data's Intimacy	76
5. Bodies into Facts	114
6. Data-Sense and Non-Sense	156
Conclusion: What Counts?	181
<i>Acknowledgments</i>	201
<i>Notes</i>	203
<i>Bibliography</i>	249
<i>Index</i>	283
<i>About the Author</i>	287

Introduction

Driven by Data

What counts as knowledge in the age of big data and smart machines?

We—we as that fiction of a collective public, we as individuals cut to ever finer pieces with each measurement—are becoming, like it or not, “data-driven.” Externally, smart machines and algorithmic prediction take the wheel, knitting together an expansive landscape of facts against which individuals are identified and judged. Internally, human drives—which, Deleuze understood, are not merely internal psychophenomena but themselves social structures¹—are measured and modulated through ubiquitous sensors. The rapid expansion of technologies of datafication is transforming what counts as known, probable, certain, and in the process, rewriting the conditions of social existence for the human subject.

The data-driven society is being built on the familiar modern promise of better knowledge: data, raw data, handled by impartial machines, will reveal the secret correlations that govern our bodies and the social world. But what happens when the data isn’t enough and the technology isn’t sufficient? The limits of data-driven knowledge lie not at the bleeding edge of technoscience but among partially deployed systems, the unintended consequences of algorithms, and the human discretion and labor that greases the wheels of even the smartest machine. These practical limits provoke an array of speculative practices, putting uncertainties to work in the name of technological objectivity. Weak indicators of human behavior and other fragmentary, error-prone data are repackaged into probabilistic “insights,” whose often black-boxed deployment now constitutes a global industry. Futuristic imaginaries of posthuman augmentation and absolute predictivity endow today’s imperfect machines with a sense of legitimacy. In the process, technologies of datafication are reshaping what *counts* as knowledge in their own image. From

self-surveillance to counterterrorism intelligence, the business of datafication hinges on redefining what kinds of data in whose hands should determine the truth of who I am or what is good for me.

The moral and political question, then, is not simply whether datafication delivers better knowledge but how it transforms what counts in our society: what counts for one's guilt and innocence, as grounds for suspicion and surveillance, as standards for health and happiness. Datafication thus reprises the enduring dilemma around the modern ideal of the good liberal subject: individuals who think and know for themselves, their exercise of reason founded on free access to information and the possibility of processing it fairly. New technologies for automated surveillance and prediction neither simply augment human reason nor replace it with its machinic counterpart. Rather, they affect the underlying conditions for producing, validating, and accessing knowledge and modifying the rules of the game of how we know and what we can be expected to know. The promise of better knowledge through data depends on a crucial asymmetry: technological systems become increasingly too massive and too opaque for human scrutiny, even as the liberal subject is asked to become increasingly legible to machines for capture and calculation.

The Duality of Fabrications

These dilemmas show that when big data and smart machines produce new predictions, new insights, what they are creating are fabrications: a process by which approximations are solidified into working certainty, guesswork is endowed with authority, and specific databases and algorithms—and all the biases and heuristics they embody—are invested with a credibility that often outstrips their present achievements. To call these activities fabrications does not mean that datafication is merely a con of epic proportions. The word originates from *fabricare*, to manufacture with care and skill; a manufacturing that every kind of knowledge system, from science to religion, undertakes in its own way. To analyze datafication in this way is to understand how data is seizing and affirming its position as truth-maker today.

Often, such fabrications involve highly accurate and sophisticated measurements that tend to perform best within tightly prescribed pa-

rameters. At the same time, their application to real-world problems often relies on arbitrary classifications, messy data, and other concealed uncertainties. Exercise trackers combine advancements in miniaturized sensors with rough heuristics, such as ten thousand steps per day—a figure originally invented by mid-twentieth century Japanese marketers to sell pedometers—to produce their recommendations. Large-scale systems, such as electronic surveillance systems for counterterrorism purposes, embed layers of human labor and decision-making into a process that is ultimately black-boxed to the ordinary citizen. The connections between data, machines and “better knowledge” remain obscure for most of us, most of the time. In many concrete cases, the claim to better, more objective knowledge through data also depends on shifting expectations around what looks and sounds like reliable truth.

Fabrications are therefore ambiguous and unstable things. Imperfect algorithms, messy data, and unprovable predictions are constantly intersected with aspirational visions, human judgment, and a liberal dose of black-boxing. Importantly, such a duality is normal to the work of datafication: a feature, not a bug. Accordingly, the solution is not as simple as a bug report that sorts out the good kinds of data-driven knowledge from the bad. Such clean and neat distinctions are not always possible and risk supporting the technocentric imagination that a few rounds of bug-fixing would allow the data to truly provide better knowledge anywhere and everywhere. Instead, this book traces underlying patterns in how such claims are made—an approach that has been crafted across areas such as sociology of knowledge, history of science and technology, and critical data studies.² The manufacturers and distributors of data-driven fabrication do not simply “cheat” truth. Rather, they are playing the game of making certain kinds of truth count. What emerges is not so much a whole new regime of knowledge but new opportunities for extending, distorting, and modifying long-standing tendencies for how we use numbers and machines to make sense of our worlds.

This approach also situates the technologies of our time in the long history of data, quantification, and social sorting. As buzzwords of the day, big data or smart machines have a short and specific life span (even if they, like artificial intelligence, often end up being recycled). But the underlying shift in what counts as knowledge often outlasts those moments in the spotlight. Joseph Weizenbaum, a pioneer of AI, had

identified this dynamic in an earlier generation of computing technologies: that far before and far in excess of computers being shaped to serve humans, humans are asked to become more compatible with the machines themselves.³ From the human body pressed into mechanical action in the Fordist factory, as immortalized in Charlie Chaplin's *Modern Times*, or the twenty-first-century population of "ghost workers"⁴ performing invisible, low-paid labor to support AI systems, the sublime spectacle of computing power constantly relies on a scaffolding of machine-compatible humans. From an epistemological standpoint, the fabrications captured in this book also echo the social life of earlier technologies for datafying bodies and lives, where the gradual normalization of modern attitudes toward numbers and statistics, then machine-driven databases, as objective fact was often achieved for and through specific political exigencies of the day.⁵

Similarly, today's fabrications are thoroughly imperfect and inescapably political. Insofar as the data-driven society is built on the bullish promise of a world run more rationally and objectively, this optimism feeds off contemporary anxieties about the seemingly growing uncertainties of modern life. There is the global diffusion of micro-threats in the "war on terror," emblemized by the almost random possibility of a "lone wolf" attack, or the heightened pressure for citizens to optimize their everyday life routines to survive the neoliberal market. Yet even as uncertainty functions as the bogeyman Other to the seductive promises of datafication, such knowledge is often achieved precisely by putting uncertainties to work. In the gaps between the fantastic promises of technology and its imperfect applications, between the reams of machine-churned knowledge and the human (in)ability to grasp it, grows a host of emergent, speculative practices that depend on the twisted symbiosis of knowledge and uncertainty.

Out There, In Here

This book examines two sites where datafication is turning bodies into facts: shaping human life, desire, and affect into calculable and predictable forms and, in doing so, changing what counts as the truth about those bodies in the first place. The first is the Snowden affair and the public controversy around the American government's electronic

“dragnet”⁶ surveillance technologies, built to quietly collect phone, email, and other electronic communications data at an unprecedented scale. The second is the rise of miniature, automated tracking devices for the monitoring of everyday life, from exercise levels to emotional status, and the subsequent analysis of that data for personalized “insights.” Surveillance by the state, surveillance by oneself—these practices reflect the expanding reach of big data’s rationality across established boundaries of the public and the private, the political and the personal.

On December 1, 2012, one “Cincinnatus”—namesake of that mythical Roman embodiment of civic virtue—contacted journalist Glenn Greenwald to request an encrypted conversation. He received no reply. Six months later, Cincinnatus was revealed to be Edward Snowden, formerly a subcontractor for the National Security Agency, now a fugitive wanted by the United States. Having eventually succeeded in reaching Greenwald, he enlisted the journalist’s help in leaking a massive cache of classified information, revealing a sprawling range of high-tech surveillance programs wielded by the US and other Western governments.

Somewhere on the way, a philosophical question had emerged: What can the public know, and what is made the public’s duty to know? The programs Snowden publicized entailed the collection of personal communications data in enormous quantities through methods designed to remain totally imperceptible to the population subject to it. This data would be harnessed toward predictive calculations whose efficacy often cannot be publicly (and, sometimes, even privately) proved. As the leaks fanned an international controversy starring lawsuits and policy debates, award-winning documentaries, and presidential speeches, the public was caught in uncertainty. One letter to *The New York Times* read: “What kind of opinion can a citizen have of his government when his government is unknown to him, or, worse, is unknown to itself? After 9/11, we found ourselves in a state of war with faceless terrorists . . . but those we have empowered to protect us use methods that we cannot see, taste or smell.”⁷ Popular book titles spoke of shadow governments, dragnet nations, and no place to hide.⁸ Such metaphors spoke to a deep sense of asymmetry: How can ordinary human subjects know the world “out there,” a world governed by increasingly vast and complex technological systems, a world that seems to begin where our personal experiences and lived worlds fall away?

As America and the world grappled with the implications of Snowden's leaks, similar dilemmas around knowledge and uncertainty were playing out through a very different set of fantasies around progress and empowerment. In September 2011, Ariel Garten took to the stage for a TED Talk—a series famous for providing slick, punchy briefs about the pressing problems of the day, and, more often than not, optimism that they can be overcome through technological and social innovations. Garten was well suited for such a stage. Juggling a life as a fashion designer, psychotherapist, artist, neuroscience researcher, and entrepreneur, she could present a figure of someone riding the waves of the newest technologies, someone standing at the threshold of the near future. Garten enthused about a wearable brainwave sensor on her forehead—an electroencephalography device that would soon go on sale by the name of “Muse.” It will tell us how focused or relaxed we are, she said, revealing aspects of ourselves that had previously been “invisible”:⁹

My goal, quite simply, is to help people become more in tune with themselves. I take it from this little dictum, “Know thyself.” . . . I'm here today to share a new way that we're working with technology to this end to get familiar with our inner self like never before—*humanising technology* [emphasis mine] and furthering that age-old quest of ours to more fully know the self.

As the American government invested massive sums into data-driven, predictive surveillance systems, its tech enthusiasts and entrepreneurs were using similar techniques to pursue an individualistic and posthuman vision: the human subject—ever a blind amnesiac, fumbling its way through the maze that is its own body and mind—would be accompanied by machines that would correct its memories and reject its excuses. Technologies of self-surveillance, overlapping across categories such as biohacking and lifelogging, use miniaturized smart machines to enable highly persistent and automated processing of human life into data-driven predictions. From the predictable, such as measures of exercise and sleep, to the bizarre, such as sex statistics (thrusts per minute), self-surveillance promises to bring home the benefits of datafication, enabling a more objective basis for knowing and improving the self.

The transformation of the everyday into a persistent backdrop of measurements and nudges promises unprecedented knowledge for the human subject precisely by shifting accepted norms around what counts as better knowledge. At one level, these machines track individuals in ways inaccessible to the human subjects' own cognition and experience—either by measuring phenomena beyond the capacity of the human senses, such as the electrical conductance in the skin, or by measuring at a frequency and granularity that people realistically cannot keep up with. The problem of what we can and must know is thus brought back from “out there” to the “in here” of the individual body and life. What does it mean to “know myself” if that knowing is achieved through mass-produced, autonomously operative devices? What kind of relationship to knowledge is produced when machines communicate ceaselessly with the body and with each other along channels that my conscious reflection cannot ever access? In many ways, the pursuit of the datafied self reenacts Weizenbaum's dictum: the capture of bodies for predictive analytics encourages those bodies to behave in ways that are most compatible with the machines around them—and, by extension, the institutions behind those machines. The good liberal subject is thus rearticulated as tech-savvy early adopters (who are willing to accept the relations of datafication before their validity is proved) and as rational, data-driven decision makers (who learn to privilege machinic sensibility above human experience).

The book traces the public presentation of state and self-surveillance across multiple sites where the technologies and their associated fantasies are proclaimed, doubted, justified, and contested. This includes the media coverage, leaked government files, lawsuits and Senate hearings around the Snowden affair (2013–), as well as advertising and promotional discourse, media coverage, and conversations with entrepreneurs and enthusiasts around the rise of self-surveillance technologies (2007–).¹⁰ I also draw on observational fieldwork of the Quantified Self (QS), an international community of self-trackers that has played a key role in popularizing the technology from a niche geek endeavor to a market of millions of users. Despite clear differences in the specific configuration of state and corporate interests, the interpellation of citizens and consumers, certain ways of thinking and dreaming about datafication recur across these contexts. Chapter 1 lays out the technological fan-

tasies that help justify, make sense of, and lend excitement to concrete systems of data-driven truth making. The promise of better knowledge is here broken down into a historically recurring faith in technoscientific objectivity, through which datafication promises a certain epistemic purity: a raw and untampered representation of empirical reality, on which basis human bodies and social problems might also be cleansed of complexity and uncertainty. These fantasies serve as navigational devices for the rest of the book.

Chapters 2 and 3 examine the predicament of the public: the people who are supposed to know for themselves, to exercise their reason, in the face of data-driven surveillance. Focusing on the Snowden affair, I argue that technologies of datafication often provoke paranoid and otherwise speculative forms of public knowledge and political participation. Ideal norms like transparent governments and informed, rational publics flounder when confronted by technological systems too large, too complex, and too opaque for human scrutiny. The Snowden files, and the electronic surveillance systems they describe, are thus recessive objects: things that promise to extend our knowledge but simultaneously manifest the contested and opaque nature of that knowledge. For both the American public and the intelligence agencies themselves, the surfeit of data provides not the clarity of predictive certainty but new pressures to judge and act in the face of uncertainty.

Chapter 4 then turns to self-surveillance and its promises of personal empowerment through the democratization of big data technologies. Paradoxically, this narrative of human empowerment is dependent on the privileging of machinic senses and automated analytics over individual experience, cognition, and affect. These new technologies for tracking and optimizing one's daily life redistribute the agency of knowing in ways that create new labors and dependencies. The chapter further traces how the Quantified Self is scaling up to the Quantified Us. Systems of fabrication first created for individual self-knowledge are gradually integrated into the wider data market, opening up new avenues of commercialization and control.

Chapters 5 digs into concrete techniques of fabrication, namely, how uncertainties surrounding terrorism and its attendant data—emails, web browsing, phone calls—are crafted into data-driven insights. Beneath and between the supermassive streams of data and metadata, impro-

vised heuristics of speculation, simulation, and deferral help produce actionable knowledge claims. These are furtive sites in which specific forms of speculation and estimation are forged into sufficiently true evidence. Such techniques, so fragile and improvised when examined up close, ride the waves of broader technological fantasies around objectivity and progress. Chapter 6 shows how self-surveillance is presented as a historically inevitable step towards a vision of posthuman augmentation that I call “data-sense.” The imperfections and contradictions of technological factmaking are thus consolidated into a normative demand that human subjects know themselves through data and thereby learn to live and reason in ways that help advance its self-fulfilling prophecy of better knowledge.

The book thus examines a number of different junctures in the social life of technologies of datafication. Chapters 4 and 5 offer a closer look at the technological side of how personal data is produced and leveraged, and their implications for human judgment in concrete sites of decision-making. Chapters 2 and 3 are more focused on how these technologies interact with existing political realities, challenging popular norms and expectations around rational publics and governments. Bookending these analyses are chapters 1 and 6, which attend to the underlying fantasies about technology and society that shape these specific practices of speculation.

Taken together, these scenes of datafication demonstrate the duality of fabrication and how the pursuit of data idealizes—and undermines—the figure of the good liberal subject. Edward Snowden justified his whistleblowing of electronic surveillance programs with the argument that the American people must learn the truth about their own datafication. Yet how can the public fulfill its Enlightenment duty—*sapere aude!*—to have the courage to use one’s own understanding—when systems of datafication recede “out there,” beyond the horizon of individual experience and knowability? The Quantified Self community explicitly cites history: the ancient Delphic maxim *gnothi seauton*, “to know thyself.” But that knowing involves brokering a very different relationship between the self that knows, the knowledge that is allegedly their “own,” and the machines that make it all possible. It is precisely such messy, speculative moments that matter for how standards of truth are being transformed. They are zones of transition, where new ways of proving and truth speaking are

accorded the status of “sufficient” certainty to meet highly practical exigencies.¹¹ It is worth remembering that big data’s “bigness” is not a matter of absolute thresholds but a relative one where qualities such as the volume and variety of the handled material exceed older bottlenecks and human limitations.¹² Yet those fleshly bottlenecks had served a function: they had slowed things down long enough for the exercise of judgment, debate, and accountability. Those opportunities for human intervention are now being systematically disrupted and overwritten. Algorithms, especially because so many tend to be classified or proprietary, themselves become sources of uncertainty because they introduce layers of mediation that become opaque to human scrutiny.¹³ Across state and self-surveillance, the pursuit of better knowledge constantly reframes the distribution of rights and responsibilities across the subject meant to know, the ever-growing panoply of machines surrounding that subject, and the commercial and governmental interests behind those machines.

Technological Defaults

The stakes of data-driven fabrications, of the changing standard of what counts as truth, cannot be confined to epistemology, but relate directly to questions of power and justice. This is a truism that bears repeating, for postwar technoscience as industry and vocation has accumulated an enduring myth of depoliticization. The idea that one merely pursues objective truth, or just builds things that work, serves as a refuge from the messiness of social problems.¹⁴ The question of what counts as knowledge leads directly to questions of what counts as intent, as prosecutable behavior, as evidence to surveil and to incarcerate? What kind of testimony is made to count over my own words and memories and experiences, to the point where my own smart machine might contest my alibi in a court of law? What constellation of smart machines, Silicon Valley developers, third-party advertisers, and other actors should determine the metrics that exhort the subject to be fitter, happier, and more productive?

Big data and smart machines push the bar toward a society in which individual human life, sensory experience, and the exercise of reason is increasingly considered unreliable. At the same time, what might other-

wise look like flaky numbers, prejudiced estimates, or dubious correlations are upgraded into the status of data-driven insights, black-boxed from public scrutiny and fast-tracked to deployment. This book argues that fantasies of machinic objectivity and pure data work to establish datafication as a technological default, where ubiquitous surveillance at both populational and personal scales are presented to the public as not only a new and attractive technology but also an inevitable future.

This default is a bottleneck for the imagination, for the ability to devise and build consensus around the kinds of policies, ethical imperatives, social norms, and even technologies that might help us manage the consequences of datafication. The answer is not, however, to run in the other direction, to romanticize humanity before the internet. Such atavism reproduces the myth of a sovereign, independent subject, one which might be resuscitated simply by detoxing ourselves from technological influence. There is no returning to the mythical time of the good liberal subject, and the transparent disclosure of the ever-expanding webs of datafication will not in itself restore the capacity for rational self-determination. Instead, we should ask of data's promise of better knowledge: What good does it really do to "know"? What other conditions, beyond the often narrowly defined metrics of accuracy and efficiency, are necessary to ensure that knowledge empowers the exercise of human reason? How can those conditions be protected as the process of knowing is increasingly overtaken by opaque systems of datafication? As I elaborate in the conclusion, asking these questions requires disrespecting the stories that data tells about itself, to refuse its rationalization of what looks like objectivity or progress, and to hold technology accountable to standards that are external to its own conditions of optimization. To refuse technology's rules of the game is to refuse the steady entrenchment of a rationality where datafication and its knowledge claims are increasingly neither by or for "us"—the human subject, the individual, the rational public—but pursues its own economic and technical priorities.

Ultimately, this book is a story of how datafication turns bodies into facts—a process that aspires to a pure and pristine objectivity but, in practice, creates its own gaps and asymmetries. The ambitious projects for state and self-surveillance reveal crucial gaps between the narrow reaches of human knowability and the vast amounts of data harvested by machines, between the public that is supposed to know

and the institutions and machines that are meant to know it in their stead, between the practical capabilities of data technologies and the wider fantasies that give them legitimacy. In each one, we find troubling asymmetries in how different bodies are treated to different kinds of factmaking. If data expands the vistas of human action and judgment, it also obscures them, leaving human subjects to work ever harder to remain legible and legitimate to the machines whose judgment they cannot understand. Caught in an expanding and consolidating data market, we cannot simply seek more and better knowledge but must rethink the basic virtues and assumptions embedded in that very word. What kind of good does knowing do? Or, rather, what must our knowledge look like that it may do good? And who are we, with what kinds of capabilities and responsibilities, with what role to play in a data-driven society? As the truth of who we are and what is good for us is increasingly taken outside ourselves and human experience, the figure of the human subject—which, Foucault had warned, is a young and temporary thing¹⁵—is flickering uncertainly, unsure of the agency and moral responsibility we had worked so hard to attach to it.

Honeymoon Objectivity

In 2014, a baby-faced, twenty-two-year-old entrepreneur named James Proud crowdfunded a sleep-tracking device that promised to automatically monitor sleep patterns, provide a numerical score, and make recommendations for sleep behavior. That such functions were already available did not escape Proud. Beddit, a sleep sensor that we will revisit in chapter 4, had been crowdfunded a year before and already released to its backers. In response, Proud chose to emphasize his device's "simple, uncomplicated and useful" qualities; designed as a slick, minimalist off-white orb, it would merge invisibly into the everyday flow of attention and reflection. "We believe technology needs to disappear," said Proud; "everything in [our device] is just designed to fade away."¹ It would carry an equally simple and no-brainer name: Sense.

In 2017, James Proud, now twenty-five, announced the end of Sense.² Panned by some tech reviewers as a "fundamentally useless" object³ and a glorified alarm clock, the device never quite delivered the quiet transformation of everyday life that its creator aspired to. Fundamentally, it proved not very good at making sense of human sleep. Users reported that any deviation from the presumed sleep scenario—for instance, a pet snuggling up in bed—would throw the device off entirely. The chaos of everyday life rarely conformed to the expectations of the tracking machine, even as its selling point was that it would discover truths about us that we cannot perceive ourselves. As Proud's team wound down operations, users began to report that their Senses were losing functionality. The orbs went mute and deaf to the data around them, a small monument to the unfulfilled promises of new technologies.

Technologies of datafication reconfigure what counts as truth and who—or what—has the right to produce it, and not simply through the success of indisputably superior machines or even their mundane

and ordinary operations in concrete practice. As relatively nascent technologies fast-tracked to the status of a global buzzword, the very idea of big data and smart machines—the spectacular keynotes and product launches, the anticipatory rhetoric, the science fiction, the projected future functions—operates as a social actor in its own right.

Proud's story, after all, is a common one. The popularization of self-surveillance technologies followed decades of anticipation (and disappointment) about a future that was always advertised as just about to arrive—a “proximate future.”⁴ A computing magazine included “smart appliances” in a 2007 piece about the “biggest technology flops,” deriding the “bubble” around smart appliances back at the turn of the century. “The bubble burst, and we haven't heard much about intelligent appliances since,” the article said.⁵ That very year, the Quantified Self (QS) community would emerge in Silicon Valley; by 2011, the Internet of Things was back on the forefront of the imagined future, featuring in the tech advisory firm Gartner's influential “Hype Cycle for Emerging Technologies” report for the year.⁶ Yet this return to the spotlight of the imminent future was not necessarily built on clear and proven cases of better knowledge. The rapid growth of the self-surveillance industry provoked public skepticism, academic research alleging negligible or backfiring effects,⁷ and even lawsuits challenging the basic accuracy of popular measuring devices (namely, *Brian H. Robb v. Fitbit Inc., et al.* 2016). The broader industry of smart machines was no better off; one internet-connected juice maker cost \$400, but its proprietary juice packs turned out to be just as squeezable by hand. A smart lock automatically updated over wireless connections and then locked users out of their homes; smart salt shakers promised voice-activated controls but were unable to grind salt. The proximate future was cobbled out of Eric Kluitenberg has called imaginary media:⁸ prototypes depicting impossible realities, products sold on the basis of never-quite-actualized functions, artists' sketches, and bullish press conferences. Even as they malfunction and disappoint, they help drag impossible functions and nonexistent relations into the realm of the sayable and thinkable. Consumers are asked to buy into the latest gadget in anticipation of its future ubiquity, to install software for its future functions, and to celebrate prototypes for their glimpse of what, surely, must be just around the corner.

Technologies of datafication seize the authority to speak truth not by achieving the improbable hype gathered around it but by leveraging those lofty goals to mobilize the public, siphon funding, and converge collective imagination. Technology thus operates not merely as tools and functions but also as a panoply of fantasies—about machines that know us better than we know ourselves, about predicting the future with pinpoint accuracy, and about posthuman cyborgs and Big Brothers. To say “fantasy” does nothing to undermine the unique importance of material facts (as if fantasies could be sustained, or even generated in the first place, without the affordances of concrete things!). But it does mean tracing the ways in which data-driven surveillance seized its claim to knowledge by mobilizing projections and estimations about technology and the future world that will necessitate those technologies. While tracking devices such as Proud’s were crafting an optimistic technofuture animated by consumerism, tugging on the broader imaginary of posthuman augmentation, state surveillance systems were warning of a future that must not happen, predictions of crime and terror that must be snuffed out through strategies of “zero tolerance.” Across both cases, fantasy takes half a step outside present reality not to escape from it but to all the more effectively guide it.⁹ Žižek once observed of the dystopian science-fiction film *Children of Men*¹⁰ that

Hegel in his aesthetics says that a good portrayal looks more like the person who is portrayed than the person itself. A good portrayal is more you than you are yourself. And I think this is what the film does with our reality.

The market projections, promotional rhetoric, bullish claims, and dystopian warnings surrounding datafication today are precisely the little doses of fiction used to make sense of these technologies and the knowledge they promise. Such beliefs are not reducible to “intellectual mistakes” by naïve or ignorant subjects. This (mis)recognition of what technology does and could do, the benefit of the doubt and the doubtful benefits, is so often a crucial part of getting technoscience off the ground.¹¹

When New Technofantasies Are Old

Round about stood such as inspired terror, shouting: Here comes the New, it's all new, salute the new, be new like us! And those who heard, heard nothing but their shouts, but those who saw, saw such as were not shouting.

—Bertolt Brecht, *Parade of the Old New*

Fantasy, in this collective, commercialized, politicized form, is never truly free-form. Datafication often falls lockstep with familiar narratives around machines and rationality, tapping into that modern drive to order the world as a taxonomy of facts for a sense of legitimacy and plausibility. As Lauren Berlant shows, these familiar anchors help stitch together the contradictions and disappointments of technology, the gaps between knowledge and uncertainty, into a sense that “the world ‘add[s] up to something,’” even when that belief is constantly displaced and disappointed.¹² The paradox throughout this book is that technologies of datafication rely so heavily on the imagined legacy of the Enlightenment, and its particular alliance of objectivity, human reason, and technological progress, even as its deployment threatens to destabilize the presumed link between information, human Reason and democratic freedoms. Since its emergence over the mid-nineteenth century, the thoroughly modern concept of technology has depicted a world whose every aspect stands ready to be flattened, standardized, and turned into problems that the ceaseless march of new inventions would render into objectively optimal states.¹³ The fabrications explored in this book leverage what we might call honeymoon objectivity: the recurring hope that with *this* generation of technological marvels, we shall establish a universal grounding for our knowledge, a bedrock of certainty, a genuine route to the raw objective layer of the world around us. By invoking this long quest, tracking technologies are able to draw together their own imperfections, uncooperative material conditions, incompatible and otherwise resistant humans into a seductive vision of better knowledge.

The objectivity invoked by data-driven surveillance constitutes no rigid dogma but a sedimented range of attitudes and affects embracing a distinctly modern way of thinking and feeling about knowledge. As extensively chronicled by Lorraine Daston and Peter Galison,¹⁴ older

renditions, such as the scholastic *obiectivus/obiectiv*, generally involved definitions starkly different from the modern. Even in Kant, objective validity meant general “forms of sensibility” that prepare experience, while the “subjective” referred to specific and concretely empirical sensations. It is only during the nineteenth century that the now-familiar juxtaposition emerges: a neutral, “aperspectival” objectivity as the privileged instrument toward truth and scientific inquiry and biased, unreliable subjectivity as its nemesis.¹⁵ By the later nineteenth century, Daston and Galison identify a dominance of “mechanical objectivity”: a regulative ideal that called for the elimination of the human observer from the process of data visualization. The critical impulse for these conceptualizations was, of course, the advent of photographic technology, which provoked new theories and standards for what counts as visual truth and who (or what) might be best equipped to produce it. Photography – despite its own long history of manipulation and contested meanings – thus spurred new linkages between automation and objectivity, producing the ideal where “machines [would be] paragons of certain human virtues” precisely by ridding themselves of human subjectivity.¹⁶

The public presentation of data-driven surveillance leverages these older ideals of objectivity, and the cultural capital it had accumulated through traditions of scientific inquiry. In self-surveillance’s effort to map the microbiome or record every moment of sleep, we find a conception of the body as an aggregation of correlations. Health, productivity, and happiness are broken down into a set of hidden but logical relationships that machines might read and catalogue—the same kind of correlations that might help predict the lone wolf terrorist, enabling an orderly distribution of risk and suspicion across the population. In this cultural imaginary, the world is an indefinite archive, and the machines of tomorrow, if not of today, will be up to the task of cataloguing it.

All this is not to say that the Enlightenment ever bequeathed a singular doctrine about technology and reason or that different practices of datafication share a totally coherent conception of a value such as objectivity. As Lorraine Daston herself noted, each historical rendition of objectivity expresses not some immutable quality rooted in natural law but a *mélange* of aspirational values that happen to occupy (or, at least, contest) a normative position at the time.¹⁷ Indeed, honeymoon objectivity describes that recurring tendency to *claim to* make new advances

toward such an immutable quality, even as the kinds of data actually produced might diverge significantly from this vision. Technologies of datafication do not subscribe neatly to any single definition but cobble together different popular imaginations of what objectivity looks and feels like. The central proposition of mechanical objectivity—the preference of nonhuman processes over subjective ones for the reliable production of knowledge—is retained as a basic article of faith, but one that is routinely transgressed and compromised in practice. The messy and flexible ways in which the virtue of objectivity is “localized” onto self-surveillance cultures reflects, above all, how broad and pliable the word has become. Like culture, objectivity exhibits a certain “strategic ambiguity.”¹⁸ Its many possible permutations allow a wide variety of interpretations and attitudes to rally behind a common banner, where more fine-grained definitions might have splintered them. Thus, the fantasy endures to pass on its allure to another institution, another machine.

Pure Data

If the pursuit of objectivity, in all its strategic ambiguity, is the well-advertised benchmark of data-driven surveillance, an equally crucial question is: What kind of regime of knowledge, what kind of social order, is it meant to deliver? This book argues that the many articulations of data’s benefits, capacities, its factmaking powers, revolve around a mythologization of data as pure and purifying. This pattern emerges not so much in efforts toward the technical definition of data, but in the public discourse, where the very question of what data *is*—or, rather, what can data *do*—again involves a messy plurality of ideas and dispositions. Data, fact, information, and knowledge are often conflated such that they are either seen to naturally follow on from each other, bolstering a sense of legitimacy.¹⁹

Three years after the first Snowden leaks, BBC4 released a documentary titled *The Joy of Data*. Its host, the mathematician Hannah Fry, boiled it down to a pyramid. From bottom to top, she explained, data is “anything that when analysed becomes information, which in turn is the raw material for knowledge—the only true path to wisdom.”²⁰ Fry left unsaid what exactly knowledge and wisdom were, but the hierarchical relation was clear: the raw objective facts gathered through new

technologies would serve as the foundation for better knowledge. This DIKW pyramid—data, information, knowledge, wisdom, in ascending order—is a fixture in many computer science textbooks. Underpinning it is a world in which everything we can or need know is reducible to positivist facts, and by descending to this atomic layer, we will be able to recover objective data for any problem.²¹

In this articulation, data and knowledge are inseparable bedfellows. Data is the ubiquitous ingredient in the buoyant dreams of better knowledge, the object unto which the hopes and fears of technological and epistemic possibility are invested. In its most elementary form, it is described as raw data: data generated by the machine but yet to undergo “secondary” processes of statistical analysis, cleaning, visualization, aggregation, and so on. It is data fresh out of the sensor, with no artificial additives. In this telling, raw data is seemingly anterior to analysis, classification, and attribution of meaning. The valorization of raw data is intimately connected to self-surveillance’s vision of empowerment through objective knowledge. In 2015, one QSer suggested raw data access as one of the three “freedoms of personal data rights”:²²

Without raw data, we are captive to the “interface” to data that a data holder provides. Raw data is the “source code” underlying this experience. Access to raw data is fundamental to giving us the freedom to use our data in other ways.

Similar sentiments were expressed by a host of prominent commentators, including QS co-founder Gary Wolf.²³ The widespread implication of raw data’s nonmediated nature translates into the fantasy of data as a purifying agent: a technology that will produce knowledge stripped clean of politics, of human bias, and of troublesome differences in opinion and establish the clear and rational path forward. Yet, as numerous scholars have pointed out, the very idea of raw data is an oxymoron.²⁴ Data only becomes data through the human design of each algorithm, relational database, and deep learning system—although there are important differences in how much detail is determined by manual design and judgment and how much is left up to machine learning.²⁵ Data is no thing-in-itself that exists prior to observation but something

to be “achieved”²⁶ through a concerted process of production that can never rid itself of human subjectivity and sensibility.

The dangerous consequence of this aspiration to purity is that the human, social, historical, and moral aspects of data’s fabrications are invisibilized—allowing familiar kinds of speculation and prejudice to reenter by the back door. Consider the effort to predict and intercept terrorists before they can cause harm. Chapter 5 examines known cases of sting operations where certain individuals—predominantly young, Muslim, of Arab descent, male—are marked out by state surveillance apparatuses for fabrication. Driven by a moral and political injunction to “zero tolerance,” in which even a single terrorist attack is an unacceptable failure of prediction and control in the wake of the September 11 attacks, counterterrorist operations do not simply wait for the data but actively work to produce the necessary proof. Thus, in the case of Sami Osmakac, FBI undercover agents supply the individual with money and the weapons and explosives to be purchased with that money, and coach him each step of the way until arrest can be justified. Meanwhile, the Snowden files reveal the surveillance programs themselves to be inevitably human. Analysts from the National Security Agency (NSA) speak of “analysis paralysis” and the struggle to handle supermassive volumes of data, while placeholder entities, such as “Mohamed Badguy” and “Mohammed Raghead,” for database-search interfaces reflect the all-too-human, all-too-crude underside of sophisticated data-driven systems. Criticisms of datafication have often invoked labels such as data doubles and doppelgängers to warn against how individual self-expression is being replaced by alternative identities recomposed from data extraction.²⁷ Alongside such “copies,” we also find a variegated ecosystem of speculative entities: the Osmakac that might have been, the Raghead in the database. Here, datafication provides no mathematical certainty but a range of possible outcomes and correlations to legitimize highly anticipatory forms of surveillance, judgment, and incarceration. The desire for epistemic purity, of knowledge stripped of uncertainty and human guesswork, ends up with concrete practices that draw perilously close to the imaginations of purifying the nation and the body politic. Suspected terrorists, brown or white, religious fanatic or ethnonationalist, end up subject to very different forms of datafication even as the technology promises a neutral illumination of truth.

The idea that raw data can access an orderly and calculable reality stripped of historical forces, social constructions, and political disputes translates into a converse fantasy: that the individual body will be purified of the elements that impair its health and productivity, and the body politic too will be secured as a transparent whole. Yet a central argument of this book is that such a fantasy of epistemic purity—of knowledge untainted, of complete certainty—itself carries political and moral biases. The belief in raw and untainted data begets not only an excessive reliance on algorithmic factmaking but also extends the older and deeper cultural desire for sorting the world into stable and discrete pieces. The recurring temptation: What if we could predict and eliminate every bad apple, every violent individual, and every criminal intent? What if we could maximize everything good about our bodies, expel all the toxins, cut off the bad friendships, and optimize every habit? Just as the pursuit of better knowledge through datafication entails a social shift in what counts as objectively true, the collective faith in the purity of data entails using the data to try to bypass important political and moral questions, to try to purify bodies through technological solutions.

The Groundless Ground

The mythologization of pure data puts into ironic relief the original Latin: *data*, meaning “that which is given.” Today, (raw, big) data’s privileged position in objective inquiry and knowledge production seeks to normalize into the woodwork, becoming “something we would *not* want to deconstruct.”²⁸ It has been called the “cathedral of computation,” or a faith in “computationalism”: the fantasy that data simply *is* and shall provide a reassuring grounding for everything else that trouble us.²⁹ This faith has immediate and practical rewards. If datafication promises objectivity, impartiality, and predictivity, all these epistemic characteristics add up to a valuable sense of *stability*. On one hand, specific processes of data-driven analytics work within narrowly defined parameters where inputs may be standardized, modeled, and otherwise manipulated. In other words, the algorithm’s truth claim itself relies on a set of grounding assumptions about the world out there and its methodological relation to data—assumptions that it agrees not to question to get the job done. At the level of data as a broader, popular imaginary,

telling here is the enduring popularity of a rather naïve extrapolation of Shannon's law of information: the idea that we can progressively eliminate uncertainty in all situations through the addition of information (which are themselves certain), each of which would reduce uncertainty in varying amounts. Both as a technical procedure and as a social imaginary, datafication thus consists not simply of truth claims but also the normalization of a new kind of grounding for knowledge claims.

This grounding, this social basis of felt certainty, was precisely the subject of Wittgenstein's final, incomplete work. In it, he asks, "What is entailed in the simple phrase, 'I know'?" There is a curious masking function: the act of saying "I know this is a tree," for instance, does not establish any comprehensive or objectively certain proof that I really do know. Yet we trust such claims on a regular basis, tacitly agreeing not to question them too far; after all, only philosophers bother to hold regular debates revisiting whether trees really exist. Our knowledge claims provide no indisputable foundation. The very act of saying "I know" seeks to "exempt certain propositions from doubt," to agree to not to look too closely.³⁰ This infrastructure of common sense is what Wittgenstein provisionally labeled world-picture, *Weltbild*: models that allow us to cope with the world, to make certainty and judgment possible.³¹

Yet herein lies an unresolvable paradox at the heart of claims to better knowledge: the groundlessness of the ground itself, or, the ways in which the demarcation of what "counts" as good knowledge is ultimately arbitrary. Wittgenstein comments that "at the foundation of well-founded belief lies belief that is not founded"³²—precisely because to claim "I know" is an act that removes its contents from the game of proof and justification. This arbitrariness is well exposed by young children not yet versed in the unspoken boundaries of the language game: "But how do you know it is a tree?" "Well, it has branches, a trunk, some leaves." "But how do you know those are branches?" "Well, if you look at an encyclopedia—" "But how do *they* know those are branches?" and so on until the frustrated adult snaps: "We just know, okay?" We might reasonably dispute whether such grounding is truly *groundless* or simply deferred and bracketed in sufficiently complex ways that it can be presumed in ordinary contexts. For our purposes, the two options have the same consequence. For ordinary subjects, navigating their everyday life, pressed to judge and form opinions about things increasingly be-

yond their phenomenological horizon (such as a vast and secret government surveillance system or bodily physiological processes beyond the human sensorium), there is a practical need to make or accept knowledge claims, to not question their indefinite regress as the child does. Hence, Wittgenstein inverts the typical model of proof to say that when I say I know this is a tree, “it is anchored in all my *questions and answers*, so anchored that I cannot touch it.”³³ To question that it is a tree, or to question how I can know any such thing, shakes too much of the edifice built above it that it typically becomes *unreasonable* to question it.

It is this ground that is being reconfigured when counterterrorism efforts blur lines with sting operations or when self-surveillance technologies are promoted as superseding human memory, cognition, and other “natural” means of datafying the natural world and their own body.³⁴ The following chapters examine how specific and often-imperfect techniques for prediction and analysis become valorized as objectively superior knowledge. Meanwhile, a growing set of assumptions—about the nature of data, the value of human thought and machinic calculation, the knowability of the world out there and the human body as an information machine—become “set apart” and invisibilized, melding into the background of everyday experience and of public discourse on data-driven knowledge. Across both state and self-surveillance, the material objects of datafication constantly seek to sink into the background of lived experience—mirroring the disappearance of data as a social construction deep into the ground. The NSA’s data collection occurs not at the embodied sites of personal communications but through undersea fiberoptic cables, restricted-access data centers deep in the Utah desert, or buildings hidden in plain sight as a brutalist New York skyscraper.³⁵ Self-surveillance devices, at first thrust into the spotlight as delightful novelties, are increasingly seeking to recede into the realm of habit and unnoticed ubiquity—where their influence on individuals no longer needs to be justified through active and spectacular use. Datafication, in short, seeks to become our groundless ground.

The groundless ground constantly encourages those who live on it to forget how contingent it is. Pointing to the most basic elements in scientific and mathematical reasoning, Ian Hacking speaks of “styles of reasoning”: nothing even so complicated as a system of measurement or a law but something as elementary as, say, the “ordering of variety

by comparison and taxonomy.”³⁶ Like the epistemic qualities Foucault charted in *The Order of Things*, these basic tendencies rarely come up for debate even as theories and ideologies are toppled. They change far more slowly and so provide stable grounding that allow us to perceive a fact as fact in the first place.³⁷ In the data-driven society, such styles of reasoning govern how we relate a number (an algorithmically generated expression of reality) to the body’s sensory experience, to conscious human testimony, and to other sources of truth. It governs how bodies are turned into facts: what kinds of bodies become eligible for what kind of datafication and how different bodies are treated to different kinds of factmaking processes. To identify the groundless ground as ultimately arbitrary and conventional is not to say that they are therefore illegitimate; such fabrication is, once again, a normal part of the social existence of things.³⁸ What it does mean is that data’s claim to better knowledge is not a given, and neither are the forms of factmaking they bequeath on society. There are important political and moral choices to be made around what kinds of authorities should serve as the groundless ground and what kinds of data, machines, and predictions should count as looking and feeling like truth.

The Data Market

The epistemic fantasies of datafication matter—not when or if they deliver on all their promises but in the present, where the mobilization of collective belief in those fantasies transform what counts as truth and certainty. The patterns and tendencies specific to contemporary state and self-surveillance stem from two important tendencies in big data analytics: indifference and recombability. Big data analytics are predicated on the ceaseless production of data indifferent to its specific nature and without a rigid presumption of its utility—because this data will always remain open further exploratory analyses, recombining different datasets and analytical methods to discover unforeseen correlations.³⁹ This is indifference to causality in favor of correlation; indifference to “intelligence,” in the sense that the data is collected without the prior establishment of an interpretive context; and, as subsequent chapters show, indifference to the human experience of the world and that context of everyday living. To be sure, indifference does not mean neutrality.

Even as many aspects of the analytical process become automated and left up to learning machines, the design of those learning processes and the initial identification of the kind of data to be gathered renders it an “interested,” if not deliberately biased, process.⁴⁰

One such driving interest is precisely the manufacture of usable, justifiable certainty. Algorithms, as Louise Amoore puts it, “allo[w] the indeterminacies of data to become a means of learning and making decisions.”⁴¹ Messy data, extracted from lived experience and social reality and reordered into machine-readable form and modeled into a comprehensible pattern, are leveraged to produce truth claims that are not simply true or false but are carefully packaged expressions of probability that harbor uncertainty by definition. These are deployed and sold as freely transportable systems for generating “insights” across different social problems. To begin with, technologies and products are often crafted for fairly specific purposes. But that very act of measuring often involves recombining whatever data that can be conveniently acquired until a useful correlation (i.e., a profitable payload) is discovered, and it is also common that such data collection later leads to new and formerly unimagined kinds of predictions. Thus, the sex-tracking app *Spreadsheets* measures “thrusts per minute,” a largely pointless value for any human assessment of sexual intercourse but one that the movement sensors on a typical smartphone are well equipped to provide. Such sensors, originally implemented for distinct features (such as the use of accelerometer and gyroscope data to allow portrait/landscape orientations on smartphone screens), create new affordances for the business of tracking. Big data analytics often has “no clearly defined endpoints or values,”⁴² precisely because its profitability hinges on the expectation that any given algorithm, any process of datafication, might potentially be exported as a standard procedure for an indefinite range of activities (and thus business opportunities).

State and self-surveillance, despite their many local differences, thus participate in a wider, cross-contextual data market. The seemingly technical tendencies of indifference and recombination work to encourage a particular set of political and economic realities. The optimism that any and every process can be improved through datafication constitutes a voracious impulse that reveals big data’s fundamental affinity with capitalism’s search for continual growth.⁴³ The larger the userbase, the more

data to be extracted, which not only refines the primary analytics but also increases the possibility of recombining that data for new uses (or for selling them on to third-party buyers). Thus PredPol, the prominent predictive analytics system for law enforcement, borrows from existing earthquake modeling techniques.⁴⁴ Palantir, a private data analytics company, was born out of funding from In-Q-Tel, the venture capital arm of the CIA, and then sold its products back to intelligence agencies.⁴⁵ It has subsequently begun to reach out to corporate clients, such as American Express and JPMorgan Chase, demonstrating the ease with which antiterrorist technologies and antiterrorist funding can be leveraged for civilian surveillance.⁴⁶ Fitbit, one of the most popular tracking devices during the mid-2010s, is piloting partnerships with insurance companies,⁴⁷ and a significant minority of products have been reported to share data with third parties,⁴⁸ following exactly in the footsteps of social media platforms' journey to profitability.

The data market advances what has been called "surveillance capitalism": the work of making the world more compatible with data extraction for recombinant value generation.⁴⁹ This perspective situates what is promoted as a technological breakthrough in a longer historical cycle of capitalist "logics of accumulation,"⁵⁰ including the postwar military-industrial complex.⁵¹ In effect, the data market constitutes an early twenty-first-century answer to capitalism's search for new sources of surplus value. Here, new technological solutions are presented as (1) a universal optimizer, which is hoped to short-circuit existing relations of production and maximize the ratio at which labor power is converted into surplus value, and (2) itself a commodity, which may be hyped up for a new round of consumerist excitement.⁵² Surplus value is located not so much in the optimization of prices and goods sales but in the optimization of data extraction and refinement.⁵³ The "profit" at the end of this process is sometimes obviously commercial, as in targeted advertising and the direct selling of consumer goods. But the profits or uses of surveillance capitalism must also be counted in the biopolitical sense, wherein state securitization seeks to identify and manage the normal population or the individual consumer is enjoined to render themselves more attractive to algorithmic decision-making systems through techniques of self-optimization. The constant traffic and recombination of data thus entail an ever-wider range of situations in which data may

substitute or override the claims of physical bodies, conscious subjects, and lived experience.

In commercialized spheres, such as self-surveillance (and even in state surveillance, where the drive to datafy produces opportunities for lucrative government contracts for private firms), the logic of accumulation is the engine that animates datafication's promise of better knowledge. In this light, the ongoing demotion of human knowing in favor of machinic measurement and data-driven insight is not simply an intellectual argument but a variation of what David Harvey called "accumulation by dispossession": the seizure of assets to release at extremely low costs, producing new opportunities for profit that predictably benefit those with incumbent capital.⁵⁴ The more devalued human intelligence, the better for selling artificial intelligence. With datafication, the deep somatic internality of the self—my desires, my intentions, my beliefs—are opened up for revaluation on terms distinctly favorable to new products and systems of datafication. Exhorting the virtues of self-surveillance requires downgrading the reliability of human memory and cognition, such that the smart machines—and the new industries of hardware sales as well as the subsequent recombination of that data—is seen as necessary to true self-knowledge.

These trends extend long-standing tendencies in the history of surveillance, both digital and otherwise. After all, Foucauldian discipline was never about the sovereign execution of coercive power through surveillance; it was itself a highly distributed and participatory practice pegged to the promises of knowledge and productivity. To be sure, embedded in the very word *surveillance*—composed from the French *sur* (above) and *veiller* (to watch)—is a specific relation: domination from "above" through optics. But alongside that straightforward image of Big Brother is a history of surveillance as a technique for producing truth, affixing subjects to the identities and roles prescribed by that truth, and, ultimately, *disciplining* subjects into general dispositions and ways of seeing. To ponder whether we are "panoptic" or "synoptic" or "post-panoptic" is to miss the broader continuity of that liberal principle in which subjects participate in their own surveillance through the internalization of a certain way of seeing.⁵⁵ The lesson shared across the panopticon, the ominous screens of 1984, the highly visible CCTV installations in London's streets, is that what really matters is not (only)

the active relation of a watching subject and the watched one but the generalization of the condition of being *under* surveillance—a condition that corrals the human body and all it does into a standing-reserve of evidentiary material for interpretation, recombination, and classification.⁵⁶ From this vantage point, what is fundamental to surveillance is not the image of an Orwellian coercive control but a set of processes by which my truth becomes defined by those other than myself through a systematic and standardizing mode of organization. Surveillance, in this sense, is inseparable from the history of large-scale communication technologies and often develops in lockstep with the reach of the latter.

This book asks what kinds of politics, what kind of subjectivity, becomes afforded through the normalization of these technological fantasies around objectivity and purity and through the cross-contextual expansion of the data market. In the data-driven society, “what counts as knowledge” so often ends up a question of what counts as *my* body, *my* truth, *my* eligibility for social services, *my* chances of being targeted for surveillance, *my* chances at a job . . . Even as the idea of big data bloomed into a ubiquitous buzzword, its ambiguous consequences continued to break out in accidents and scandals. Some were told through the popular annals of outrageous stories: the man who was fired by algorithms,⁵⁷ the African Americans categorized as gorillas by Google Images.⁵⁸ Other controversies were more wide-ranging and enduring, such as the Snowden affair itself. It has been described as the “data wars”: the growing social conflict over how people’s algorithmic identities are determined and by whom.⁵⁹ Like the culture wars, what is at stake is the distribution of labels and associations by which we can identify, sort, and make judgments on individuals.

The trouble is that even as big data and smart machines invoke the thoroughly modern and Enlightenment imagery of technological progress and societal reform, this generalization of indifferent and recombinant factmaking often serves to retrench politics and economics as usual. The mix of naïve liberal individualism and technocracy that fuels the visions of machine-optimized futures provides no fresh political vision for the distribution of resources or the organization of collectives. There is only the conceit that with new technologies, we can finally achieve a fully automated luxury capitalism. Indeed, the very idea of “optimizing” reflects one of capitalism’s essential assumptions: that there

is always another world beyond this one to plunder, that there is no end to expansion, and that we shall not run out of resources, of new conquests, new sources of value.⁶⁰ That capitalism, just like technology, just needs the next upgrade, the next invention, to really fulfill its pure vision of totally frictionless transactions and truly melt all that is solid into air. Dressed in the shining garb of technological novelty, datafication proves most of all the difficulty of proposing a coherent alternative to capitalism and the good liberal subject.⁶¹ The push for datafication thus extends and depends on enduring fantasies around liberal values, even as its implementation often reprises old roadblocks and compromises. We now turn to one such impasse in the Snowden affair, where an unanswered question looms above all the debates around transparency and secrecy, surveillance and privacy: Can the public truly know for itself in the age of nonhuman technologies? If not, what kind of politics remains?

NOTES

INTRODUCTION

- 1 Daniel W. Smith, "Deleuze and the Question of Desire: Toward an Immanent Theory of Ethics," *Parrhesia* 2 (2007): 66–78.
- 2 For instance, the study of epistemic cultures in science examines "how we know what we know" and what kinds of "epistemic machinery" furnish the actual claims (such as equations or laws). Karin Knorr Cetina, *Epistemic Cultures—How the Sciences Make Knowledge* (Cambridge, MA: Harvard University Press, 1999). I also draw on Ian Hacking's notion of styles of reasoning later.
- 3 Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (New York: W. H. Freeman and Company, 1976).
- 4 Mary L. Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Boston: Houghton Mifflin Harcourt, 2019).
- 5 For examples in the twentieth century, see Colin Koopman, *How We Became Our Data: A Genealogy of the Information Person* (Chicago: Chicago University Press, 2019); Orit Halpern, *Beautiful Data: A History of Vision and Reason since 1945* (Durham, NC: Duke University Press, 2014); Paul Erickson, Judy L. Klein, Lorraine Daston, Rebecca Lemov, Thomas Sturm, and Michael D. Gorin, *How Reason Almost Lost Its Mind* (Chicago: University of Chicago Press, 2013).
- 6 Long used as a generic term for police techniques such as stop-and-frisk, the dragnet became a popular label for key elements of NSA electronic surveillance programs unveiled by Snowden. For instance, the NSA's upstream interception of phone call records from "backbone chokepoints" of the communications infrastructure is described as a dragnet in lawsuits brought by civil society against the NSA (*Clapper v. Amnesty International* [2013]; *Wikimedia Foundation v. NSA* [2015]). There is, of course, no universal dragnet for the complete capture of all communications data. Rather, the dragnet expresses the relative jump in the size and breadth of the NSA's numerous collection programs and the many ways in which the data captured necessarily precedes and exceeds the identification of specific targets. Also see Benjamin Wittes, "The Problem at the Heart of the NSA Disputes: Legal Density," *Lawfare*, February 14, 2014, <https://www.lawfareblog.com/problem-heart-nsa-disputes-legal-density>.
- 7 Bruce Neuman, "When a Senator and the C.I.A. Clash," *The New York Times*, March 12, 2014, <https://www.nytimes.com/2014/03/13/opinion/when-a-senator-and-the-cia-clash.html>.

- 8 Respectively, Tom Engelhardt, *Shadow Government: Surveillance, Secret Wars, and a Global Security State in a Single-Superpower World* (Chicago: Haymarket Books, 2014); Julia Angwin, *Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance* (New York: Times Books, 2014); Greenwald, *No Place to Hide: Edward Snowden, the NSA and the Surveillance State* (London: Penguin Books, 2014).
- 9 Ariel Garten, “Know Thyself, with a Brain Scanner,” filmed at TEDxToronto, Toronto, Canada, September 2011, video, 14:49, www.ted.com.
- 10 This included the full population of media coverage on self-surveillance and the QS from a selection of major publications—*The Atlantic*, *Fast Company*, *Harvard Business Review*, *Inc.*, *National Review*, *The New York Times*, *The Washington Post*, and *Wired*—between 2007 and 2015. These were supplemented by coverage from other news sources; the QS official website archives; promotional material on commercial self-surveillance products, from product websites to crowdfunding campaigns to scientific research papers published by developers.
- 11 This approach draws on Funtowicz and Ravetz’s analysis of what kinds of uncertainties are integrated into or excised from the normative bandwidth of scientific inquiry. See Silvio O. Funtowicz and Jerome R. Ravetz, “The Emergence of Post-Normal Science,” in *Science, Politics and Morality: Scientific Uncertainty and Decision Making*, ed. Rene von Schomberg (Springer-Science+Business Media, B.V., 1993), 85–123.
- 12 For a systematic description of big data in these terms, see Rob Kitchin, “Big Data, New Epistemologies and Paradigm Shifts,” *Big Data & Society* 1, no. 1 (2014), <https://www.theoryculturesociety.org/kittler-on-the-nsa/>; Rob Kitchin and Gavin McArdle, “What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets,” *Big Data & Society* 3, no. 1 (2016): 1–10.
- 13 Mei-po Kwan, “Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge,” *Annals of the American Association of Geographers* 106, no. 2 (2016): 274–282.
- 14 Consider, for instance, Steven Shapin’s history of Mertonian moral equivalence and other reshaping of scientific subjectivity vis-à-vis the increasing institutionalization of “Big Science.” Steven Shapin, *The Scientific Life: A Moral History of a Late Modern Vocation* (Chicago: University of Chicago Press, 2008).
- 15 Michel Foucault, *The Order of Things: An Archaeology of the Human Sciences* (London: Routledge, 2002).

CHAPTER 1. HONEYMOON OBJECTIVITY

- 1 “Meet Sense,” Hello.is, accessed April 13, 2016, <https://hello.is/videos#meet-sense>.
- 2 James Proud, “Goodbye, Hello,” *Medium*, June 12, 2017, <https://medium.com/@hello/goodbye-hello-c62eaf58d13>.
- 3 Fahrad Manjoo, “Mysteries of Sleep Lie Unsolved,” *The New York Times*, February 24, 2015, http://www.nytimes.com/2015/02/26/technology/personaltech/despite-the-promise-of-technology-the-mysteries-of-sleep-lie-unsolved.html?_r=0.

- 4 Genevieve Bell and Paul Dourish, "Yesterday's Tomorrows: Notes on Ubiquitous Computing's Dominant Vision," *Personal and Ubiquitous Computing* 11, no. 2 (2007): 133–143.
- 5 David Haskin, "Don't Believe the Hype: The 21 Biggest Technology Flops," *Computerworld*, April 4, 2007, <https://www.computerworld.com/article/2543763/computer-hardware/don-t-believe-the-hype--the-21-biggest-technology-flops.html>.
- 6 Jackie Fenn and Hung LeHong, "Hype Cycle for Emerging Technologies, 2011," Gartner, July 28, 2011, <https://www.gartner.com/doc/1754719/hype-cycle-emerging-technologies>.
- 7 For example, John M. Jakicic, Kelliann K. Davis, and Renee J. Rogers, "Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-Term Weight Loss," *JAMA: The Journal of the American Medicine Association* 316, no. 11 (2016): 1161–1171.
- 8 Eric Kluitenberg, "On the Archaeology of Imaginary Media," in *Media Archaeology: Approaches, Applications, and Implications*, ed. Erkki Huhtamo and Jussi Parikka (Berkeley: University of California Press, 2011), 48–69.
- 9 A point well made in anthropological investigations of ritual. See Roy Rappaport, *Ritual and Religion in the Making of Humanity* (Cambridge: Cambridge University Press, 1999); Victor Turner, "Liminal to Liminoid," in *Play, Flow, Ritual: An Essay in Comparative Symbolology*, ed. Victor Turner (New York: Performing Arts Journal Publishing, 1982), 53–92.
- 10 Benjamin J. Muller, "Securing the Political Imagination: Popular Culture, the Security Dispositif and the Biometric State," *Security Dialogue* 39, no. 2–3 (2008): 211.
- 11 Suzanne L. Thomas, Dawn Nafus, and Jamie Sherman, "Algorithms as Fetish: Faith and Possibility in Algorithmic Work," *Big Data & Society* 5, no. 1 (2018).
- 12 See the definition of fantasy given in Lauren Berlant, *Cruel Optimism* (Durham, NC: Duke University Press, 2011), 2.
- 13 Here I am drawing on existing arguments by historians of science and technology regarding the historical emergence of technology as word and idea, its expansion from specific crafts and solutions into a broader and autonomous force, the gradual prioritization of progress as for its own sake, an originary link between the focus on "rational manipulation of the environment" and the penchant for technological solutions, and America's historical depiction of technology through, paradoxically, often aesthetic and fantastical imagery of rational mastery. For examples, see Leo Marx, "Technology: The Emergence of a Hazardous Concept," *Technology and Culture* 51, no. 3 (2010): 561–577; Langdon Winner, *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought* (Cambridge, MA: MIT Press, 1977); Leo Marx, "The Idea of 'Technology' and Postmodern Pessimism," in *Does Technology Drive History? The Dilemma of Technological Determinism*, ed. Merritt Roe Smith and Leo Marx (Cambridge, MA: MIT Press, 1994),

- 237–258; David Landes, *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present* (Cambridge: Cambridge University Press, 1969); David E. Nye, *American Technological Sublime* (Cambridge, MA: MIT Press, 1994).
- 14 Lorraine J. Daston and Peter Galison, *Objectivity* (New York: Zone Books, 2007).
 - 15 Also see Lorraine J. Daston, “Objectivity and the Escape from Perspective,” *Social Studies of Science* 22 (1992): 597–618.
 - 16 Daston and Galison, *Objectivity*, 123.
 - 17 Lorraine Daston, “The Moral Economy of Science,” *Osiris* 10 (1995): 3–24.
 - 18 Eric M. Eisenberg, “Ambiguity as Strategy in Organisational Communication,” *Communication Monographs* 51, no. 3 (1984): 227–242.
 - 19 Also see Claudia Aradau and Tobias Blanke, “The (Big) Data-Security Assemblage: Knowledge and Critique,” *Big Data & Society* 2, no. 2 (2015): 1–12.
 - 20 Catherine Gale, prod. and dir., *The Joy of Data*, aired July 20, 2016, on BBC Four.
 - 21 See Martin Frické, “The Knowledge Pyramid: A Critique of the DIKW Hierarchy,” *Journal of Information Science* 35, no. 2 (2009): 131–142.
 - 22 Ernesto Ramirez, “QS Access: Personal Data Freedom,” *Quantified Self*, February 11, 2015, <http://quantifiedself.com/2015/02/qs-access-personal-data-freedom/>.
 - 23 Sara M. Watson, “You Are Your Data,” *Slate*, November 12, 2013, www.slate.com/articles/technology/future_tense/2013/11/quantified_self_self_tracking_data_we_need_a_right_to_use_it.html; Gary Wolf, “A Public Infrastructure for Data Access,” *Quantified Self*, March 8, 2016, <http://quantifiedself.com/2016/03/larry-smarr-interview/>.
 - 24 Lisa Gitelman and Virginia Jackson, “Introduction,” In *Raw Data Is an Oxymoron*, ed. Lisa Gitelman (Cambridge, MA: MIT Press, 2013), 1–14.
 - 25 The degree and type of human intervention depend on the actual methods involved, of course. It is one thing to manually specify categories in a database and code for automatic sorting of new entries, and it is quite another to train artificial neural nets to develop their own categories as they look for the optimal structure to arrive at desired outcomes (which are, in this case, set manually by human designers). But any specific automated process is part of a larger process in which human coding, selection, categorization, and cleaning of data often plays an important role. Alternatively, could data be “raw” in the sense that it represents the independent objective truth of certain “elementary” natural facts—for example, the weight of an apple. Without delving into the ontological debates, it suffices to point out that the machinic apparatus and its ability to recognize “weight” is, from their inception, designed by humans in terms of what such phenomena mean for us.
 - 26 Josh Bersen, *Computable Bodies: Instrumented Life and the Human Somatic Niche* (London: Bloomsbury Academic, 2015).
 - 27 For example, Kevin D. Haggerty and Richard V. Ericson, “The Surveillant Assemblage,” *British Journal of Sociology* 51, no. 4 (2000): 605–622; John Cheney-Lippold, *We Are Data—Algorithms and the Making of Our Digital Selves* (New

- York: New York University Press, 2017); Sara M. Watson, “Data Doppelgängers and the Uncanny Valley of Personalization,” *Atlantic*, June 16, 2014, <https://www.theatlantic.com/technology/archive/2014/06/data-doppelgangers-and-the-uncanny-valley-of-personalization/372780/>; Bernard Harcourt, *Exposed: Desire and Disobedience in the Digital Age* (Cambridge, MA: Harvard University Press, 2015).
- 28 Daniel Rosenberg, “Data before the Fact,” in *Raw Data Is an Oxymoron*, ed. Lisa Gitelman (Cambridge, MA: MIT Press, 2013), 15–40.
 - 29 Respectively, Ed Finn, *What Algorithms Want: Imagination in the Age of Computing* (Cambridge, MA: MIT Press, 2017); N. Katherine Hayles, *My Mother Was A Computer: Digital Subjects and Literary Texts* (Chicago: University of Chicago Press, 1999); David Golumbia, *The Cultural Logic of Computation* (Cambridge, MA: Harvard University Press, 2009).
 - 30 Ludwig Wittgenstein, *On Certainty*, ed. G. E. M. Anscombe and G. H. von Wright (New York: Harper Torchbooks, 1969), sec. 88.
 - 31 See Schulte, “World-Picture and Mythology,” *Inquiry: An Interdisciplinary Journal of Philosophy* 31 (1988): 323–334.
 - 32 Wittgenstein, *On Certainty*, sec. 253.
 - 33 Wittgenstein, *On Certainty*, sec. 103.
 - 34 These regimes, it should be clear, are often maintained not through explicit propositions, as in a code of law or in a textbook on the scientific method; such expressions must themselves be sketched by looking for the shared assumptions and ways of making claims underlying more solid monuments of the regime. It is, Wittgenstein says, a *form of life*. See Wittgenstein, *On Certainty*, sec. 358.
 - 35 This was “Project X” or the “Long Lines Building,” a concrete-and-granite skyscraper in Manhattan that reveals nothing about its identity to the city around it. Snowden-leaked files strongly suggest that it was used as an NSA surveillance site for housing and tapping into phone call routing systems. See Ryan Gallagher and Henrik Moltke, “Titanpointe: The NSA’s Spy Hub in New York, Hidden in Plain Sight,” *The Intercept*, November 16, 2016, <https://theintercept.com/2016/11/16/the-nsas-spy-hub-in-new-york-hidden-in-plain-sight/>.
 - 36 Ian Hacking, “Statistical Language, Statistical Truth and Statistical Reason: The Self-Authentication of a Style of Scientific Reasoning,” in *Social Dimensions of Science*, ed. Ernan McMullin (Notre Dame, IN: University of Notre Dame Press, 1992), 130–157.
 - 37 David Rabouin, “Styles in Mathematical Practice,” in *Cultures Without Culturalism: The Making of Scientific Knowledge*, ed. Karine Chemla and Evelyn Fox Keller, 196–223 (Durham, NC: Duke University Press, 2017), 207.
 - 38 A point also made frequently in science studies. See Bruno Latour, *Pandora’s Hope: Essays in the Reality of Science Studies* (Cambridge, MA: Harvard University Press, 1999).
 - 39 See, for example, Michael Betancourt, “The Demands of Agnotology::Surveillance,” *Ctheory*, 2014, http://ctheory.net/ctheory_wp/the-

- demands-of-agnotologysurveillance; Shoshana Zuboff, “Big Other: Surveillance Capitalism and the Prospects of an Information Civilization,” *Journal of Information Technology* 30, no. 1 (2015): 75–89.
- 40 Bernhard Rieder, “Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality,” *Information, Communication & Society* 20, no. 1 (2017): 100–117.
- 41 Louise Amoore, “Doubt and the Algorithm: On the Partial Accounts of Machine Learning,” *Theory, Culture & Society* 28, no. 6 (2011): 24–43.
- 42 Orit Halpern, *Beautiful Data: A History of Vision and Reason since 1945* (Durham, NC: Duke University Press, 2014), 26.
- 43 Jim Thatcher, David O’Sullivan, and Dillon Mahmoudi, “Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data,” *Environment and Planning D: Society and Space* 34, no. 6 (2016): 991. Also see Nick Srnicek, *Platform Capitalism* (Cambridge: Polity Press, 2017).
- 44 Mark Andrejevic, “To Preempt a Thief,” *International Journal of Communication* 11 (2017): 890.
- 45 These public–private partnerships are animated not only by the commercial logic but also by the state’s interests in securitization. The US government, frequently an open admirer of the Valley during much of this period, repeatedly sought its technologies; In-Q-Tel also funded a mapping company that Google would purchase to build Google Earth. Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019), 117.
- 46 Peter Waldman, Lizette Chapman, and Jordan Robertson, “Palantir Knows Everything About You,” *Bloomberg Businessweek*, April 19, 2018, <https://www.bloomberg.com/features/2018-palantir-peter-thiel/>.
- 47 Erika Pearson, “Smart Objects, Quantified Selves, and a Sideways Flow of Data” (presentation at *ICA 2016*, Fukuoka, Japan, 2016).
- 48 For example, Linda Ackerman, “Mobile Health and Fitness Applications and Information Privacy,” 2013 (San Diego, CA: Privacy Rights Clearinghouse); Kate Kaye, “FTC: Fitness Apps Can Help You Shred Calories—and Privacy,” *AdAge*, May 7, 2014, <http://adage.com/article/privacy-and-regulation/ftc-signals-focus-health-fitness-data-privacy/293080/>.
- 49 For example, Zuboff, *The Age of Surveillance Capitalism*.
- 50 Zuboff, “Big Other,” 75.
- 51 John Bellamy Foster and Robert W. McChesney, “Surveillance Capitalism: Monopoly-Finance Capital, the Military-Industrial Complex, and the Digital Age,” *Monthly Review* 66, no. 3 (2014), <https://monthlyreview.org/2014/07/01/surveillance-capitalism/>.
- 52 See David Harvey, *Marx, Capital, and the Madness of Economic Reason* (New York: Oxford University Press, 2018).
- 53 Srnicek, *Platform Capitalism*.

- 54 David Harvey, *The New Imperialism* (Oxford: Oxford University Press, 2003). Elsewhere, he terms it a madness, a Hegelian bad infinity: Harvey, *Marx, Capital, and the Madness of Economic Reason*, 172–173. The idea has been taken up more commonly with respect to Big Tech data brokers such as Google and Facebook. For more, see Thatcher, O’Sullivan, and Mahmoudi, “Data Colonialism through Accumulation by Dispossession”; Zuboff, *The Age of Surveillance Capitalism*, 99.
- 55 For example, see Zygmunt Bauman and David Lyon, *Liquid Surveillance: A Conversation* (Cambridge: Polity, 2013), 51–52.
- 56 Rosen and Santesso’s historical study also emerges with a definition of surveillance not tied to centralizing, coercive power, but “the monitoring of human activities for the purposes of anticipating or influencing future events.” David Rosen and Aaron Santesso, *The Watchman in Pieces—Surveillance, Literature, and Liberal Personhood* (New Haven, CT: Yale University Press, 2013), 10.
- 57 Namely, Ibrahim Diallo, a contractor who was fired by an errant data input that set in motion a set of termination procedures that neither his manager nor departmental director could stop.
- 58 For more on this subject, see Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018), 7.
- 59 Cheney-Lippold, *We Are Data*, 19.
- 60 An anxiety that has been with us since at least the later nineteenth century. See Rosalind H. Williams, *The Triumph of Human Empire: Verne, Morris, and Stevenson at the End of the World* (Chicago: University of Chicago Press, 2013).
- 61 See Mark Fisher, *Capitalist Realism: Is There No Alternative?* (Winchester, UK: O Books, 2009).

CHAPTER 2. THE INDEFINITE ARCHIVE

- 1 *Last Week Tonight with John Oliver*, featuring John Oliver, Tim Carvell, James Taylor, and Jon Thoday, aired April 5, 2015, on HBO.
- 2 Timothy Morton, *Hyperobjects: Philosophy and Ecology after the End of the World* (Minneapolis: University of Minnesota Press, 2013).
- 3 The irony being, of course, that Kant’s valorization of “public use of reason” was restricted to actions outside each subject’s “private,” professional activity, in which the overriding virtue was to obey. One suspects Snowden’s actions would not have passed this test. See Immanuel Kant, “An Answer to the Question: What Is Enlightenment?” in *What Is Enlightenment? Eighteenth-Century Answers and Twentieth-Century Questions*, ed. James Schmidt, translated by James Schimdt (Berkeley: University of California Press, 1996), 58–77.
- 4 Kant, *What Is Enlightenment?*, 63. He was, of course, referring to Julien Offray La Mettrie’s *L’Homme machine*, then and now widely seen as a culmination of a radical materialism in the Enlightenment. La Mettrie’s invocation of machine extended “a general trend . . . towards rationalisation and secularisation” in the