INTRODUCTION

>

We are "well filled with data" in today's networked society.[1] If you don't believe me, open your computer and roam the web for five minutes. In a period of time only slightly longer than the average television commercial break, you will have generated, through your web activity, an identity that is likely separate from the person who you thought you were. In a database far, far away, you have been assigned a gender, ethnicity, class, age, education level, and potentially the status of parent with *x* number of children. Maybe you were labeled a U.S. citizen or a foreigner. There's even a slight chance you were identified as a terrorist by the U.S. National Security Agency.

This situation is simultaneously scary and intriguing. It's scary because of the very real power that such classifications hold: having a SIM card match the data signature of a suspected terrorist can put someone at the receiving end of a drone missile strike. Having Internet metadata that identifies a user as a foreigner means she may lose the right to privacy normally afforded to U.S. citizens. And it's intriguing because there's something gallingly, almost comically presumptuous about such categorizations. Who would have thought class status could be algorithmically understood? How can something as precise as citizenship be allocated without displaying one's passport? And how audacious is it to suggest that something even less precise, like ethnicity, could be authoritatively assigned without someone having the decency to ask?

We live in a world of ubiquitous networked communication, a world where the technologies that constitute the Internet are so woven into the fabrics of our daily lives that, for most of us, existence without them

seems unimaginable.[2] We also live in a world of ubiquitous surveillance, a world where these same technologies have helped spawn an impressive network of governmental, commercial, and unaffiliated infrastructures of mass observation and control.[3]

Today, most of what we do in this world has at least the *capacity* to be observed, recorded, analyzed, and stored in a databank. As software developer Maciej Ceglowski explains, "The proximate reasons for the culture of total surveillance is clear. Storage is cheap enough that we can keep everything. Computers are fast enough to examine this information, both in real time and retrospectively. Our daily activities are mediated with software that can easily be configured to record and report everything it sees upstream."[4] A simple web search from even the most unsophisticated of smart phones generates a lengthy record of new data. This includes your initial search term, the location of your phone, the time and day when you searched, what terms you searched for before/after, your phone's operating system, your phone's IP address, and even what apps you installed on your phone. Add onto this list everything else you do with that phone, everything else you do on your computer, and everything else that might be recorded about your life by surveilling agents.

This resulting aggregation of our lives' data founds the discursive terrain of our digital environments. We live in what legal scholar Frank Pasquale has termed a "black box society," where algorithms determine the contours of our world without us knowing. Within this society, "a predictive analytics firm may score someone as a 'high cost' or 'unreliable' worker, yet never tell her about the decision."[5] What "high cost" and "unreliable" mean is up to the algorithms' authors. It's an output that we feel only as we wait for a job interview that will never come. In the case of identity online, it is this categorical output that speaks for you—not you, yourself.

Indeed, you are rarely "you" online. "We are data" is not a claim that we, individually, are data. Rather, we are temporary members of different emergent categories, like "high cost" or "celebrity," from this book's

preface, according to our data. The future of identity online is how we negotiate this emergence. Accordingly, the arguments in this book deliberately attend to the level of the category itself, not the "you" of the user.

Through various modes of algorithmic processing, our data is assigned categorical meaning without our direct participation, knowledge, or often acquiescence. As Pasquale puts it, "the values and prerogatives that the encoded rules enact are hidden within black boxes."[6] Which is to say that our social identities, when algorithmically understood, are really not social at all. From behind the black box, they remain private and proprietary. Yet when employed in marketing, political campaigns, and even NSA data analytics, their discursive contents realign our present and futures online.

Who we are in this world is much more than a straightforward declaration of self-identification or intended performance. Who we are, following Internet researcher Greg Elmer's work on "profiling machines," is also a declaration by our data as interpreted by algorithms.[7] We are ourselves, plus layers upon additional layers of what I have previously referred to as algorithmic identities.[8]

Algorithmic interpretations like Google's "celebrity" identify us in the exclusive vernacular of whoever is doing the identifying. For the purposes of my analysis, these algorithmic categorizations adhere to what philosopher Antoinette Rouvroy calls "algorithmic governmentality"—a logic that "simply ignores the embodied individuals it affects and has as its sole 'subject' a 'statistical body'. . . . In such a governmental context, the subjective singularities of individuals, their personal psychological motivations or intentions do not matter."[9] Who we are in the face of algorithmic interpretation is who we are computationally calculated to be. And like being an algorithmic celebrity and/or unreliable, when our embodied individualities get ignored, we increasingly lose control not just over life but over how life itself is defined.

This loss is compounded by the fact that our online selves, to borrow the overused phraseology of pop psychology, is a schizophrenic

phenomenon. We are likely made a thousand times over in the course of just one day. Who we are is composed of an almost innumerable collection of interpretive layers, of hundreds of different companies and agencies identifying us in thousands of competing ways. At this very moment, Google may algorithmically think I'm male, whereas digital advertising company Quantcast could say I'm female, and web-analytic firm Alexa might be unsure. Who is right? Well, nobody really.

Stable, singular truth of identity, also known as authenticity, is truly a relic of the past. Our contemporary conception of authenticity, as argued by feminist scholar Sarah Banet-Weiser, has become malleable, even ambivalent. What used to be sold to us as "authentic," like the marketed promise of a corporate brand, is now read as polysemic multiplicity.[10] Google's, Quantcast's, and Alexa's interpretations of my data are necessarily contradictory because they each speak about me from their own, proprietary scripts. Each is ambivalent about who I am, interpreting me according to their individual algorithmic logics.

But in the algorithmic identifications of our gender, unreliability, or celebrity status, we are given little recourse. We most often have no way to say "no!" or "yes, but . . ." Nor can we really know who we are online, as our algorithmic identities change by the input: minute by minute and byte by byte.

In other words, online you are not who you think you are. Indeed, one of the key consequences of our algorithmic identities is how they recast the politics around identity into the exclusive, private parlance of capital or state power. If you have a Google account, go into your "Settings for Google Ads" to see what Google infers your age and gender to be (www.google.com/ads/preferences). These gender and age formulations are not based on your voluntary identification, physical performance, or amount of times you have revolved around the sun. Instead, Google's assignments of your gender and age come from the collection of web pages you have visited over the course of your Google career.

And whether you recognize it or not, these identifications affect our lives. Your search results and advertisements will be subsequently gen-

dered and aged. Websites will take the fact that you went to a certain site as evidence of your identity as, say, a middle-aged man. And online news editors may then see your visit as proof that the site's new campaign aimed at middle-aged-men-centered content is succeeding. The different layers of who we are online, and what who we are means, is decided for us by advertisers, marketers, and governments. And all these categorical identities are functionally unconcerned with what, given your own history and sense of self, makes you *you*.

Theorists Geoffrey Bowker and Susan Leigh Star write that "classification systems are often sites of political and social struggles, but these sites are difficult to approach. Politically and socially charged agendas are often first presented as purely technical and they are difficult even to see."[11] The process of classification itself is a demarcation of power, an organization of knowledge and life that frames the conditions of possibilities of those who are classified. When Google calls you a man or celebrity, this is not an empty, insignificant assessment. It is a structuring of the world on terms favorable to the classifier, be it as a member of a market segment for profit or as a categorized public figure to avoid the legal morass of European privacy law.

We witness this favorable structuring in legal scholar C. Edwin Baker's concept of "corruption" that "occurs when segmentation reflects the steering mechanisms of bureaucratic power or money rather than the group's needs and values."[12] Consider, for example, how your own gender identity interfaces with the complexities of your lived experience. When Google analyzes your browsing data and assigns you to one of two distinct gender categories (only "male" or "female"), your algorithmic gender may well contradict your own identity, needs, and values. Google's gender is a gender of profitable convenience. It's a category for marketing that cares little whether you really are a certain gender, so long as you surf/purchase/act like that gender.

Google's category, moreover, speaks with a univocality that flattens out the nuance and variety of the lived experience of gender. And that corrupt category, says Baker, "undermines both common discourse

and self-governing group life."[13] More comprehensively, an algorithmic gender's corrupt univocality substitutes for the reflexive interplay implicit in gender's social constructionism.

As an example, I could identify and perform as a man, while Google thinks I'm a woman (this is true). Or I could be in my thirties, possess a driver's license to prove it, but Google could think I'm sixty-five (this is also true). In these examples, I am not merely listing instances of misrecognition or error on the part of an algorithm. Machines are, and have been, "wrong" a lot of the time—often more than the techno-enthusiasts among us would like to admit. Even the most expensive computer software can crash, and biometric technologies routinely fail despite their touted infallibility. The point of this example, rather, is to highlight the epistemological and ontological division between my gender and how Google defines and operationalizes my algorithmic gender.

And precisely because Google's gender is Google's, not mine, I am unable to offer a critique of that gender, nor can I practice what we might refer to as a first-order gendered politics that queries what Google's gender means, how it distributes resources, and how it comes to define our algorithmic identities. Here I offer an alternative to political economist Oscar Gandy's claim that "because identity is formed through direct and mediated interaction with others, individuals are never free to develop precisely as they would wish."[14] When identity is formed without our conscious interaction with others, we are never free to develop—nor do we know how to develop. What an algorithmic gender signifies is something largely illegible to us, although it remains increasingly efficacious for those who are using our data to market, surveil, or control us.

Of course, interpretations of data have always mediated identity, whether it be through the applications of census records, econometrics, and even IQ test results. From philosopher Ian Hacking's work on statistics "making up people" to the cyberculture studies of media theorist Mark Poster's "database discourses," the nominal underpinning of "we are data" is hardly an unprecedented phenomenon.[15]

What *is* new about the categories that constitute us online is that they are unknown, often proprietary, and ultimately—as we'll later see—modulatory: Google's own interpretation of gender, faithful to nothing but patterns of data, can be dynamically redefined according to the latest gendered data.

These categories also operate at—and generate—different temporalities. As a general rule, Gandy reminds us that "the use of predictive models based on historical data is inherently conservative. Their use tends to reproduce and reinforce assessments and decisions made in the past."[16] This type of categorization delimits possibility. It shuts down potential difference for the sake of these historical models. It constructs what digital media theorist Wendy Hui Kyong Chun has called "programmed visions" that "extrapolate the future—or, more precisely, a future—based on the past."[17]

However, the myriad flows of ubiquitous surveillance reorient these visions. For companies like Google, algorithms extrapolate not just a future but a present based on the present: of near real-time search queries, web browsing, GPS location, and metadata records.[18] This change in temporality reframes the conservatism of categorical knowledge to something more versatile, similar to what geographer Louise Amoore's calls a "data derivative"—"a specific form of abstraction that distinctively correlates more conventional state collection of data with emergent and unfolding futures."[19]

When algorithms process near real-time data, they produce dynamic, pattern-based abstractions that become the new, actionable indices for identity itself. These abstractions may be invested not in extrapolating a certain future or enacting a singular norm but in producing the most efficacious categorical identity according to available data and algorithmic prowess.

Correspondingly, in an online world of endlessly overhauled algorithmic knowledge, Google's misrecognition of my gender and age isn't an error. It's a reconfiguration, a freshly minted algorithmic truth that cares little about being authentic but cares a lot about being an effective

metric for classification. In this world, there is no fidelity to notions of our individual history and self-assessment.

As philosopher Alexander Galloway observes about the video game *Civilization III*, "the modeling of history in computer code . . . can only ever be a reductive exercise of capture and transcoding," whereas "'history' . . . is precisely the opposite of history . . . because the diachronic details of lived life are replaced by the synchronic homogeneity of code pure and simple."[20] The complexity of our individual histories cannot be losslessly translated into a neat, digital format. Likewise, our self-assessments come from layers upon layers of subjective valuations, all of which are utterly unintelligible as ones and zeros.

In this algorithmic reality, there is instead a dependency on something else, a data-based model of what it means to be 'famous,' 'not famous,' 'man,' 'woman,' 'gay,' 'straight,' 'old,' 'young,' 'African American,' 'Hispanic,' 'Caucasian,' 'Asian,' 'other,' 'Democrat,' 'Republican,' 'citizen,' 'foreigner,' 'terrorist,' or 'college educated.' I offset all of these algorithmically produced categories with an unattractive use of quotation marks precisely because they are not what they say they are. Like the sardonic use of air quotes to emphasize an ironic untruth, each quotation-marked classification is an algorithmic caricature of the category it purportedly represents. These algorithmic caricatures, or what I call *measurable types*, have their own histories, logics, and rationales. But these histories, logics, and rationales are necessarily different from our own. Google's 'gender' is not immediately about gender as a regime of power but about 'gender' as a marketing category of commercial expedience.

Crucially, algorithmic categories do not substitute for their non-quotation-marked peers but rather function—sometimes in concert, sometimes in tension—with them as an additional layer of identity. I might be a man, but I am *also* a 'woman.' In my day-to-day life, I might be a boring professor. But if Google determines I'm a 'celebrity,' I lose the right to be forgotten. In this layered approach, I must attend to both the offline and the online, as both have impact on my life, and both

bleed into each other. This collapse subsequently disallows any clean conceptual separation between life as data and life as life.

And given this layered interaction, our algorithmic identities will certainly impact us in ways that vary according to our own social locations "offline," which are likewise determined by classifications that are often not of one's choosing but which operate according to different dynamics from their algorithmic counterparts. Similarly, to the extent that one's online and offline identities align or misalign, the effects of this interface will vary according to the relative status assigned to each category and one's own power to contest or capitalize on it.

All of which is to say, companies like Google use their algorithms and our data to produce a dynamic world of knowledge that has, and will continue to have, extraordinary power over our present and futures. And as we also continue to be well filled with data, this algorithmic logic produces not just the world but us. *We Are Data* aims to extend the scholastic lineage that connects the social construction of knowledge with the layers upon layers of technical, quotation-marked constructions of knowledge—a union where essential truths do not and cannot exist.

ON DATA'S TERMS

Algorithmic agents make us and make the knowledges that compose us, but they do so on their own terms. And one of the primary terms of an algorithm is that everything is represented as data. When we are made of data, we are not ourselves in terms of atoms. Rather, we are who we are *in terms of data*. This is digitization, the term that MIT Media Lab founder Nicholas Negroponte employs to talk about the material conversion of "atoms to bits."[21] It is also biomedia, the "informatic recontextualization of biological components and processes" of philosopher Eugene Thacker.[22] And it is ultimately datafication: the transformation of part, if not most, of our lives into computable data.[23]

Importantly, the "we" of "we are data" is not a uniform totality but is marked by an array of both privileging and marginalizing difference.

As digital theorist Tyler Reigeluth reminds us, we need to see digital technology "in continuity with 'previous' or existing social, political and economic structures, and not only in terms of change, revolution or novelty."[24] And as *all* data is burdened by this structural baggage, any interpretive classification of datafied life necessarily orders and organizes the world in the shadows of those structures' effects.

It is significant to note that these shadows have, for centuries, proliferated across the datafied world. From the state's use of DNA to support claims of "authentic" racial character (to reference the work of critical scholars like Kim Tallbear and Alondra Nelson) to now-debunked histories of phrenology, hegemonic forms of empiricism have long buttressed the "corrupted" identification of race to make sense of, and thus validate, dominant classifications of identity.[25] More recently, surveillance theorist Simone Browne's concept of "digital epidermalization" reminds us that algorithmic interpretations of race corrupt as well: "the exercise of power cast by the disembodied gaze of certain surveillance technologies . . . can be employed to do the work of alienating the subject by producing a truth about the racial body and one's identity (or identities) despite the subject's claims."[26]

Indeed, all algorithmic interpretations produce their own corrupted truths—not just about race—in ways particular to their technological capacity and programmed direction. Digital media scholar Lev Manovich defines this as "transcoding," or how cultural concepts, when brought onto the data/algorithm ontology of the computer, must follow the "established conventions of the computer's organization of data."[27] To perceive the world on data's terms is to recognize how "the logic of a computer can be expected to significantly influence the traditional cultural logic of media."[28] We are not simply well filled of data but made of data that is interpreted, conferred truth, and disseminated for motives of profit, organization, and/or control. The resulting classifications become the discursive terrain from which we, and others, compose our digital selves.

Consider two seemingly separate, but algorithmically connected, examples. If I smile, a computer doesn't see a cheerful man like a human

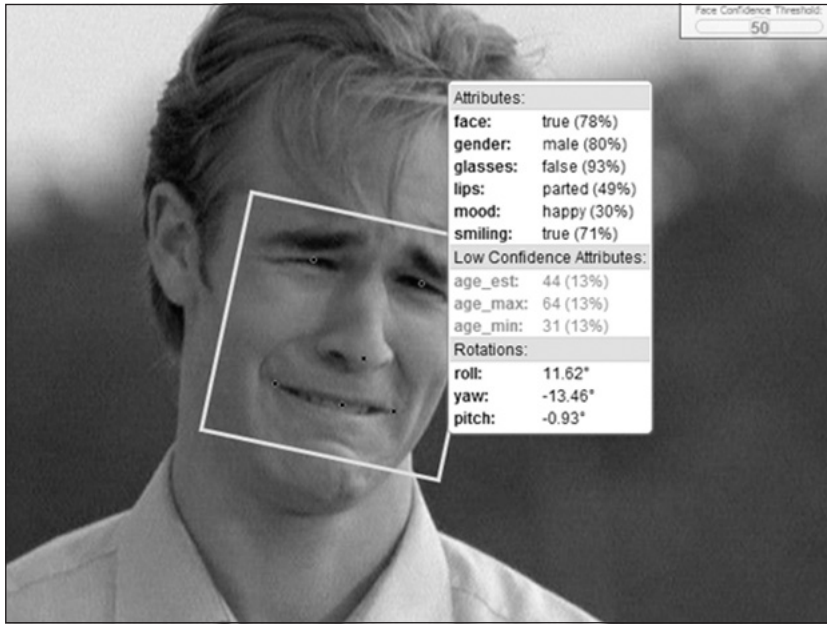

FIGURE I.1. Face.com's software analyzed photos according to a predefined set of attributes: 'face,' 'gender,' 'glasses,' 'lips,' 'mood,' 'smiling,' and 'age.' Source: Kenneth Butler, "New Face.com Tool Can Guess Your Age, Determine Gender and Mood [Hands-On]," *Laptopmag*, April 2, 2012, http://blog.laptopmag.com.

would. A computer can "see" my smile only upon interpreting the discrete pixels forming the shapes believed to be my 'eyes.'[29] It then crops, rotates, and scales the image to produce a standardized 'face.' Next, it encodes that modified image, using several algorithmic filters according to prototypical 'smiling' photos, into a single number that evaluates me as either 'smiling' or 'not smiling.'[30] Voilà, a data-'smile.'

In July 2011, this type of technology is exactly what the facial-recognition company Face.com unveiled with its innovative facial-analytics software that detects "moods and key facial expressions" in digital photographs.[31] Not limited to 'smiling,' Face.com claimed that its algorithms would be able to evaluate a 'face's' 'mood,' from 'surprised' to 'angry,' from 'sad' to 'neutral' (figure I.1). Even 'kissing' is available as

a facial category, where a 'face' can have 'lips' that are 'parted,' 'kissing,' or 'sealed.'

All of these descriptive states are accompanied by an associated confidence percentage, the statistical certainty that the 'face' the machine sees is, in fact, interpreting the correct 'emotion.' Our data-'smiles' are only smiling if Face.com's algorithms statistically say so. Quite unsurprisingly, Face.com's recognition technology was purchased in 2012 for an estimated $80 million by the owner of the largest repository of photographs on the planet, Facebook.[32]

Emotions have become datafied. But unlike the computational wizardry of turning "atoms into bits," our emotions are not understood only as atoms. Like identity, they're a complex, culturally situated, and materially incoherent experience that technically overwhelms that which is computable. Defining 'emotion' via algorithm is an interpretation of the world that ushers us into a distinct suite of knowledge, one that orders and understands the ineffable chaos of our world as operational bits of data.[33]

But as noted earlier, when we are made of data, we are not just represented but also regulated by data. And not all data is treated alike. Face.com's algorithms, for example, are not infallible truth tellers. Like latent spirits, our data might be present but remain unseen. By this, I mean the computer encounters us and represents us in pixel (or data) form but fails to recognize us. There might be a face in an image that we, feeble humans, can visually attest is there. But there isn't a data-'face.'

This kind of nonrecognition may occur because computers are unable to identify every particularity and dimension that the world has to offer.[34] Despite attempts otherwise, notably by proponents of the field of digital philosophy, it's quite difficult to assume and represent the entire world according to ones or zeros.[35] What is "seen" and "not seen" is more than a technological phenomenon or limitation but an algorithmic consequence shaped by history: who is empowered to look, what is made visible, and what is made invisible?

Like digital epidermalization, the algorithms that facilitate facial recognition—following the work of surveillance scholar Kelly Gates—are burdened with a politics, a programmed rule set that administers the designated identity of a 'face.'[36] An algorithm might disadvantage some data while privileging others, through either technological failure and/or an innate bias of the algorithm's authors. What can be seen is what can be made intelligible, and what can be intelligible determines who we can be.

To repurpose René Descartes' dictum "I think, therefore I am," if we are made of data, but our data is not "seen," do we exist?[37] Or to less clichédly relate this to computer science, we are encountering a world where "negation as failure" becomes the central mode of existence: "not *p* can be inferred if every possible proof of *p* fails."[38] "Not *p*" is unknowable—it can only be assumed when every alternative is shown to be unsuccessful.

This type of (unintended) negation is precisely what happened in a piece of viral media that made its rounds in late 2009. In the YouTube video "HP Computers Are Racist," two people are shown speaking into a new, HP Mediasmart computer with an embedded motion-tracking and facial-recognition camera. These individuals, identified as "Black Desi" Cryer and "White Wanda" Zamen, are employees at an unnamed Texas retail store. The video (figure I.2) begins with Cryer introducing himself and the technology and explicitly stating his race and the issue at hand with the computer: "I'm black. . . . I think my blackness is interfering with the computer's ability to follow me."[39]

Indeed, the facial-recognition software seems unable to identify Cryer's black face. He asks his white coworker Zamen to pop into frame. The absurdity of the video begins when the camera almost immediately recognizes Zamen, zooming in and following her as she floats around the screen. Cryer, who narrates through all of this, visually reintroduces himself within the camera's view and instantly breaks the facial-recognition software: the frame stops following Zamen, and the camera moves back to its default position.

FIGURE I.2. Desi Cryer and Wanda Zamen appear in the 2009 YouTube video "HP Computers Are Racist." Source: Wanda Zamen, "HP Computers Are Racist," YouTube.com, December 10, 2009.

The undisguised immediacy of the computer's identification/non-identification based on skin color provides a powerful example of the embedded digital epidermalization within HP's software.[40] Whiteness was legible to the machine; blackness was not. In fact, not only was Cryer's blackness illegible, but it also spoiled Zamen's experience, halting the software that was attentively recognizing and tracking whiteness.

HP responded to the video with a brief statement on its website: "We are working with our partners to learn more. The technology we use is built on standard algorithms that measure the difference in intensity of contrast between the eyes and the upper cheek and nose. We believe that the camera might have difficulty 'seeing' contrast in conditions where there is insufficient foreground lighting."[41] This response is compelling for many reasons, not least of which is the technically descriptive explication of why the problem occurred in the first place. HP talks about the algorithm's difficulty in "seeing" the contrast of eyes, upper cheek, and nose due to "insufficient foreground lighting." Impor-

tantly, computers don't see; they compute.[42] And while the metaphor of "seeing" functions as a physical, and near-universally accessible, analogy for the nonvisual algorithmic processing that calculates HSV (hue, saturation, value) contrast, the metaphor's imprecision runs up against the sharp edge of power.[43]

When feminist theorist Donna Haraway writes that the "'eyes' made available in modern technological sciences shatter any idea of passive vision," she points squarely to the impossibility for neutral sight.[44] "Difficulty 'seeing'" is not just the consequence of bad eyesight. To "see" is itself an act saturated in politics, writes visual culture scholar Nicholas Mirzoeff, a manifestation that "sutures authority to power and renders this association 'natural.'"[45]

Nevertheless, HP employs the sense of sight to emphasize its camera's presumed baseline neutrality. And this neutrality is then believed to let HP off the hook. According to the company's account, HP computers weren't racist; they were merely technologically/optically limited. The computer failed to recognize blackness not because of the latter's history or cultural encoding but because of a simple, resolvable lighting problem. Doubling down on the ruse of postracial optimism, HP argued that its computers would surely "see" the various hues of racial difference once this technical problem was addressed.

Cryer and Zamen's experience with HP's technology gracefully intersects the seemingly discrete fields of humanities and social science with the more hard-science approaches of computer science and engineering. Here, it becomes even more apparent, citing digital media theorist Beth Coleman, that "what used to be a matter of flesh and blood is now highly abstracted data. Race has been made information."[46] And while unintentional, the racial dimension of HP's algorithm disadvantaged dark-skinned users. Cryer jokes at the end of the video, "I bought one of these for Christmas. . . . I hope my wife doesn't see this video."[47] Full use of HP's computer was denied to users with darker bodies—a distinction based not on traditional, prejudicial racism but on a functionalist decision that put values on HSV contrast, not skin color.[48]

This functionalist decision wrapped the human body with a new, racialized layer. In the preceding case, 'face,' and thus 'race,' is represented in the digital language of HSV contrast. Waning may be the days of unadorned racist stereotypes and "one-drop" rules, for HP's 'race' is not black and brown versus white but intelligible versus unintelligible data.[49] It's a graphical template that faces either fit or frustrate. The corruption implicit in HP's classification is not rooted in explicit racial bias, and yet the classification nonetheless reflects and reinforces what digital scholar Lisa Nakamura calls a "digital racial formation" that reifies whiteness as normal and blackness as abnormal.[50]

The pattern that authenticates White Wanda's 'face' but denies Black Desi's is not white because it enjoys playing golf and listening to Jimmy Buffett. It's 'white' thanks to a politics of computation that clearly culls from the traces of its offline predecessor: we can only imagine that Cryer would not have had any issue if the HP engineering team and its beta testers were majority dark-skinned. But the realities of the U.S. tech industry forbid that possibility: in 2012, citing U.S. Department of Labor Statistics, Silicon Valley's ten largest companies employed only 6 percent Hispanic and 4 percent black workers. Those numbers are quickly halved to 3 percent and 1 percent, respectively, when we consider only those who hold executive and top-managerial positions.[51]

The unspoken and unintentional presumption of whiteness in HP's facial-recognition technology exemplifies white privilege as more than societal advantage. It embeds seemingly neutral technological and infrastructural projects with (often white-supremacist) racial logics at their most ground level.[52] Accordingly, when whiteness is operationalized as the norm, and thus as the default, the possibility for a universally consistent "we" as data is disrupted.

But while a face cannot be perfectly transcoded into 'face,' neither can users' offline positionalities. Even if "we" is not a homogeneous entity, the asymmetries of the we do not perfectly overlap the asymmetries of the 'we.' More simply put, HP's algorithm might be unable to recognize a white user with a dark tan but could easily authenticate

a black user with lighter skin. It is on the terms of data, and how racial differences are converted into data, that recognition occurs—not the essentializing terms of racialization itself.

DATA WARS AND DATAVEILLANCE

Of course, who we are as data draws from much more than our names and our faces. But a user's mundane datafied behavior—an email, a GPS location, or a visit to sports site ESPN.com—is quite meaningless in and of itself. It is when this data becomes aggregated, tethered together and operationalized, that we, as data, become worthwhile. If that same user also goes to tech blog Arstechnica.com, home-repair site Homedepot.com, and jejune men's magazine Maxim.com, that quick tour around the Internet can become quite meaningful. It might even suggest that this user is 92 percent confidently 'male.'

Our datafied lives, when aggregated and transcoded into quotation-marked categories, increasingly define who we are and who we can be. But to make sense of this aggregation requires a new definition of 'man,' one based on website visits instead of performance or self-identification. In this case, a user is a 'man' according to how closely 'his' data stacks up to preexisting models of 'man.' These models are what I call measurable types, or interpretations of data that stand in as digital containers of categorical meaning. And in order for companies or governments to construct these measurable types, they need a lot of our data.

It was 2007, and the "Data Wars" had just begun.[53] Actually, the exact day was Friday, April 13, and search supergiant Google had just acquired targeted-advertising supergiant DoubleClick for $3.1 billion in cash, ushering in a massive agglomeration of data between Google's search information and DoubleClick's user-tracked marketing data. Rumors that Microsoft and Yahoo! were planning to scoop Google preceded the deal, and the eventual purchase, valued at twenty times DoubleClick's annual revenue, achieved what we could indulgently call a coup "d'ata" for Google.

With this announcement, data itself had become a business, maybe even the central commodity for digital capital.[54] Prior to the deal, Google had access only to my email and search history. Now, blessed by DoubleClick's extensive network, Google also would know which pages I visited, how often I visited them, how long I stayed on those pages when I did, and which pages I went to before and after visiting a certain site. This aggregated data would end up making Google "staggering" amounts of money.[55]

More structurally, by purchasing DoubleClick, Google had woven an Internet-wide surveillance network. This "surveillant assemblage" conjoined the various, decentralized vestiges of data about us and our online behaviors—things we might not care about and/or things we might not even share with our closest confidant—and funneled their flows into Google's own private databases.[56] Today, Google records data from more than a billion Google users, more than three billion search queries a day, more than 425 million Gmail accounts, and traffic from an estimated one million websites, including almost half of the ten thousand most visited.[57]

Indeed, in 2007, one could focus on the potential dangers of how this data could be used—and some did. The acquisition's announcement was met with immediate resistance. Advocacy groups including the Center for Digital Democracy appealed to the U.S. Federal Trade Commission and European Union to stop the merger on grounds of privacy and anticompetition. Marc Rotenberg, executive director of the Electronic Privacy Information Center, testified in a Senate hearing against the acquisition. Even Microsoft, in an ironic postscript to its famed 2001 *United States v. Microsoft* antitrust case, critiqued the proposed merger because it "raise[d] serious competition and privacy concerns."[58]

But the Data Wars were already in full swing. Four months later, Microsoft went on to spend a whopping $6.3 billion to buy digital-marketing parent company aQuantive. In the interim, Yahoo! had purchased two companies at a more sober price. Full ownership of

ad exchange Right Media and advertising network Blue Lithium had set the company back a cool $1.15 billion. Clearly, these "wars" weren't about companies spending more than $10 billion just to invade your privacy. For Google, Microsoft, and Yahoo!, this was about a new type of market dominance: collecting enough data to grow from mere search engines into much more profitable advertising behemoths, capable of providing query results next to exactly defined commercial propaganda.[59] They wanted to be sovereigns over what digital humanist Frédéric Kaplan later called linguistic capitalism, or the commodification of our textual lives.[60]

With all this available data, Google, Microsoft, and Yahoo! could found and fund their own proprietary digital kingdoms at the same time. Yes, we were being surveilled, but this grand aggregation of data wasn't the surveillance of Orwell's "Big Brother" or the FBI's COINTELPRO. Nor did it resemble the kind of two-hundred-year-old portmanteau that paired the French word *sur*, meaning "on/over," with the even Frenchier word *vellier*, meaning "watch." Surveillance in this frame prompts docility: populations are seen as spatially underneath power, both observed and manipulated by a larger, towering, and eternally vigilant entity.

The surveillance embedded in the technical structure of these billion-dollar acquisitions is subtler, more mundane, yet ever more extensive in reach. Google is "watching" us, but we're also voluntarily logging into Google's servers, quickly and wantonly accepting terms of service agreements that relinquish our formal rights to privacy. This is a type of surveillance, surely, but it's also something different. It's what computer scientist Roger Clarke calls dataveillance, "the systematic monitoring of people's actions or communications through the application of information technology" that reconfigures the character of surveillance and its subjects.[61]

For thinkers like Clarke, the introduction of information technology into the surveilling relationship fundamentally changes both what is being watched and how that watched object is perceived. It is our data that is being watched, not our selves. And when dataveillance observes,

saves, pattern analyzes, and uses our own data to profile us, its relationship to liberal theories of privacy becomes muddled. The data that Google, Microsoft, and Yahoo! collect is sometimes encrypted, usually anonymized, and mostly "not personally identifiable," leading some privacy theorists to argue "no harm, no foul" if such data cannot be tracked back to an individual, discrete, personally nameable self.[62]

But what dataveillers know about me is not just a catalogue of different data elements. These elements must be aggregated, cross-referenced, and algorithmically analyzed to produce knowledges about my life that, while not discernible at face value, can then be used by marketers, political campaigns, spy agencies, big-data researchers, and even police departments.[63] As we'll learn in the coming pages, Google's dataveillance creates the datafied templates, or measurable types, for what it means to be an 'old' 'man.' Web-audience measurement company Quantcast constructs the datafied idea of users who are 'Hispanic.' And the NSA uses its incredible government dragnet to pattern analyze who is a 'terrorist.'

These productions of life are the hidden stakes of the Data Wars, the consequences of which go well beyond Google (or Microsoft or Yahoo!) as a unique institution and its role in the general erosion of privacy online.

For example, take the trend toward "predictive policing" in U.S. police departments. In this techno-enthused strategy, police departments use crime statistics to generate maps highlighting five-hundred-by-five-hundred-square-foot areas (one city block) where crimes are likely to occur.[64] A block might be "high crime" at two a.m. on a Friday but "low crime" at two p.m. on a Tuesday. Or we could think like the Chicago Police Department (CPD) in 2013 and channel this logic—*Minority Report*–style—in order to assign "criminality" not to blocks but to people.[65]

Criminology research finds that behavioral traits—like being a victim of a previous shooting, having an arrest record, or being friends with others who are similarly affected by crime—are closely correlated with being either a perpetrator or a victim of a violent crime.[66] Research-

ers have thus theorized that if you or your associates have experience with the criminal justice system and/or crime, you're more likely to experience the criminal justice system and/or crime. At face value, this assumption is quite obvious. Crime and incarceration are structural issues that affect communities, not individual people; they are not random occurrences.[67]

What is more noteworthy is how this algorithmic assessment led the CPD to create a "heat list" of four hundred 'at-risk' individuals, an algorithmic category populated by both 'victims' and 'offenders.' To be algorithmically identified as 'at risk' means that you will physically receive a visit from Chicago's "finest." And to be visited by Chicago's "finest" means you will actually be at risk: a CPD officer will "[warn] those on the heat list individually that further criminal activity, even for the most petty offenses, will result in the full force of the law being brought down on them."[68]

Here, we encounter two foundational interpretations that establish the measurable type of 'at risk.' One, there's an explicit value judgment that social connections themselves are the most efficacious indicators of 'at-risk' behavior. And two, there's an implicit value judgment about what data is available, and eventually used, in processing. It is impossible to divorce the CPD's "heat map" from centuries of preceding racist and classist policing practices, as those who are assigned by the department to be 'at risk' are unlikely to be affluent whites from Chicago's Gold Coast neighborhood. The data used to generate 'at risk' is not representative of a universal population and thus does not treat all populations equally. Nevertheless, to have someone knock on your door because your data is seen to be 'at risk' reaffirms some of the worst fears we might have about this new, datafied world: our data is spoken for louder than we can speak, and it is spoken for on its own terms.

One of these terms is what I have been calling the measurable type. This concept appropriates the sociological design of an "ideal type," or what social theorist Max Weber defines as "the one-sided accentuation of one or more points of view . . . arranged according to those one-

sidedly emphasized viewpoints into a unified analytical construct."[69] Measurable types like 'at risk' are actionable analytical constructs of classificatory meaning, based exclusively on what is available to measure. When "we are data," we are indebted to both how our lives are datafied and how that data is algorithmically interpreted.

Communication scholar Tarleton Gillespie writes that algorithms "are not barometers of the social. They produce hieroglyphs: shaped by the tool by which they are carved, requiring of priestly interpretation, they tell powerful but often mythological stories—usually in the service of the gods."[70] We can think of a measurable type like 'at risk' as a hieroglyph, not a truth of identity but a priestly interpretation. It is not simply an officer who decides our fate in any given encounter with the police. Rather, it's an algorithmic interpretation of our own, datafied social networks that enacts police suspicion.

This brings us back to Pasquale's *Black Box Society*. In a diagram of web surveillance, we encounter the metaphor of "one-way mirrors," in which Internet users remain ignorant of how their data is used while site owners are privileged with near-universal access to that data.[71] Our algorithmic identities are similarly made behind a one-way mirror: it is largely impossible to know what our 'gender' is, how 'old' we are, and possibly if we're 'at risk' (as well as what all of those measurable types actually mean). But there are nonetheless powerful effects to this algorithmic unidirectionality. One effect is that while we know ourselves and exist as beings in highly politicized worlds, we don't know ourselves as beings in highly politicized algorithmic worlds. In most instances of algorithmic identification, we are seen, assessed, and labeled through algorithmic eyes, but our reaction is often available only to be obliquely felt. Like a sudden draft of air from an unseen window, we may very well feel we are being watched but never see what sees.

Accordingly, our own algorithmic sight can never be complete. Instead, it is an impaired sight plus an added *else*. This concept of the else is unquantifiable. Like scopaesthesia, we are affected by the thought, we know that something else is going on, but we're not exactly sure

what it may be. We can't really understand how we're talked about. We can't really experience, directly and knowingly, what our algorithmic identities are. But they're there regardless, and their epistemic corruption alerts us to their presence.

Our algorithmic identities emerge from a constant interplay between our data and algorithms interpreting that data. This dynamism echoes a larger argument, best detailed in philosopher Judith Butler's work around gender performativity, that rejects any stable truth of identity.[72] What we call gender, and what we can call identity at large, is derived from our own particular performance of gender—and not an attachment to some totalizing essentialism around, for example, what it means to be a man. Queer theorist J. Halberstam locates this gender performativity within an algorithmic logic: it is a "learned, imitative behavior that can be processed so well that it comes to look natural."[73]

Gender's naturality is a processed output, one that necessarily changes over time to accommodate new behaviors. What becomes prototypically masculine at the moment is contingent on what men are doing right now. Gender—and, I propose, identity at large—is a processed construct; it "has a technology."[74]

Online, this axiom of performative power shifts from a largely theoretical argument to one that is squarely empirical. Data about who we are becomes more important than who we *really* are or who we may choose to be. We intentionally perform our data when we fill out our name, date of birth, and gender when buying plane tickets online: those pieces of data allow the U.S. government to identify who is traveling from Caracas to Atlanta. And we unintentionally perform our data when we buy food at the grocery store, use our cell phone to invite a friend to dinner, or even just move around our neighborhood (our phone's internal GPS alerts our mobile carrier, Facebook, and our other mobile applications where we are, at what time we are there, how long we stay, and potentially who we are with). Our data is constantly being "watched" by, or in media scholar Mark Andrejevic's words, "interacts" with, algorithmic machines.[75]

But in this uncertainty, we can also find incredible possibility. When identity is made in terms of data, the antiessentialism of measurable-type meaning lets us play with its instability. Google does not discipline us back into being a "good 'man,'" nor do we know what the discursive parameters of Google's 'man' are. Similarly, we can confound the CPD's 'at risk' algorithm by managing not who we are friends with but how those friendships get datafied. In the coming chapters, we'll further explore the logic of algorithmic identification, its operation, and how we can play, experiment, and ultimately resist it.

OUR ALGORITHMIC IDENTITIES

Our algorithmic identities are based on near-real-time interpretations of data. And as we produce more and more pieces of data, these interpretations must necessarily change—the foundry of who we are online lacks epistemic stability. For example, an individual previously not on the Chicago Police Department's radar might become 'at risk' because she made new friends and put their contact information into her phone. You might have previously been unrecognizable according to HP's facial-recognition algorithm, but after the purchase of a new office lamp, you now have a 'face.' And as a user surfs from one site to another online, that user might change from being 92 percent 'male' to 88 percent 'male' within a few seconds. What the world looks like to you today, and who you are seen to be this afternoon, is constructed from the datafied scraps of what you did last night.

This type of dynamism of categorical membership sets the stage for what philosopher Gilles Deleuze has called the "societies of control." In these societies, control regulates its subjects with constant contact to power, whereby structures of constraint move according to "ultrarapid forms of free-floating control."[76] This constraint makes and remakes itself, a process Deleuze describes as modulation: "like a self-deforming cast that will continuously change from one moment to the other, or like a sieve whose mesh will transmute from point to point."[77] As sub-

jects to algorithmic interpretation, who we are also changes "from point to point," moving from 'woman' to 'man' or 'old' to 'young.'

But because these measurable types are also made exclusively from data, new data necessarily changes what defines 'old woman' and 'young man,' as well. Not only does your algorithmic identity readjust when you visit a new website, but the measurable types that are used to compose you modulate as well. "We are data" means that our data, produced in accelerating quantities and with increasing intimacy, is not just data but constitutive material for interpretative, structuring, and ultimately modulatory classifications.

Theorist Tiziana Terranova traces these dynamic algorithmic productions in terms of what she calls "network power." A company like Google doesn't talk to a specific person. Google listens to and talks to that person's data, the subindividual units that we can think of as dividuals: "the decomposition of individuals into data clouds subject to automated integration and disintegration."[78] We are automatically integrated/disintegrated according to the needs of whoever possesses these clouds. For Terranova, when individuals are replaced by dividuals, the categories of identity that we normally think of as politically owned by us, like gender, race, and citizenship (what she calls "macrostates"), become nonlinearly connected to an endless array of algorithmic meaning, like web use and behavior data (what she calls "microstates").[79] In other words, we produce the dividual, microstate data that Google uses to make our algorithmic, macrostate templates. And these macrostates are the conceptual forebears to what I have been referring to as measurable types.

More directly, as Alexander Galloway writes, "on the Internet there is no reason to know the name of a particular user, only to know what that user likes, where they shop, where they live, and so on. The clustering of descriptive information around a specific user becomes sufficient to explain the identity of that user."[80] Our individualities as users may be quite insignificant. Rather, what our dividualities can be algorithmically made say is how we are now seen.

Furthermore, much like "there is no reason to know" a user's name so long as that user is sufficiently profiled, any enduring identity of 'man' itself is similarly ignored. As new, descriptive information (microstates) for what makes a 'man' changes, the resulting measurable type (macrostate) realigns 'man' in accordance. Galloway defines his concept of "protocol" in a parallel way: "an algorithm, a proscription for structure whose form of appearance may be any number of different diagrams or shapes."[81]

When 'gender' is made through different diagrams and shapes according to flows of near-real-time data, 'gender' ceases to be a stable referent of identity. A user might be 92 percent 'male' at 9:30 p.m. But eight hours later, at 5:30 a.m., after spending the evening asleep and visiting no new sites, that user could now be 88 percent 'male.' What 'man' looks like today and who 'men' are seen to be later this afternoon are also constructed from the datafied scraps of what 'men' did last night. When who we are is made from shifting founts of measurable-type meaning, the network power of algorithmic machines, citing Galloway and philosopher Eugene Thacker, "set[s] the terms within which practices may possibly occur."[82] This mode of shifting algorithmic identity is what I have previously called *soft biopolitics*, which I will describe further in chapter 2.[83]

In the preceding analysis of user profiling, I'm not referring to critiques of personalization technologies like those spelled out in legal scholar Cass Sunnstein's *Republic.com 2.0* or author Eli Pariser's *The Filter Bubble*, in which data profiles construct self-affirming echo chambers of targeted advertisements and content according to one's presumed identity, political affiliation, or interests.[84] I'm instead describing a form of control that is much more indirect and unapparent. It's a form that moves the goal posts of what is and what is not true, algorithmically regulating discursive construction largely beyond our gaze and most often without our comprehension.

These profiling algorithms continuously redefine the "knowledge" element in philosopher Michel Foucault's power/knowledge doublet.[85]

They measure and react to streams of available data, diligently following users' new datafied, microstate tendencies all while those users remain largely unaware of their formative role in macrostate construction. More to the point, algorithms delimit the discursive substrates that determine what can and cannot be said, dynamically modifying what Judith Butler calls the "limits of acceptable speech."[86] With ubiquitous surveillance, a macrostate like Google's 'gender' is constantly made and remade, updated based on the infinite fount of new microstate inputs coming from our datafied lives.

And with hundreds of other companies similarly assigning each of us different 'genders,' 'races,' and even 'classes,' there is no single, static sense of us but rather an untold number of competing, modulating interpretations of data that make up who we are.

When profiling algorithms interpret and reinterpret our data, the resulting macrostate emergence comes to define and redefine the conditions of possibility of who we are seen to be online: a Google 'woman,' a CPD 'at risk,' and potentially an HP 'white,' ad infinitum. And as our measurable-type identities shift according to changes in data, and those measurable types dynamically adapt to these changes as well, we encounter an algorithmic version of what artist Mary Flanagan calls "creatures of code," a "virtual body" for which "information has taken over," all while that body's "racial, aged, and sexual categories . . . are constructed as personality features rather than historical or culturally specific categories."[87]

Here, an 'at-risk white woman' is not necessarily an at-risk white woman reared in the structural particularities of that subject position. 'At-risk white woman' is instead a template of datafied performances, a "feature" as opposed to a set of historically or culturally grounded subjective experiences. The gap between these two interpretive modes emphasizes the disempowering asymmetry, quoting digital theorist Anna Watkins Fisher, between how we "use" and how we are "used."[88]

In the book *The Googlization of Everything*, cultural historian Siva Vaidhyanathan writes that while the Internet giant may proclaim users'

control over their use of the system, "as long as control over our personal information and profiles is granted at the pleasure of Google and similar companies, such choices mean very little. There is simply no consistency, reciprocity, or accountability in the system."[89] The one-sidedness of algorithmic identification normalizes this inconsistency at every datafied step.

We Are Data is an attempt to continue a long analytical tradition that interrogates how the terms of information become the terms of our emergent subjectivity. This theoretical move isn't to displace the corporeal from our understanding of ourselves, to be posthuman or a cyborg in such a way that erases the body or affirms a neodualism of mind and body. Rather, I want to explore how companies, governments, and researchers use algorithms to augment our already-posthuman selves with layers upon additional layers of new, datafied identifications.

Our algorithmic identities are made through data and only data. It is a process that gleans from databases; it reads our behaviors as data, our social ties as data, and our bodies as data. And it conditions our present and futures on the basis of a dynamic interpretation of that reading. When media theorist Friedrich Kittler writes in *Discourse Networks* that technologies actively produce the discourses that we as subjects speak and are made through, he's pointing to what I argue is the kernel of algorithmic identity.[90] What is new, though, is how these very discourses are now composed from a near-real-time, dynamic terrain of statistics and commonality models—not from life's subjective experience and the conflicts therein.

And here we encounter the political rub of "we are data": we lack the vocabulary needed to enact a politics around our algorithmic identities. To repurpose a phrase from the field of computer science, we lose expressive power.[91] To think politically about algorithmic 'race,' 'gender,' and 'class' is to attach a politics onto something that is both predominantly unknowable and changes without our knowing. We can't think about 'gender' in the same way as we think about gender. We can't talk about 'race' in the same way either. But what we can do is appreciate

the changes that happen when categorical definitions are reconfigured through an algorithmic locum.

Measurable types of identity like Google's 'gender' or the CPD's 'at risk' are dispersed, dividuated, and saved across a range of databases in many different countries. But more importantly, they are also ever changing, dependent on flows of available data. They are what we can think of as "just-in-time" identities that are made, ad hoc, in an ongoing conversation between our data and the various algorithms that process it.[92] To revise the web-surfing exercise that began this chapter, a simple five-minute sojourn across the online universe fails to construct for us a single identity. It rather assigns us an unstable, multivariate set of modulating identifications.

Paired with other instances of algorithmic processing (facial-recognition technologies, semantic analyses of my writing on my Facebook wall or Twitter feed, a history of purchases on my credit card, etc.), I am utterly overdetermined, made and remade every time I make a datafied move. Through my data alone, I have entered into even more conversations about who I am and what who I am means, conversations that happen without my knowledge and with many more participants involved than most of us could imagine.

To make the world according to algorithm is, to paraphrase one of historian Melvin Kranzberg's laws of technology, neither good nor bad nor neutral.[93] But it is new and thus nascent and unperfected. It's a world of unfastening, temporary subject arrangements that allows a user to be, for example, 23 percent confidently 'woman' and 84 percent confidently 'man' at the same time. Categories of identity of any sort, quotation-marked or not, might not be superficial, but they also aren't impossibly profound. As we move "point to point," the daisy chain of modulation does as well, requiring a buoyancy to categorical meaning and identity as life's changes alter their discursive contents.

But our algorithmic identities are not a recasting of our single, offline self into a digital self. This type of work has already been well documented by sociologist David Lyon, in his work on data doubles, or what

legal scholar Daniel Solove dubs a "digital dossier."[94] Instead, we should locate our algorithmic identities in the spirit of how data sociologist Evelyn Ruppert theorizes her database subject: "the subject is made up of unique combinations of distributed transactional metrics that reveal who they are and their capacities, problems and needs."[95] When "we are data," we become fluidly responsive to these datafied transactions, unsettled like a sailor lost at sea who is slowly lobbed to and fro by the water's silent undulations. Rather than seeing our online identities as aberration or misidentification, I propose we acknowledge that they merely signal a great diversity of who we are and can be, even as different technologies of power desperately try to wrangle us into preexisting boxes of identity.

METHODOLOGY

The research that grounds this book comes from years of investigatory work around an object of study that remains difficult to pin down. The process by which algorithms produce knowledge about us is both unwieldy and often impossible to know. The companies, governments, and researchers that collect, evaluate, and algorithmically interpret our data are agents invested in keeping their interpretations secret. Google's 'gender' and 'age' algorithms are proprietary, as are HP's 'face' and Face.com's 'emotions.'

In response to this methodological limit, I rely on a series of empirical examples to demonstrate the theoretical schematic by which we are made of data. This patchwork, cultural studies approach attempts to identify how different algorithmic logics produce the knowledge by which we live our digital lives. As information technologies increasingly datafy the world and its inhabitants, what our data comes to mean is determined by neither a person nor (as yet) a global, universal "master algorithm," such that "all knowledge—past, present, and future—can be derived from data by a single, universal learning algorithm."[96]

Rather, we are pushed and pulled by different interpretations of our data through different algorithmic instructions. And how those interpretations are made and how those interpretations change according to new data mean that any algorithmic object of study is a moving target. New technologies, new algorithms, and especially new cultural implementations of algorithmic judgment confound any notion of a single "algorithm."

The theoretical arguments that this book makes are not claims about algorithms at large. Rather, they are attempts to show how data about us is used to produce new versions of the world—versions that might differ greatly from their nonalgorithmic counterparts. For this reason, I move from example to example, measurable type to measurable type, in order to paint a fuller picture of how "we are data."

CHAPTER OVERVIEW

This book has two intersecting purposes. First is to understand how algorithms transcode concepts like gender, race, class, and even citizenship into quantitative, measurable-type forms. Second is to recognize how these measurable types reconfigure our conceptions of control and power in a digitally networked world. The political battles that surround structures of patriarchy, white supremacy, and capitalism writ large must necessarily attend to the terms of algorithm.

In other words, while HP's facial-recognition cameras might promote asymmetrical usages across racial lines, an algorithmic 'race' complicates this asymmetry. HP's response is indicative of this complication, in that HSV contrast, not skin color, shared history, or even DNA, stands in for the concept of race. We're talking about HP's construction of a 'white' 'face,' not White Wanda's white body. But we're talking about whiteness nonetheless. Race is incessantly being made and remade, and we must focus on how digital technology also makes and remakes 'race' on algorithmic terms.

Explicating the terms that underpin this making/remaking is the ultimate goal of the book. In the following pages, we'll talk about jazz, terrorists, HP being racist (again), marketing, the NSA, citizenship, and even Santa Claus. The shift to the data/algorithm ontology of the computer conceptually moves identity past explicit, policed boundaries that require negation and exclusivity (either male or female, at risk or not, black or white).

This move lays the foundations for a plane of smoothness, an open set of possibilities where we play on the limits of established truth. Algorithmic identity doesn't declare that you are just 'male' or 'female.' Statistical confidence and probability, even the chance that this book will spontaneously combust, can never be 100 percent anything. Rather, you're likely to be 92 percent confidently 'male' and 32 percent confidently 'female.' Algorithmic 'race' and 'gender' isn't about being a white man. It's about being a 'Caucasian' 'man' with a confidence measure of 87 percent. In algorithmic identity, we confirm the inorganic realities of Donna Haraway's cyborg, one who is "not afraid of permanently partial identities and contradictory standpoints."[97]

Chapter 1. Categorization: Making Data Useful

In order to compute something like 'woman' or 'smiling,' one needs to first make data useful. In chapter 1, I describe the how-to of algorithmic knowledge production. This how-to centers on how computers create categories through patterns in data, which then construct algorithmically transcoded ideas about the world that I call measurable types. Algorithms are neither magical nor mysterious. Instead, they make data useful through a very intricate but, I promise, also very interesting constellation of different technologies (like metadata or marimbas) that then create different algorithmic identifications (like 'terrorist' or 'John Coltrane').

Chapter 2. Control: Algorithm Is Gonna Get You

Measurable types are much more than descriptive containers for algorithmic meaning. They also play formative roles in how life and knowledge is controlled. With the aid of Gilles Deleuze's concept of modulation, I theorize how the deluges of data we produce online help enact a form of control. This type of control substitutes the strongly worded, hard-coded prohibitory "no!" of traditional modes of power in exchange for what some scholars have called "control without control"—and that I call soft biopolitics. These soft biopolitics describe how our algorithmic identities can regulate life without our direct participation or realization.

Chapter 3. Subjectivity: Who Do They Think You Are?

Soft-biopolitical measurable types structure our lives' conditions of possibilities every time we buy a plane ticket, cross a border, or translate a document on Google Translate. While we are ceaselessly made subject to different arrangements of algorithmic knowledges, these *datafied subject relations* are foreign to our most immediate experiences. We are not individuals online; we are dividuals. And without the philosophical anchor of the individual to think alongside, we are often at a loss in how we interpret ourselves as users. This chapter explores how algorithms make us subject in ways unique to online, algorithmic life.

Chapter 4. Privacy: Wanted Dead or Alive

How does one practice privacy in a world where not only is almost everything surveilled but that surveillance is rarely, if ever, felt? I evaluate privacy's legacy and outline its origins in the nineteenth-century phrase "right to be let alone," in order to bring that history into conversation with the exigencies of our contemporary era. I argue that privacy cannot just be about whether you have a password on your email or

whether there are doors on a bathroom stall. Privacy must be a practical response to the lived restriction and control implicit in ubiquitous surveillance. In this way, I theorize a *dividual privacy* that focuses especially on how the freedom in being "let alone" might translate to a datafied, algorithmic world.

Conclusion: Ghosts in the Machine

At the end of the book, I return to my central arguments: online we are made, read, interpreted, and intelligible according to data. Our world, and the knowledge that gives it its meaning, is increasingly a datafied world. We are subsequently understood in the datafied terms of dynamic, soft-coded, and modulating measurable types. The contemporary encounters we have with ubiquitous surveillance suggest a new relationship to power that I term soft biopolitics. And the resulting ubiquity and emergent configurations of these different types of knowledge force us to rethink how subjectivity functions and what it is that privacy can practically defend.