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 structures[1]—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 parameters. 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," emblematized 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, improvised 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 explicity 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 otherwise 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 thing[15]—is flickering uncertainly, unsure of the agency and moral responsibility we had worked so hard to attach to it.

1

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" object[3] 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