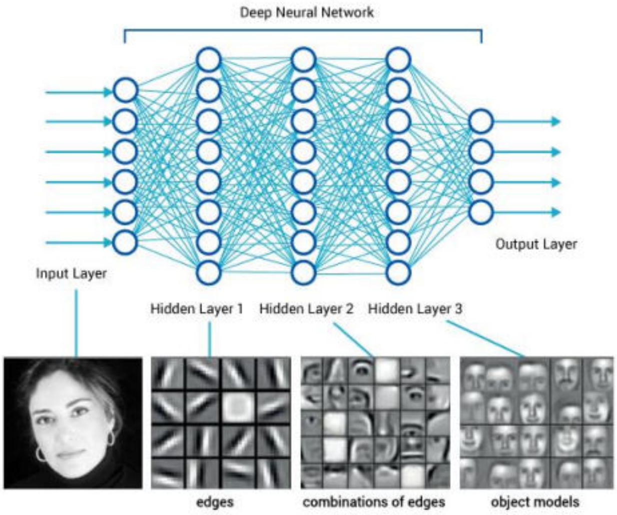
Septembre 2020

Marie Lechner & Yves Citton

Angles morts du numérique ubiquitaire



Sélection de lectures, volume 2

Fondamentaux & Domaines

Sommaire

Fondamentaux

Domaines

Article

Science, Technology, & Human Values
1-25
© The Author(s) 2015
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0162243915606523
sthv.sagepub.com
SAGE

Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness

Mike Ananny[1]

Abstract

Part of understanding the meaning and power of algorithms means asking what new demands they might make of ethical frameworks, and how they might be held accountable to ethical standards. I develop a definition of networked information algorithms (NIAs) as assemblages of institutionally situated code, practices, and norms with the power to create, sustain, and signify relationships among people and data through minimally observable, semiautonomous action. Starting from Merrill's prompt to see ethics as the study of "what we ought to do," I examine ethical dimensions of contemporary NIAs. Specifically, in an effort to sketch an empirically grounded, pragmatic ethics of algorithms, I trace an algorithmic assemblage's power to convene constituents, suggest actions based on perceived similarity and probability, and govern the timing and timeframes of ethical action.

[1]University of Southern California, Los Angeles, CA, USA

Corresponding Author:
Mike Ananny, University of Southern California, 3502 Watt Way, Los Angeles, CA 90089, USA.
Email: ananny@usc.edu

Keywords
ethics, information algorithms, networked technology, probability, similarity, automation, time, Internet

What new approach to media ethics might algorithms require? In comparison to concerns over how to produce or circulate media ethically, train ethical media professionals, or ethically regulating media industries, what might it mean to take an algorithmic assemblage—a mix of computational code, design assumptions, institutional contexts, folk theories, user models—with semiautonomous agency as a unit of ethical analysis?

This essay is an attempt to define a networked information algorithm (NIA) and suggest three dimensions for scrutinizing its ethics: the ability to *convene* people by inferring associations from computational data, the power to judge *similarity* and suggest *probable* actions, and the capacity to organize *time* and influence when action happens. I argue that such a framework might give starting points for holding algorithmic assemblages accountable and develop this argument through critical readings of NIAs in contemporary journalism, online commerce, security and policing, and social media.

Three Approaches to the Intersection of Information Technology and Ethics

Most basically, ethics is “the study of what we ought to do” (Merrill 2011, 3) and is usually divided into three subareas. The first, associated with Kant’s ([1785]2002) call for categorically guided action through reason, is a *deontological* approach: a fixed set of duties, rules, and policies define actions as ethical. Break these rules and you have behaved unethically. The second, associated with the utilitarian philosophies of Jeremy Bentham and John Stuart Mill and related to the American school of pragmatism, is a *teleological* approach focused on the consequences. Ethics should help people choose “the action that will bring the most good to the party the actor deems most important” (Merrill 2011, 11). Finally, the *virtue* model of ethics (Hursthouse 1999) is unconcerned with duties or consequence, focusing instead on the subjective, idiosyncratic and seemingly nonrational impulses that influence people in the absence of clear rules and consequences. It is “more spontaneous” and “motivated by instinct or a spiritually motivated will” (Merrill 2011, 12).

These frameworks have rough parallels to dominant ways of understanding the ethical dimensions of technologies. The first, rooted in policies and regulations, attempts to codify the ethical development and use of technologies, creating standards for punishing errors, teaching best practices, and preventing future failures. For example, the rapid proliferation of intercontinental ballistics spurred the Computer Professionals for Social Responsibility group to create a "Ten Commandments of Computer Ethics" (Computer Ethics Institute 2011) for engineers to ethically develop and use computational weapons systems. Such codes have become the key techniques for teaching engineering students how to ethically build and use semiautonomous cybernetic systems, decision support technologies (Cummings 2006), and robotic "artificial moral agents" (Wallach and Allen 2008).

Instead of applying ethical rules to technologies, a second approach tries to anticipate ethical concerns raised by technological innovation. For example, bioethics emerged as a field largely because new technologies were introduced "with great hopes but little forethought" into a world in which "physicians had almost total control of information and decision-making power" (Levine 2007, 7). It was impossible to apply the existing ethical frameworks because new technologies were fundamentally reconfiguring relationships among doctors, nurses, technicians, patients, and families; new questions about risk, health, life, and death stretched beyond the scope of the existing ethical framework. Similarly, the definition of ethical journalism as the disinterested pursuit of neutral facts for broad consumption emerged, in part, from sociotechnical innovations. The telegraph made it possible to think of stories as the transmission of "pure" information for rational consumption (Carey 1989), and mass-scale advertising and distribution regimes rewarded risk-averse newspapers that appealed to the widest possible array of audience preferences (Schudson 1978). Technologies and economics thus created a journalistic objectivity that outstripped the profession's existing professional frameworks (Schiller 1979), showing of any era's definition of ethical journalism always reflects rapidly coevolving press tools and practices.

The third approach focuses on the values and beliefs of technologists themselves. Grounded in the claim that artifacts with "political qualities" (Winner 1986, 20) give certain people, ideas, and events more visibility and power than others, it asks how "designers and producers include values, purposively, in the set of criteria by which the excellence" of their artifacts are judged (Flanagan, Howe, and Nissenbaum 2008, 322). Such approaches trace the clues that designers leave about their *own* ethical standards in everything from web browser cookie management systems, workplace plasma displays, and urban simulation software (Friedman, Kahn, and Borning 2006).

Such standards are not explicit in official codes of conduct but exist at the level of individual, seemingly idiosyncratic practice. They emerge informally as designers create systems with "value levers" (Shilton 2012) that users can use to enact what *designers* see as acceptable and desirable applications. Akin to the "virtue approach," this approach takes the designer and his or her context as the primary units of analysis, tracing how ethics emerges not from formal standards or broad institutional patterns, but from a technologist's own values and choices.

In reality, technology ethics emerges from a mix of institutionalized codes, professional cultures, technological capabilities, social practices, and individual decision making. Indeed, ethical inquiry in any domain is not a test to be passed or a culture to be interrogated but a complex social and cultural *achievement* (Christians et al. 2009). It entails anticipating how the intersecting dynamics of a sociotechnical system—design, interpretation, use, deployment, value—"matter" for the future (Marres 2007)—and figuring out how to hold these intersections accountable in light of an ethical framework.

Media ethics usually frames accountability in terms of two questions: "accountable *for what*?" and "accountable *to whom*?" (Glasser 1989, 179), but these questions are usually asked of mature media systems (McQuail 2003)—technologies, institutions, and professions that are relatively stable and understood well enough to describe how they behave and how they should be regulated. There may be little consensus on how *exactly* to hold newspaper, television, radio, or cable television industries accountable, but their form, power, meaning, and genres are understood clearly enough to debate with some clarity which standards and people should hold them accountable.

But when technologies and media systems like algorithms are new—before the "wider social-cultural milieu" has prevented them from having "more than one interpretation" (Pinch and Bijker 1984, 409)—they need ethical critiques that keep flexible and contestable their fundamental forms, power, and meanings. Before "social interactions between and within relevant social groups" have made systems "less and less ambiguous" (Bijker 1995, 270-71) and harder to reinterpret, there is an opportunity to intervene and influence their ethics. If what they are or might be can be placed clearly and creatively in terms of an ethical framework, we may discover new ways of holding them accountable before forces of "closure and stabilization" (p. 279) limit debate about how they work and what they mean.

Defining NIAs

Computer science defines an algorithm as a "description of the method by which a task is to be accomplished" (Goffey 2008, 15). Rendered in any programming language and judged according to how quickly and reliably they transform known inputs into desired outcomes, algorithms are generic solutions for well-defined problems. They are the clearest evidence of computation's power to be "a positivistic dominant of reductive, systemic efficiency and expediency" (Galloway 2011, 100).

But this computational definition belies algorithms' sociological and normative features, for example, their power to:

- sort and rank the social web, signaling search quality (Mager 2012) and organizing online communities (Bucher 2012);
- spur commercial activity and direct flows of online capital (Webster 2010);
- organize people into audiences (C. W. Anderson 2011) while automatically creating (Carlson 2015), recommending (Beam 2014), and reading news (Kleinnijenhuis et al. 2013) with little human oversight (Diakopoulos 2015);
- optimize international online labor markets (Kushner 2013);
- create "cyborg finance" (Lin 2013) systems that operate faster than human comprehension (Arnuk and Saluzzi 2012);
- direct military drones to target locations before requesting firing authority from human operators (Calo 2015).

I use the term "networked information algorithm" for two reasons: to distinguish the object of study in this article from computer science's purely mathematical, mechanistic focus and to make it possible to consider the ethics of the sociotechnical *relationships* producing, interpreting, and relying upon the formation processed by computational algorithms. The aim is to describe a unit of ethical analysis—a target for media accountability—that is not a code or a human action on code but, rather, an intersection of technologies and people that makes some associations, similarities, and actions more likely than others.

Algorithms "govern" because they have the power to structure possibilities. They define which information is to be included in an analysis; they envision, plan for, and execute data transformations; they deliver results with a kind of detachment, objectivity, and certainty; they act as filters and mirrors, selecting and reflecting information that make sense within an algorithm's computational logic and the human cultures that created that logic Gillespie (2014). Algorithms do not simply *accelerate* commerce, journalism, finance, or other domains—they are a discourse and culture of knowledge that is simultaneously social and technological, structuring how information is produced, surfaced, made sense of, seen as legitimate, and ascribed public significance (Beer 2009; Bucher 2012; Striphas 2015).

Various types of resistance and dissent are emerging in response to such power. Some criticize the intellectual property and professional norms that keep algorithms private and call for transparent code (Diakopoulos 2015; Pasquale 2011). Others challenge algorithms as unconstitutional when they make "editorial decisions that are neither obvious nor communicated to the reader" (chilling speech) or "single out speakers" without their consent (invading privacy; Benjamin 2013, 1446). Others suggest hiding from algorithms by de-indexing files from search engine crawlers or using anonymous currencies like bitcoin (Maurer, Nelms, and Swartz 2013). Others audit them to derive their inner workings (Sandvig et al. 2014) or purposefully give "misleading, false, or ambiguous data with the intention of confusing" algorithms (Brunton and Nissenbaum 2011, np).

Part of the challenge of critiquing and resisting algorithms is locating them in the first place. Like infrastructure (Star and Ruhleder 1996), algorithms are embedded within the sociotechnical structures; they are shaped by communities of practice, embodied in standards, and most visible when they fail. But, distinct from infrastructure, the relevance, quality, and stability of algorithms depend upon end users. Machine learning algorithms need a great deal of data before they are useful or reliable, social network algorithms require a significant number of nodes before they are able to describe or influence an online community, and recommendation and prediction algorithms observe data flows for long periods of time before they create useful forecasts. It matters little if the "black boxes" of algorithm code (Pinch and Bijker 1984) are opened or comprehensible since they only become *ethically* significant in relation to others.

Understanding how algorithmic ethics is relationally achieved can be helped by applying frameworks designed to trace networks of sociotechnical power. Latour (2005) traces how humans and nonhumans together create and stabilize controversies, produce knowledge and associations, and surface ethical tensions. Similarly, "neo-institutional" studies of organizational technologies (Orlikowski 2010) show how "loosely coupled arrays of standardized elements" (DiMaggio and Powell 1991, 14)—individuals, laws, norms, professional ideals, economic priorities—combine to make technologies that a network *sees* as workable or socially acceptable (or not). Napoli (2014) goes so far as to define algorithms *as* institutions because of their power to structure behavior, influence preferences, guide consumption, produce content, signal quality, and sway commodification.

With these relationships in mind, I define an NIA as an assemblage (DeLanda 2006; Latour 2005) of institutionally situated computational code, human practices, and normative logics that creates, sustains, and signifies relationships among people and data through minimally observable, semiautonomous action. Although code, practices, and norms may be observed *individually* in other contexts, their full "meaning and force . . . can only be understood in terms of relations with other modular units" (Chadwick 2013, 63). For example, Google News' results differ as the page rank algorithm changes, as it is personalized for different individual user profiles, and as Google judges some different news as more worthy of indexing than others. It makes more sense to talk about the ethics of a particular Google News assemblage than the ethics of its algorithm.

Studying the ethics of such assemblages entails not just reading black boxes of code for values (Steen 2014) but also criticizing assemblages "in ways that might serve the ends of freedom and justice" (Winner 1993, 374-76). Such an ethics ignores the unanswerable question of whether code is biased or not (Edelman 2011) and instead asks whether different assemblages "help us get into satisfactory relation with other parts of our experience" (James 1997, 100). The crux of this ethics, of course, rests upon a rich and diverse debate about what "satisfactory relation" means and assemblages create the conditions under which an algorithm might be seen as "wrong" (Gillespie 2012). This pragmatic focus answers Latour's (2004) call for studies of science and technology to move beyond "matters of fact"—deconstructing and explaining sociotechnical systems—to "matters of concern."

Critiquing NIAs

In identifying the matters of algorithm concern, my approach breaks down Merrill's claim—that ethics is the study of "what we ought to do"—into constituent concepts that can be traced across algorithmic assemblages. This critique is not intended as a comprehensive account of algorithmic ethics—other ethical claims could be operationalized and other assemblage dimensions could be analyzed—but it attempts to move *toward* a model of algorithm ethics by asking *when*, *how*, and for *whom* NIAs work. Specifically, how do NIAs convene a “we” (a collective of ethical concern)? How do algorithms encode chance and certainty, suggesting what should probably happen (the likely set of influences and outcomes needing ethical critique)? And how does an assemblage’s construction of timing and timeliness influence when action is taken (creating timeframes over which ethical concerns can play out)?

Convening Constituents by Algorithmically Inferring Associations

Publics emerge when technologies create associations by aggregating people. “Who is inside and outside, who may speak, who may not, and who has authority and may be believed” (Marvin 1990, 4) depend on communication technologies that see some people as like or unlike others, despite variations the technologies cannot capture. Maps, newspapers, museums, and censuses help people see themselves as part of a common group, eliding differences and excluding those not represented in these media (B. Anderson 1983). Opinion polls and market surveys collapse contentious disagreements or subtle variations into binaries and predefined categories that underpin political action (Herbst 1995) and create commercial markets (Igo 2007). Such technologies efficiently align interests and enable a type of collective action—but they also have the power to artificially limit a group’s size (Dahl and Tufte 1973), “compel” association where none is chosen (Rosenblum 2000), and aggregate people into groups without their consent (Salmon and Glasser 1995).

NIAs exercise this aggregative power by semiautonomously sorting data into categories and drawing inferences, through surveillance infrastructures that most people never encounter directly (McKelvey 2014). For example:

- The National Security Agency (NSA) uses cell GPS data to infer individual locations and relationships (Soltani and Gellman 2013) and Google’s Advertising algorithmically labels people as potential terrorists (Soltani, Peterson, and Gellman 2013).
- Analyzing Facebook data, researchers at the Massachusetts Institute of Technology observed that “the percentage of a given user’s friends who self-identify as gay male is strongly correlated with the sexual orientation of that user” (Jernigan and Mistree 2009, np), algorithmically inferring unrevealed orientations.

- Analyzing phone metadata of a relatively small population, Mayer and Mutchler (2014) correctly inferred caller identities, relationships, occupations, medical conditions, religious affiliations, and political beliefs.
- An ethically controversial study automatically filtered Facebook users' content to be "positive" or "negative" to show that the emotional content of people's subsequent posts could be algorithmically influenced (Kramer, Guillory, and Hancock 2014).
- Computer scientists recently produced "images that are completely unrecognizable to humans, but that state-of-the art [deep neural networks] believe to be recognizable objects with 99.99% confidence" (Nguyen, Yosinski, and Clune 2014, 1).

Each of these examples entails algorithms deriving categories and creating associations by sensing and combining aspects of the world they have been programmed to see (Cheney-Lippold 2011). People who fail to leave data that can be categorized are effectively invisible to the database and algorithm (Lerman 2013), but those who leave few traces can still be categorized: reliable pattern-matching often does not require "big data" but small amounts of densely connected metadata that an algorithm is programmed to see as related.

A *deontological* critique would ask how much such algorithmic samples look like broader demographic categories: Does Twitter's distributions of genders and ethnicities match those of the United States? How do Facebook's 1 billion-plus users align with global population patterns? Do high-frequency trading algorithms simply speed up the transactions people would have made anyway? A *teleological* critique of algorithmic convening is rooted in pragmatism. It asks whether the algorithms of Facebook, Twitter, the NSA, or high-frequency trading produce "satisfactory relations with other parts of our experience" (James 1997, 100) without worrying whether algorithms recreate the existing demographic patterns. A *virtue-based* critique of convening would ask how designers think people *should* be aggregated, what comparison and association they build into their designs, and how audiences interpret the associations algorithms present them. Deontologically acceptable NIAs correspond with how standards *outside* the assemblage have already sorted the world, teleologically acceptable NIAs produce associations that people see as efficacious, and acceptable virtue-based algorithms align with designers and users' local, idiosyncratic hopes for and expectations of the world.

Algorithmic convening thus poses a complex ethical challenge. It is difficult to criticize algorithmic convening on deontological grounds because the inner workings of algorithms are proprietary and thus hard to compare to other types of associational technologies (like the census or opinion polls). It is difficult to criticize algorithmic convening on teleological grounds since the effects of a single assemblage are not universally distributed—different people experience different algorithmic assemblages differently. Finally, it is difficult to criticize the virtue of algorithmic convening because we can usually only evaluate what algorithms *produce*, with little insight into the dynamics of the cultures that created them. Most insights we have into the priorities, values, and compromises that determine how an algorithm convenes groups come from corporate self-reporting (Facebook 2013; Google n.d.), post hoc analyses (Bucher 2012), auditing (Sandvig et al. 2014), or reverse engineering (Seaver 2014).

An ethical critique of an algorithmic assemblage that convenes people could be multidimensional, analyzing how well its aggregates adhere to external standards, how its affiliations are interpreted and deployed, and what kind of assumptions and values underpin the cultures that create such associational technologies.

Governing Action by Judging the Probability of Similarity

The second aspect of understanding how NIAs govern ‘‘what we ought to do’’ rests upon understanding how they judge similarity and probability. How closely and confidently do they see a situation resembling a previous one?

Recommendations based on probable similarity raise ethical concerns because when unobservable and seemingly objective computational logics equate two or more instances, people see ‘‘resemblances between certain acts’’ as ‘‘completely natural and self-evident.’’ This makes it harder for them to recognize ‘‘genuine differences,’’ generate alternatives, defend unsuggested actions, or argue for exceptions to similarity (Hofstadter and Sander 2013, 10). Many search algorithms organize their outputs by relevance, but the ethical provenance or significance of such judgments is often unclear. For example, Facebook can help ‘‘lenders discriminate against certain borrowers based on the borrower’s social network connections’’ (Sullivan 2015) and online advertisers can use racial stereotypes to create targeted ads (Sweeney 2013)—but to criticize or resist such predictions means understanding how algorithms create and associate *categories* like ‘‘friends with,’’ ‘‘credit risk,’’ ‘‘black-identifying names.’’

Categories give people ‘‘the feeling of understanding a situation,’’ helping them ‘‘to draw conclusions and to guess about how a situation is likely to evolve" (Hofstadter and Sander 2013, 14-15). They are shared impressions of the world and shortcuts that reduce the risk of misinterpreting new data or situations. But categories are also evidence of the power to strip "away the contingencies of an object's creation," to put "the thing that does not fit into one bin or another ... into a 'residual' category" that signals marginality, impurity, or an outlier accident (Bowker and Star 1999, 299-300). Algorithmic categories raise ethical concerns to the extent that they signal certainty, discourage alternative explorations, and create coherence among disparate objects—categorically narrowing the set of socially acceptable answers to the question of what ought to be done. Consider the following examples:

- Google's Autocomplete (Garber 2013) algorithm finishes people's search queries by comparing them to content and people it sees as similar, reinforcing cultural stereotypes (Baker and Potts 2013) and dissuading people from unpopular searches (Gannes 2013).
- Facebook algorithms track users across the web, watching what they click on, read, share, and comment on to create a personal preference history that organizes Facebook's News Feed and suggests actions (Gerlitz and Helmond 2013). It recommends purchases it sees as similar to users' profiles and suggests news it sees as consistent with past reading behavior (Nielsen and Schrøderb 2014).
- Amazon.com product recommendations are primarily based on how similar an *item* is to those that others have purchased, rated, or viewed (Linden, Smith, and York 2003). This "item-to-item" approach makes it easy to make recommendations to customers who have purchased little, overcoming the lack of "transactional data" (Beer and Burrows 2013) to suggest purchases consistent with similarities among products. Recommendations for what ought to be purchased come not from the similarities among people or consistency with past behavior but from categorical resemblances among objects.

These examples raise ethical concerns because each case—recommending a search, standardizing a user's online behaviors, and suggesting a purchase—involves unseen, categorical, computational judgments about which searches, articles, or purchases should *probably* come next. Users are not offered limitless options but are, in fact, given a narrowly construed set that comes from successfully fitting other people, past actions, and inanimate objects into categories—using categories to discipline action.

Such algorithmic assemblages are simply the latest version of computational systems disciplining users within a narrow set of actions the computer expects (Suchman 1994). Efficient and scalable systems *require* stable categories of people who have learned to say certain words, click certain sequences, and move in predictable ways. This is the ethical power of algorithms: to create a disciplined *network* of humans and machines that resembles and recreates probabilities, making the set of possible outcomes the *model* anticipates likely and reasonable (Mackenzie 2015). Efficient—but not necessarily ethical—algorithmic assemblages use such probabilities to suggest what ought to be done.

Such similarity systems can fail and be resisted, though. Targeted advertisements, for example, made people "uncomfortable if [they] seemed to know too much of their past behavior" but were acceptable again if they "perfectly aligned" people's interests (Wohn and Sarkar 2014, 577). The discomfort with such "uncanny valleys" (Mori 1970) of similarity may not only be the evidence of failed algorithms but starting points for investigating the *ethical* limits of similarity. That is, algorithms that produce results judged as too similar—or the "wrong" kind of similar—may represent moments when people find algorithms' ends, means, or values as too inconsistent with personal codes, too unhelpful for navigating social relationships, or too misaligned with their ethical idiosyncrasies. For example, my Facebook connections may indeed reliably predict my credit risk, but the algorithm driving this prediction may be ethically dubious if it simply accepts similarities between social connections and financial behaviors without seeing structural racism and socioeconomic discrimination as mediators—judgments, categories, and similarities that may be hard to computationally encode.

The *ethics* of a probabilistic system cannot only be judged by "the degree of belief warranted by evidence" it provides (how much it can be trusted) or its ability to "produce stable relative frequencies" (how often it should be trusted; Hacking 2006, 1). What is *also* required is a sensitivity to the categories it uses and a sufficiently creative imagination able to envision other, *better* types of similarity that might produce more "satisfactory relations with other parts of our experience" (James 1997, 100).

Setting Deadlines and Governing Rhythms

Algorithmic assemblages can also suggest *when* action should be taken, but such suggestions depend on how quickly and confidently an assemblage produces results with an acceptable risk of error. Computer scientists use "big-O" notation to indicate "whether a given algorithm will be able to run in a reasonable amount of time on a problem of a given size," suggesting how much error might be tolerated at any moment in the algorithm's operation (Skiena 1998, 16).[1] Such notation is a shared language for analyzing the temporal dynamics of code, a way to quantify the risk of interrupting an algorithm. If slow and fast algorithms are stopped after the same amount of time, the slow algorithm may have produced *more* error-prone results than the fast algorithm (because its conclusion is based on fewer pieces of data), or it may have produced *less* error-prone results (because it has more confidence in the answers it did have time to give). If you know how a code works, you can calculate the probability that an algorithm's results are correct at any point in time.

It is harder, though, to time an *assemblage's* results—to understand how long a mix of code, people, practices, and norms requires to produce meaningful, trustworthy results. For example:

- Twitter's "Trends" algorithm "identifies topics that are immediately popular, rather than topics that have been popular for a while or on a daily basis" (Twitter 2014). A small number of users who frequently tweet is responsible for most of these trends (Asur et al. 2011) and Twitter staff sometime intervene to hand-curate trends (Gillespie 2012). A trend's ethical significance—how its patterns might suggest action at any particular moment—depends on momentary confidence in the trend, on actors' power to interrupt the algorithm, freeze its results, act on answers, or wait for more data. The Twitter assemblage's preference for immediacy (sensitivity to frequent tweeters, the code's design, staff interventions) makes it less useful for taking action supported by longer-term views.
- News organizations frequently use algorithms to list the "most e-mailed" or "most read" articles on their websites. But, unlike the rhythms that have traditionally organized news publishing (morning and evening newspapers, six-o'clock newscasts; Schudson 1986), the actions of distributed users determine which list items persist or decay. The rhythms that produce clicks, forwards, tweets, likes, and posts from other parts of the web are beyond the control of news organizations and susceptible to third-party algorithms that surface stories (e.g., Twitter trends, Facebook News Feed, Google News), making it impossible to reassemble an online audience (Lehmann et al. 2013). If networked news organizations earn their democratic legitimacy, in part, from convening and sustaining conversations with distributed audiences, they

have an ethical imperative to break news, update audiences, issue corrections, and give a historical context. But implementing this imperative depends upon an algorithmic assemblage of networked news time: people, code, practices, and norms extending far beyond the newsroom that create the networked press's rhythms and timeliness.

- Algorithms can also anticipate future actions. Police departments in Los Angeles (Berg 2014) and New York use "predictive policing" algorithms to combine historical crime data with real-time, geo-located tweets, deploying officers "where and when crime is most likely to occur" (Morrison 2014). And Pennsylvania is considering allowing judges to use statistical estimates of future offenses to determine an inmate's current sentence—punishing them not only for crimes they have committed but crimes that algorithms think they *might* commit (Barry-Jester, Casselman, and Goldstein 2015). Algorithmic ethics resemble actuarial ethics: a prediction's legitimacy is based not only on the probable correctness of a current calculation but on the risk of applying that calculation in the future. If "risk is a product of human imaginations disciplined and conditioned by an awareness of the past" (Jasanoff 2010, 15), predictive algorithms are a key element of disciplining and conditioning ethical imagination—of envisioning what might or ought to be done.
- Algorithms can also influence memory. The Internet Archive (2001) lets sites opt out of its index by including the following lines of code in its webserver's "robot.txt" file:

```
User-agent: ia_archiver

Disallow: /
```

- The *Washington Post* (2014) uses this code to prevent the archive from indexing its site, while the *New York Times* (2014) uses similar code to prevent the Associated Press and Reuters from archiving its site. Even without these blocks, Thelwall and Vaughan (2004) show how the Internet Archive algorithmically narrows its own archive: since its crawler algorithm privileges sites that already have links to them, countries with less densely linked websites can fail to appear in the archive altogether. Similarly, researchers collecting tweets using Twitter's own Application Programming Interface report having incomplete data sets compared to accessing the full archive through the Twitter's exclusive data "firehose" (Driscoll and Walker 2014)—the same moment can

> be remembered differently depending on the sampling algorithm used. If data-based decisions about what *should* happen are to align with—or purposefully differ from—records of what *has* happened, then we need to understand how algorithms organize the past and thus influence memories.

Unlike algorithmic convening (when algorithms construct the "we") or algorithmic similarity (when algorithms create the space of probable action), algorithmic *timing* entails prediction, interruption, and anchoring—using algorithms to suggest when an event will likely happen, the relevant time frames, the memories to recall. What does it mean if public attention assembled by an algorithm appears only briefly and dissipates before it can be understood? If public attention no longer exists, does it need to be accounted for? If there is no record of public attention, how can it be recreated or prevented from reoccurring? Since Google Search, Facebook News Feed, and Twitter Trends continually change their algorithms without public oversight, which *versions* of an assemblage should be held responsible for ethically questionable outcomes?

Answering these questions requires seeing how algorithmic assemblages create what Durkhein called a consensus on "temporal orientation'" (Durkheim [1912] 1954, 440). Consensus is not necessarily agreement but, rather, the product of forces battling to mark time, to define stops and starts, to make interruptions, to say that enough is known to act. For example, understanding contemporary, networked "news time" means tracing how the power to structure time is distributed among news organizations, social media companies, and their respective practices, code, actors, and norms. Part of holding the media ethically accountable for its organization of people's time and attention means appreciating how algorithmic assemblages order events, suggest causes, orient attention, recall memories so that some actions might be taken over others, some consequences secured and others avoided. (Dewey 1954, 12)

Conclusion

Starting from an admittedly simplistic notion of ethics as "the study of what we ought to do," my aim has been to sketch an ethics of NIAs. Specifically, how algorithms convene a "we," judge similarity, and create time—all in order to suggest which actions are likely to happen, and when.

My definition of NIAs as *assemblages* of institutionally situated code, human practices, and normative logics may seem overly broad, but it is intended to narrow the empirical study of algorithmic ethics to the linkages *among* empirical sites. I unpacked the simple definition of ethics as ‘‘the study of what we ought to do’’ into its conceptual constituents—convening, probability, time—to create concepts that can only be fully appreciated in relationships among algorithmic code, practices, and norms. The assemblages governing the question of ‘‘what we ought to do’’ might, therefore, be seen as a three-by-three matrix of concepts (convening, probability, time) and actants (code, practices, norms)—potential actions and their ethical significance exist at this matrix's intersections. To be sure, the concepts and actants might change or be reformulated in response to different ethical theories and new empirical contexts. The framework offered here is meant only as a step toward analyzing the empirical and normative dynamics at play in NIAs.

Such frameworks are urgently required because media are increasingly susceptible to algorithmic assemblages. Algorithms are created by professionals with shifting boundaries (software designers move among social media, ecommerce, and networked news platforms), algorithmic technologies have unpredictable outcomes (outputs cannot be understood by any single programmer or controlled by any one organization), and algorithmic ecosystems are increasingly personalized (media reaches consumers through myriad and opaque rules and values). The existing approaches to media accountability that assume stable technologies and clear questions are outstripped by the dynamic and contested nature of algorithmic assemblages. Some see accountability existing as code transparency, others seek state regulation of companies with algorithmic monopolies, and others aim to build algorithmic literacy among end users. Each unit of analysis is important but considering the ethics of each on isolation misses appreciating the full power of algorithmic assemblages.

Unlike other media technologies whose ethical dynamics might be evaluated when they are designed, deployed, or interpreted, NIAs and their ethical dimensions are moving targets. A purely deontological approach might be applied to the entire assemblage—asking whether its rules and policies adhere to ethical principles—but it may be difficult to trace which parts of an assemblage adhere to or deviate from deontological guidelines. A strictly teleological approach focused on ends and consequences may be the most effective for large-scale, complex assemblages, but it begs questions about who is inside or outside of an assemblage—who is the maker and who is its target when algorithms dynamically adapt to the users they encounter?

Should users be held partly accountable for an algorithm's output if they knowingly provided it with data? A virtue model seems promising since it questions the seemingly idiosyncratic sociotechnical dynamics of assemblages—seeing each as a particular ethical arrangement—but this approach is difficult to scale in the context of fast-moving, algorithmic assemblages with myriad, unseen code, actors, and norms. A combination of all three approaches is likely needed.

My aim has been to show that even though algorithms are unstable objects of study, their ethics might still be investigated systematically by redescribing an ethical framework in terms of traceable, operationalized concepts and then looking for evidence of such concepts among the elements of algorithmic assemblages. This approach does not require—but nor does it eschew—code transparency. Seeing inside a black box is sometimes necessary, but never sufficient, for holding an algorithmic assemblage accountable. Rather, this framework focuses on the pragmatic question of how an entire assemblage *acts*. Its code may be transparent, its designers may have good intentions, and its institution may be well regulated, but an algorithmic assemblage might only be considered ethical if some combination of its means, ends, and virtues helps "us get into satisfactory relation with other parts of our experience" (James 1997, 100).

While this might seem like a hedge or ethical relativism—what does "satisfactory" mean, which parts, and are all experiences to be considered equally valid?—this approach is meant to connect the lived, relational dynamics of algorithmic assemblages (code, practices, norms) to an operationalized conception of ethics (convening, probability, time) so that any approach to accountability might answer the question: how are groups, similarities, and time lines governed by algorithmic assemblages creating (un)satisfactory relations? This is an argument against equating the ethics of algorithmic assemblages with the transparency of algorithmic code—an argument *for* a more expansive model of algorithmic ethics, taking up Dewey's (1891, 196) observation that "to do truly is to regard the whole situation as far as one sees it, and to see it as far as one can."

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

[1] For example, if the time, **T**, an algorithm requires to work on a data set of size ***n*** is **2*n***, then the time required to complete the algorithm increases *linearly* with the size of the data set (the algorithm is said to have linear big-O time, written as **T(*n*)=O(*n*)**).

References

Anderson, B. 1983. *Imagined Communities*. Revised ed. London, UK: Verso.

Anderson, C. W. 2011. "Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of Its Public in an Age of Audience Transparency." *International Journal of Communication* 5:19. Accessed September 8, 2015. http://ijoc.org/index.php/ijoc/article/view/884.

Arnuk, S. L., and J. C. Saluzzi. 2012. *Broken Markets: How High Frequency Trading and Predatory Practices on Wall Street Are Destroying Investor Confidence and Your Portfolio*. New York: FT Press.

Asur, S., B. A. Huberman, G. Szabo, and C. Wang. 2011. "Trends in Social Media: Persistence and Decay." Paper presented at the AAAI Conference on Weblogs and Social Media, Association for the Advancement of Artificial Intelligence, Barcelona, Spain, July 17–21, 2011.

Baker, P., and A. Potts. 2013. "'Why Do White People Have Thin Lips?' Google and the Perpetuation of Stereotypes via Auto-complete Search Forms." *Critical Discourse Studies* 10 (2): 187-204. doi:10.1080/17405904.2012.744320.

Barry-Jester, A. M., B. Casselman, and D. Goldstein. 2015. "The New Science of Sentencing." *The Marshall Project*. Accessed August 10, 2015. https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing.

Beam, M. A. 2014. "Automating the News: How Personalized News Recommender System Design Choices Impact News Reception." *Communication Research* 41 (8): 1019-41. doi:10.1177/0093650213497979.

Beer, D. 2009. "Power through the Algorithm? Participatory Web Cultures and the Technological Unconscious." *New Media and Society* 11 (6): 985-1002.

Beer, D., and R. Burrows. 2013. "Popular Culture, Digital Archives and the New Social Life of Data." *Theory, Culture and Society* 30 (4): 47-71.

Benjamin, S. M. 2013. "Algorithms and Speech." *University of Pennsylvania Law Review* 161 (6): 1445-94.

Berg, N. 2014. "Predicting Crime, LAPD-style." *The Guardian*. Accessed August 20, 2014. http://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report.

Bijker, W. E. 1995. *Conclusion: The Politics of Sociotechnical Change of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA: The MIT Press.

Bowker, G. C., and S. L. Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press.

Brunton, F., and H. Nissenbaum. 2011. "Vernacular Resistance to Data Collection and Analysis: A Political Theory of Obfuscation." *First Monday* 16 (5): 1-16.

Bucher, T. 2012. "Want to be on Top? Algorithmic Power and the Threat of Invisibility on Facebook." *New Media and Society* 14 (7): 1164-80.

Calo, M. R. 2015. "Robotics and the New Cyberlaw." *California Law Review* 103 (4): 101-46.

Carey, J. W. 1989. *Communication as Culture: Essays on Media and Society*. New York: Routledge.

Carlson, M. 2015. "The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority." *Digital Journalism* 3 (3): 416-31. doi:10.1080/21670811.2014.976412.

Chadwick, A. 2013. *The Hybrid Media System: Politics and Power*. Oxford, UK: Oxford University Press.

Cheney-Lippold, J. 2011. "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control." *Theory, Culture and Society* 28 (6): 164-81.

Christians, C. G., T. L. Glasser, D. McQuail, K. Nordenstreng, and R. A. White. 2009. *Normative Theories of the Media*. Urbana: University of Illinois Press.

Computer Ethics Institute. 2011. *The Ten Commandments of Computer Ethics.* Accessed May 1, 2013. http://cpsr.org/issues/ethics/cei/.

Cummings, M. L. 2006. "Automation and Accountability in Decision Support System Interface Design." *The Journal of Technology Studies* 32 (1): 23-31.

Dahl, R. A., and E. R. Tufte. 1973. *Size and Democracy*. Stanford, CA: Stanford University Press.

DeLanda, M. 2006. *A New Philosophy of Society: Assemblage Theory and Social Complexity*. New York: Bloomsbury Academic.

Dewey, J. 1891. "Moral Theory and Practice." *International Journal of Ethics* 1 (2): 186-203.

Dewey, J. 1954. *The Public and Its Problems*. New York: Swallow Press.

Diakopoulos, N. 2015. "Algorithmic Accountability: Journalistic Investigation of Computational Power Structures." *Digital Journalism* 3 (3): 1-18. doi:10.1080/21670811.2014.976411.

DiMaggio, P. J., and W. W. Powell. 1991. "Introduction." In *The New Institutionalism in Organizational Analysis*, edited by W. W. Powell and P. J. DiMaggio, 1-38. Chicago: The University of Chicago Press.

Driscoll, K., and S. Walker. 2014. “Working within a Black Box: Transparency in the Collection and Production of Big Twitter Data.” *International Journal of Communication* 8:20. Accessed September 8, 2015. http://ijoc.org/index.php/ijoc/article/view/2171.

Durkheim, E. (1912) 1954. *The Elementary Forms of the Religious Life*. Translated by J. W. Swain. Glencoe, IL: Free Press.

Edelman, B. 2011. “Bias in Search Results? Diagnosis and Response.” *The Indian Journal of Law and Technology* 7 (1): 16-32.

Facebook. 2013. “News Feed FYI: A Window into News Feed.” *Facebook*. Accessed April 1, 2014. https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed.

Flanagan, M., D. Howe, and H. Nissenbaum. 2008. “Embodying Values in Technology: Theory and Practice.” In *Information Technology and Moral Philosophy*, edited by J. van den Hoven and J. Weckert, 322-53. Cambridge, UK: Cambridge University Press.

Friedman, B., P. H. Kahn, and A. Borning. 2006. “Value Sensitive Design and Information Systems.” In *Human-computer Interaction in Management Information Systems: Foundations*, edited by P. Zhang and D. Galletta, 348-72. London, UK: M.E. Sharpe.

Galloway, A. 2011. “Are Some Things Unrepresentable?” *Theory, Culture and Society* 28 (7-8): 85-102.

Gannes, L. 2013. “Nearly a Decade Later, the Autocomplete Origin Story: Kevin Gibbs and Google Suggest.” *All Things D*. Accessed January 29, 2014. http://allthingsd.com/20130823/nearly-a-decade-later-the-autocomplete-origin-story-kevin-gibbs-and-google-suggest/.

Garber, M. 2013. “How Google’s Autocomplete was … Created / Invented / Born.” *The Atlantic*. Accessed March 3, 2014. http://www.theatlantic.com/technology/archive/2013/08/how-googles-autocomplete-was-created-invented-born/278991/.

Gerlitz, C., and A. Helmond. 2013. “The Like Economy: Social Buttons and the Data-intensive Web.” *New Media and Society* 15 (8): 1348-65. doi:10.1177/1461444812472322.

Gillespie, T. 2012. “Can an Algorithm be Wrong?” *Limn: Crowds and Clouds*. Accessed January 2, 2014. http://limn.it/can-an-algorithm-be-wrong/.

Gillespie, T. 2014. “The Relevance of Algorithms.” In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by T. Gillespie, P. Boczkowski, and K. A. Foot, 167-94. Cambridge, MA: MIT Press.

Glasser, T. L. 1989. “Three Views on Accountability.” In *Media Freedom and Accountability*, edited by E. E. Dennis, D. M. Gillmor, and T. L. Glasser, 179-93. New York: Praeger.

Goffey, A. 2008. "Algorithm." In *Software Studies: A Lexicon*, edited by M. Fuller, 15-20. Cambridge, MA: MIT Press.

Google. n.d. *How Search Works: From Algorithms to Answers*. Accessed January 6, 2014. http://www.google.com/insidesearch/howsearchworks/thestory/.

Hacking, I. 2006. *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*. Cambridge, UK: Cambridge University Press.

Herbst, S. 1995. *Numbered Voices: How Opinion Polling Has Shaped American Politics*. Chicago: The University of Chicago Press.

Hofstadter, D., and E. Sander. 2013. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. New York: Basic Books.

Hursthouse, R. 1999. *On Virtue Ethics*. Oxford, UK: Oxford University Press.

Igo, S. 2007. *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*. Cambridge, MA: Harvard University Press.

Internet Archive. 2001. *Removing Documents from the Wayback Machine.* Accessed March 1, 2014. https://archive.org/about/exclude.php.

James, W. 1997. "What Pragmatism Means." In *Pragmatism: A Reader*, edited by L. Menand, 93-111. New York: Random House.

Jasanoff, S. 2010. "Beyond Calculation: A Democratic Response to Risk." In *Disaster and the Politics of Intervention*, edited by A. Lakoff, 14-41. New York: Columbia University Press.

Jernigan, C., and B. F. T. Mistree. 2009. "Gaydar: Facebook Friendships Expose Sexual Orientation." *First Monday*. Accessed September 4, 2015. http://firstmonday.org/article/view/2611/2302.

Kant, I. [1785] 2002. *Groundwork for the Metaphysics of Morals*. Translated by A. W. Wood. Binghamton, NY: Vail-Ballou Press.

Kleinnijenhuis, J., F. Schultz, D. Oegema, and W. van Atteveldt. 2013. "Financial News and Market Panics in the Age of High-frequency Sentiment Trading Algorithms." *Journalism* 14 (2): 271-91. doi:10.1177/1464884912468375.

Kramer, A. D. I., J. E. Guillory, and J. T. Hancock. 2014. "Experimental Evidence of Massive-scale Emotional Contagion through Social Networks." *PNAS* 111 (24): 8788-90. doi:10.1073/pnas.1320040111.

Kushner, S. 2013. "The Freelance Translation Machine: Algorithmic Culture and the Invisible Industry." *New Media and Society* 15 (8): 1241-58. doi:10.1177/1461444812469597.

Latour, B. 2004. "Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern." *Critical Inquiry* 30 (2): 225-48.

Latour, B. 2005. *Reassembling the Social: An Introduction to Actor-network-theory*. Oxford, UK: Oxford University Press.

Lehmann, J., C. Castillo, M. Lalmas, and E. Zuckerman. 2013. "Transient News Crowds in Social Media." Paper presented at the ICWSM 2013. Accessed September 4, 2015. http://chato.cl/papers/lehmann_castillo_lalmas_zuckerman_2013_transient_news_crowds.pdf.

Lerman, J. (2013). "Big Data and Its Exclusions." *Stanford Law Review*. Accessed September 4, 2015. http://www.stanfordlawreview.org/online/privacy-and-big-data/big-data-and-its-exclusions.

Levine, C. 2007. "Analyzing Pandora's Box: The History of Bioethics." In *The Ethics of Bioethics: Mapping the Moral Landscape*, edited by L. A. Eckenwiler and F. G. Cohn, 3-23. Baltimore, MD: Johns Hopkins University Press.

Lin, T. C. W. 2013. "The New Investor." *UCLA Law Review* 60 (3): 678-735.

Linden, G., B. Smith, and J. York. 2003. "Amazon.com Recommendations: Item-to-item Collaborative Filtering." *IEEE Internet Computing* 7 (1): 76-80.

Mackenzie, A. 2015. "The Production of Prediction: What Does Machine Learning Want?" *European Journal of Cultural Studies* 18 (4-5): 429-45. doi:10.1177/1367549415577384.

Mager, A. 2012. "Algorithmic Ideology." *Information, Communication and Society* 15 (5): 769-87.

Marres, N. 2007. "The Issues Deserve More Credit: Pragmatist Contributions to the Study of Public Involvement in Controversy." *Social Studies of Science* 37 (5): 759-80.

Marvin, C. 1990. *When Old Technologies Were New: Thinking About Electric Communication in the Late Nineteenth Century*. Reprint ed. New York: Oxford University Press.

Maurer, B., T. C. Nelms, and L. Swartz. 2013. "When Perhaps the Real Problem Is Money Itself!: The Practical Materiality of Bitcoin." *Social Semiotics* 23 (2): 261-77. doi:10.1080/10350330.2013.777594.

Mayer, J., and P. Mutchler. 2014. "MetaPhone: The Sensitivity of Telephone Metadata." *Web Policy*. Accessed April 1, 2014. http://webpolicy.org/2014/03/12/metaphone-the-sensitivity-of-telephone-metadata/.

McKelvey, F. 2014. "Algorithmic Media Need Democratic Methods: Why Publics Matter to Digital Media Research." *Canadian Journal of Communication* 39 (4): 597-613.

McQuail, D. 2003. *Media Accountability and Freedom of Publication*. Oxford, UK: Oxford University Press.

Merrill, J. C. 2011. "Theoretical Foundations for Media Ethics." In *Controversies in Media Ethics*, 3rd ed., edited by A. D. Gordon, J. M. Kittross, J. C. C. Merrill, W. Babcock, and M. Dorsher, 3-32. New York: Routledge.

Mori, M. 1970. "The Uncanny Valley." *Energy* 7 (4): 33-35.

Morrison, K. 2014. "The NYPD Will Use Twitter to Predict Street Crime." *Social Times*. Accessed December 1, 2014.https://socialtimes.com/nypd-twitter-predict-crime_b147775.

Napoli, P. M. 2014. "Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption." *Communication Theory* 24 (3): 340-360. doi:10.1111/comt.12039.

Nguyen, A., J. Yosinski, and J. Clune. 2014. "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images." *arXiv - Computer Vision and Pattern Recongition*. Accessed December 12, 2014. http://arxiv.org/abs/1412.1897

Nielsen, R. K., and K. C. Schrøderb. 2014. "The Relative Importance of Social Media for Accessing, Finding, and Engaging with News: An Eight-country Cross-media Comparison." *Digital Journalism* 2 (4): 472-89. doi:10.1080/21670811.2013.872420.

Orlikowski, W. 2010. "Technology and Organization: Contingency All the Way Down." *Research in The Sociology of Organizations* 29: 239-46. doi:10.1108/S0733-558X(2010)0000029017.

Pasquale, F. 2011. "Restoring Transparency to Automated Authority." *Journal on Telecommunications and High Technology Law* 9 (235): 235-54.

Pinch, T. J., and W. E. Bijker. 1984. "The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." *Social Studies of Science* 14 (3): 399-441.

Rosenblum, N. L. 2000. *Membership and Morals*. Princeton, NJ: Princeton University Press.

Salmon, C. T., and T. L. Glasser. 1995. "The Politics of Polling and the Limits of Consent." In *Public Opinion and the Communication of Consent*, edited by T. L. Glasser and C. T. Salmon, 437-58. New York: The Guilford Press.

Sandvig, C., K. Hamilton, K. Karahalios, and C. Langbort. 2014. "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms." Paper presented at the Data and Discrimination: Converting Critical Concerns into Productive: A preconference at the 64th Annual Meeting of the International Communication Association, Seattle, WA.

Schiller, D. 1979. "An Historical Approach to Objectivity and Professionalism in American News Reporting." *Journal of Communication* 29 (4): 46-57.

Schudson, M. 1978. *The Ideal of Objectivity Discovering the News: A Social History of American Newspapers*. New York: Basic Books.

Schudson, M. 1986. "Deadlines, Datelines, and History." In *Reading News*, edited by R. K. Manoff and M. Schudson, 79-108. New York: Pantheon Books.

Seaver, N. 2014. “On Reverse Engineering: Looking for the Cultural Work of Engineers.” *Medium*. Accessed April 3, 2014. https://medium.com/anthropology-and-algorithms/d9f5bae87812.

Shilton, K. 2012. “Value Levers: Building Ethics into Design.” *Science, Technology, and Human Values* 38 (3): 374-97.

Skiena, S. S. 1998. *The Algorithm Design Manual*. Berlin, Germany: Springer.

Soltani, A., and B. Gellman. 2013. “New Documents Show How the NSA Infers Relationships Based on Mobile Location Data.” *The Washington Post*. Accessed April 2, 2014. http://www.washingtonpost.com/blogs/the-switch/wp/2013/12/10/new-documents-show-how-the-nsa-infers-relationships-based-on-mobile-location-data/.

Soltani, A., A. Peterson, and B. Gellman. 2013. “NSA Uses Google Cookies to Pinpoint Targets for Hacking.” *The Washington Post*. Accessed January 2, 2014. www.washingtonpost.com/blogs/the-switch/wp/2013/12/10/nsa-uses-google-cookies-to-pinpoint-targets-for-hacking/.

Star, S. L., and K. Ruhleder. 1996. “Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces.” *Information Systems Research* 7 (1): 111-34.

Steen, M. 2014. “Upon Opening the Black Box and Finding it Full: Exploring the Ethics in Design Practices.” *Science, Technology, and Human Values* 40 (3): 389-420. doi:10.1177/0162243914547645.

Striphas, T. 2015. “Algorithmic Culture.” *European Journal of Cultural Studies* 18 (4-5): 395-412. doi:10.1177/1367549415577392.

Suchman, L. 1994. “Do Categories Have Politics?” *Computer Supported Cooperative Work* 2 (3): 177-94.

Sullivan, M. 2015. “Facebook Patents Technology to Help Lenders Discriminate Against Borrowers Based on Social Connections.” *Venture Beat*. Accessed August 6, 2015. http://venturebeat.com/2015/08/04/facebook-patents-technology-to-help-lenders-discriminate-against-borrowers-based-on-social-connections/.

Sweeney, L. 2013. “Discrimination in Online Ad Delivery.” *Communications of the ACM* 56 (5): 44-54. doi:10.1145/2447976.2447990.

Thelwall, M., and L. Vaughan. 2004. “A Fair History of the Web? Examining Country Balance in the Internet Archive.” *Library and Information Science Research* 26 (2): 162-76.

The New York Times. 2014. *Robots.txt.* Accessed April 15, 2014. http://www.nytimes.com/robots.txt.

Twitter. 2014. “FAQs About Trends on Twitter.” *Twitter Help Center*. Accessed April 20, 2014. https://support.twitter.com/articles/101125-faqs-about-trends-on-twitter#.

Wallach, W., and C. Allen. 2008. *Moral Machines: Teaching Robots Right from Wrong*. Oxford, UK: Oxford University Press.

The Washington Post. 2014. *Robots.txt.* Accessed April 15, 2014. http://www.washingtonpost.com/robots.txt.

Webster, J. G. 2010. "User Information Regimes: How Social Media Shape Patterns of Consumption." *Northwestern University Law Review* 104 (2): 593-612.

Winner, L. 1986. *The Whale and the Reactor*. Chicago: Chicago University Press.

Winner, L. 1993. "Upon Opening the Black Box and Finding it Empty: Social Constructivism and the Philosophy of Technology." *Science, Technology, and Human Values* 18 (3): 362-78.

Wohn, D. Y., and C. Sarkar. 2014. "The Uncanny Valley Effect in Behavioral Targeting and Information Processing of Peripheral Cues." Paper presented at the iConference 2014, iSchools Organization, Berlin, Germany, March 4–7, 2014.

Author Biography

Mike Ananny (PhD, Stanford) is an assistant professor at the University of Southern California's Annenberg School for Communication & Journalism and an affiliated faculty with USC's Science, Technology, and Society cluster. He studies how the design and sociotechnical dynamics of networked news systems encode normative theories of the press.

Chris Anderson

The End of Theory:
The Data Deluge Makes the Scientific Method Obsolete

Wired Magazine, June 23, 2008

So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies like Google, which have grown up in an era of massively abundant data, don't have to settle for wrong models. Indeed, they don't have to settle for models at all.

Sixty years ago, digital computers made information readable. Twenty years ago, the Internet made it reachable. Ten years ago, the first search engine crawlers made it a single database. Now Google and like-minded companies are sifting through the most measured age in history, treating this massive corpus as a laboratory of the human condition. They are the children of the Petabyte Age.

The Petabyte Age is different because more is different. Kilobytes were stored on floppy disks. Megabytes were stored on hard disks. Terabytes were stored in disk arrays. Petabytes are stored in the cloud. As we moved along that progression, we went from the folder analogy to the file cabinet analogy to the library analogy to — well, at petabytes we ran out of organizational analogies.

At the petabyte scale, information is not a matter of simple three- and four-dimensional taxonomy and order but of dimensionally agnostic statistics. It calls for an entirely different approach, one that requires us to lose the tether of data as something that can be visualized in its totality. It forces us to view data mathematically first and establish a context for it later. For instance, Google conquered the advertising world with nothing more than applied mathematics. It didn't pretend to know anything about the culture and conventions of advertising — it just assumed that better data, with better analytical tools, would win the day. And Google was right.

Google's founding philosophy is that we don't know why this page is better than that one: If the statistics of incoming links say it is, that's good enough. No semantic or causal analysis is required. That's why Google can translate languages without actually "knowing" them (given equal corpus data, Google can translate Klingon into Farsi as easily as it can translate French into German). And why it can match ads to content without any knowledge or assumptions about the ads or the content.

Speaking at the O'Reilly Emerging Technology Conference this past March, Peter Norvig, Google's research director, offered an update to George Box's maxim: "All models are wrong, and increasingly you can succeed without them."

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.

The big target here isn't advertising, though. It's science. The scientific method is built around testable hypotheses. These models, for the most part, are systems visualized in the minds of scientists. The models are then tested, and experiments confirm or falsify theoretical models of how the world works. This is the way science has worked for hundreds of years.

Scientists are trained to recognize that correlation is not causation, that no conclusions should be drawn simply on the basis of correlation between X and Y (it could just be a coincidence). Instead, you must understand the underlying mechanisms that connect the two. Once you have a model, you can connect the data sets with confidence. Data without a model is just noise.

But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete. Consider physics: Newtonian models were crude approximations of the truth (wrong at the atomic level, but still useful). A hundred years ago, statistically based quantum mechanics offered a better picture — but quantum mechanics is yet another model, and as such it, too, is flawed, no doubt a caricature of a more complex underlying reality. The reason physics has drifted into theoretical speculation about *n*-dimensional grand unified models over the past few decades (the "beautiful story" phase of a discipline starved of data) is that we don't know how to run the experiments that would falsify the hypotheses — the energies are too high, the accelerators too expensive, and so on.

Now biology is heading in the same direction. The models we were taught in school about "dominant" and "recessive" genes steering a strictly Mendelian process have turned out to be an even greater simplification of reality than Newton's laws. The discovery of gene-protein interactions and other aspects of epigenetics has challenged the view of DNA as destiny and even introduced evidence that environment can influence inheritable traits, something once considered a genetic impossibility.

In short, the more we learn about biology, the further we find ourselves from a model that can explain it.

There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

The best practical example of this is the shotgun gene sequencing by J. Craig Venter. Enabled by high-speed sequencers and supercomputers that statistically analyze the data they produce, Venter went from sequencing individual organisms to sequencing entire ecosystems. In 2003, he started sequencing much of the ocean, retracing the voyage of Captain Cook. And in 2005 he started sequencing the air. In the process, he discovered thousands of previously unknown species of bacteria and other life-forms.

If the words "discover a new species" call to mind Darwin and drawings of finches, you may be stuck in the old way of doing science. Venter can tell you almost nothing about the species he found. He doesn't know what they look like, how they live, or much of anything else about their morphology. He doesn't even have their entire genome. All he has is a statistical blip — a unique sequence that, being unlike any other sequence in the database, must represent a new species.

This sequence may correlate with other sequences that resemble those of species we do know more about. In that case, Venter can make some guesses about the animals — that they convert sunlight into energy in a particular way, or that they descended from a common ancestor. But besides that, he has no better model of this species than Google has of your MySpace page. It's just data. By analyzing it with Google-quality computing resources, though, Venter has advanced biology more than anyone else of his generation.

This kind of thinking is poised to go mainstream. In February, the National Science Foundation announced the Cluster Exploratory, a program that funds research designed to run on a large-scale distributed computing platform developed by Google and IBM in conjunction with six pilot universities. The cluster will consist of 1,600 processors, several terabytes of memory, and hundreds of terabytes of storage, along with the software, including IBM's Tivoli and open source versions of Google File System and MapReduce.[1] Early CluE projects will include simulations of the brain and the nervous system and other biological research that lies somewhere between wetware and software.

Learning to use a "computer" of this scale may be challenging. But the opportunity is great: The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.

There's no reason to cling to our old ways. It's time to ask: What can science learn from Google?

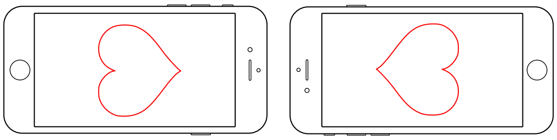
Chris Anderson (canderson@wired.com) *is the editor in chief of* Wired.

The Fibreculture Journal

DIGITAL MEDIA + NETWORKS + TRANSDISCIPLINARY CRITIQUE

issn: 1449-1443

issue 25: Apps and Affect.



FCJ–187 The Droning of Experience.

doi: 10.15307/fcj.25.187.2015

Mark Andrejevic
Pomona College.

Abstract:

The figure of the drone is invoked as a contemporary avatar for the logics of distributed networking at a distance, automated sense-making and automated response associated with interactive platforms more generally. The case of affective apps is forwarded as an example of the generalised logic of 'droning' in the register of control via the modulation of affect. The argument focuses on examples of data mining in the affective frontier: apps that generate data about affective states in order to more effectively categorise and target users. The conclusion argues for a reconfigured approach to questions of alienation as one way of critiquing drone infrastructures and the logics of droning.

Recent debates over the fate of automated weaponry raise the question of pre-empting pre-emption: might it be possible to thwart the seeming ineluctable development of so-called 'killer robots,' that can respond to perceived threats more efficiently and rapidly than humans? The processes of disarmament and pre-emption collided in the 'bold action' of a top United Nations official who issued a call to ban the ominously-acronymed Lethal Autonomous Weapons (LAWs). 'You have the opportunity to take pre-emptive action and ensure that the ultimate decision to end life remains firmly under human control,' UN Director-General Michael

Moller told the participants in a 2014 conferenced on killer robots (Agence France Presse, 2014). The difference between LAWs and other lethal weapons lies in the command decision – that is, the final determination regarding whether to fire (bomb, destroy, etc.). If the command decision always incorporates a human at some point in the command chain, the promise of the LAW is to codify priorities so that the human element can be programmed in advance (and thereby bypassed). Suggestively, automated application has long been a fantasy of the law – that is, the prospect that a law might carry within itself the principle of its application, thereby obviating the need for the all-too-human category of judgement. LAWs, in a sense, literalise the fantasy of automated application, by trying, sentencing, condemning and executing all at once.

In more general terms, the promise of LAWs recapitulates that of frictionless automation in which the resistance that slows down decision-making processes takes the forms of humans themselves. We, in all our sluggish, fleshly, humanity, are gumming up the machine by preventing it from operating as efficiently at it might otherwise do, freed from the vagaries of our desires and the hesitations of our decisions. The fantasy of friction-free capitalism outlined by Bill Gates (1995), for example, is one in which intelligent 'agents' speed up the consumption process, seeking out information about products, prices, and eventually about human desire so that it can be filled automatically. This same fantasy underwrites current developments in predictive analytics designed to distribute goods to particular locations before they have been ordered, to know what consumers want better than they themselves do. The prospect of LAWs envisions something similar – a process of automated warfare that can take place in an ongoing fashion at a pace that outstrips the limitations of human command and control. The friction-free conceit behind a LAW is that it can 'outperform and outthink a human operator' (Foust, 2013). As one university researcher put it in what sounds like a parody of contemporary Gradgrindianism:

> *If a drone's system is sophisticated enough, it could be less emotional, more selective and able to provide force in a way that achieves a tactical objective with the least harm… A lethal autonomous robot can aim better, target better, select better, and in general be a better asset with the linked ISR [intelligence, surveillance, and reconnaissance] packages it can run. (Foust, 2013)*

The same logic can be turned around on humans themselves through the process of what might be described as self-droning: finding ways to transform humans into networked, sensing devices. Consider, for example, the HSARPA-funded cortically coupled computer vision system that seeks to make human image scanners more efficient by tracking brain responses in real time. The goal is to make intelligence analysts, among others (including shoppers, of course), more efficient by bypassing the need for conscious recognition.

The program's lead researcher, Paul Sajda, claims to be able to show images of drone footage or surveillance satellite photos to analysts more rapidly than they can consciously process in order to use their brains, hooked up to EEG monitors, as a detection device. The resulting technology, according to researchers, can at least triple search speeds. Sajda describes it this way: 'The system latches on to individual perceptions and trains the computer to know what the user means by "interesting"' (Daley, et al, 2011). Building on this research, The U.S. Army is reportedly interested in creating a direct interface from drivers' brains to automated forms of reaction and response.

> *A driver might see something peculiar on the roadside. Maybe it is an improvised explosive device. His C3Vision headgear would register the brain waves associated with the suspicious object and inject them into the vehicle's driving system. When the system sees other things out there that look similar, it would automatically evade them. Likewise, security guards might use such gear to spot suspicious activity on surveillance video. (Daley, et al, 2011)*

Related research explores the ability of such systems to improve response times in jet pilots: the construction of LAWs by other means.

Unsurprisingly, in our convergent world, the technology is also envisioned to have consumer applications: a miniaturised, wireless version of the device might be used to identify consumer items or even specialty shops that catch your fancy as you walk down a city street. 'Just a quick glance at a dress in a window, for instance, might elicit a neural firing pattern sufficient to register with the system. A program could then offer up nearby stores selling similar items or shops you might want to investigate' (Daley, et al, 2011). It sounds like a ready-made app for an EEG-equipped Google Glass: the promise to realise the fantasy that neuromarketers have been pushing: a direct feedback system routed through the affective register to bypass self-conscious thought altogether. If Bill Gates envisioned automated consumption via 'intelligent agents' that determined our tastes and shopped for us, the C3 system promises to turn us into our own intelligent agents by bypassing the forms of conscious reaction and deliberation that threaten to introduce 'friction' into the system.

The goal of aligning these examples with one another is to highlight a shared logic that coalesces around a version of experience that literalises the post-psychoanalytic disentanglement of language and desire. A particular version of the materialisation of desire – (its subtraction from the realm of language and therefore its 'post-humanisation') – fits neatly with the forms of monitoring and manipulation envisioned by the coming generation of affective applications and platforms. What model of experience corresponds to this reconfiguration and generalisation of desire? The work of Ian Bogost led me to this question in reverse largely through the attempt to discern what the appeal of the model of experience he proposes might be. He raises the relevant question in the subtitle of his 2012 book, Alien Phenomenology: Or What It's Like to Be a Thing. In a sense, being a thing is precisely what the C3 system starts to envision. Bogost proposes an object-neutral definition of experience under which we might subsume all forms of interaction in terms of an expression familiar to the denizens of the data mine: the monitoring of the 'exhaust' of things. As Bogost puts it, 'The experience of things can be characterized only by tracing the exhaust of their effects on the surrounding world' (2012: 100). That is, things can only experience other things by tracing their 'exhaust' – and their own experience is whatever reaction they might have to this exhaust, a reaction that generates further exhaust. We might describe this as the meta-datafication of everything, a sensor-based model of experience, insofar as anything that is, in any sense, impacted by anything else becomes in the broadest interpretation of the term, a sensor. I'm inclined to push this reframing a bit farther and call it a kind of drone experience, in part because of the agentic sense with which Bogost infuses this flattened-out concept of experience, in part because of his fascination with various imaging technologies, and in part because of the treatment of the object as a probe: the attempt to experience the experience of the object that motivates the analysis.

The drone model of experience invokes the notion of a sensor-database-algorithmic formation that might be summed up by using the figure of the drone broadly construed: not just in the form of a flying, weaponised, surveillance device, but as the combination of a distributed sensor equipped with automated data analysis and response capabilities. Discussions of 'big data,' 'data mining,' and new forms of monitoring and surveillance often emphasise the figure of the database: the place where the data is stored, rather than that of the infrastructure that makes data collection possible. In part this is because of the distributed and heterogeneous character of the various sensors that comprise the monitoring 'assemblage' – but in part it is because of what might be described as the turn away from infrastructure that has characterised the fascination with so-called 'immaterial' forms of activity. This turn is echoed in the rhetoric of immateriality that characterises discussions of the 'cloud' (in 'cloud computing') and cyberspace more generally. Such formulations are symptomatic of anti-infrastructural thought. The figure of the drone, by contrast, focuses attention back upon the interface device that serves as mediator for information collection, automated analysis, and automated response at a distance.

The underlying claim here is that one of the reasons the figure of the drone has so rapidly captured the popular and media imagination is that, in addition to reviving what might be described as the ballistic imaginary once associated with technological gadgetry (in the Popular-Science vision of personal jet packs and rocket-ships), it encapsulates the emerging logic of portable, always-on, distributed, ubiquitous, and automated information capture: the droning of experience and response. The promise of the drone as hyper-efficient information technology is four-fold: it extends, multiplies, and automates the reach of the senses or the sensors, it saturates the time and space in which sensing takes place (entire cities can be photographed 24-hours a day), it automates the sense-making process, and it automates response. In this regard, the figure of the drone, generalised, stands for that of the indefinitely expandable and distributable probe that foregrounds the seemingly inevitable logic of algorithmic decision-making. The model of the signature strike (directed toward targets that 'fit a profile' rather than uniquely identified targets – that is, named and identified individuals) is an increasingly familiar one in the realm of data mining generally – whether for the purposes of health care, surveillance, marketing, policing, or security. Identification takes a back seat to data analytics: one needn't know the name of a particular individual to target him or her, merely that he or she fits the target profile. This is why the category of Personally Identifiable Information is becoming an increasingly vexed one. Data analytics are subsumed and accounted for by the broader ensemble represented by drone logic, which unites sensing, analytics, and response. The figure of the drone, then, serves as icon of the (inter)face of new forms of monitoring, surveillance, and response: an exemplar of emerging forms of digital 'interactivity.'

It is with this broader conception of the drone in mind that we might approach the affective frontier of data collection and monitoring: the fascination with so-called mood monitoring and sentiment analysis. The hallmark of the drone as a material object is – like so many of the digital devices that have come to permeate the daily life of technologically saturated societies – its mobility and miniaturisation, that is, its anticipated efficiency as ubiquitous, always-on probe. We might use the notion of the signature strike and its analogue in target marketing as an example: identification falls by the wayside, as do those aspects of the legacy version of experience associated with accounts of intentionality, motivation, and desire in ways that recall Chris Anderson's paean to the power of big data: 'Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people [and things] do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity' (2008). Such logic, like the signature strike, isn't interested in biographical profiles and backstories, it does not deal in desires or motivations: it is post-narratival in the sense conjured up by Bogost as one of the virtues of Object Oriented Ontology: 'the abandonment of anthropocentric narrative coherence in favor of worldly detail' (2012: 42).

Experiencing the data flow becomes, necessarily, the job of various kinds of distributed objects. Perhaps this is the appeal of Bogost's theory in the digital era: the excavation of the forms of post-human experience that characterise automated data collection. The interest in capturing all available data – as exemplified by a fascination with open-ended, random lists – embraces what Bogost describes as ‘a general inscriptive strategy, one that uncovers the repleteness of units and their interobjectivity’ (2012: 38). He calls this process ontography: the writing of being, which ‘involves the revelation of object relationships without necessarily offering clarifying description of any kind’ (2012: 38). This formulation bears a certain resemblance to Anderson’s diagnosis of the ‘end of theory’ wherein data mining might generate actionable but emergent information that is both unpredictable and, inexplicable (in the sense that it neither needs nor generates an underlying explanatory model).

The defining attribute of the kind of ‘knowledge’ envisioned in Anderson’s Big Data manifesto is the process of emergence itself – the fact that data mining by definition generates un-model-able outcomes and thereby puts emergence to work. So we start to see the outlines of a particular form of so-called knowledge emerging: a post-comprehension, post-referential (in the sense of referring to an underling cause or explanation), data-exhaust driven way of ‘knowing.’ It is with this in mind that I want to turn to the theme of emotion and relate it to the non-human version of materiality outlined by, for example, Jane Bennett (2009). She describes a version of affect that is, ‘not only not fully susceptible to rational analysis or linguistic representation but that is also not specific to humans, organisms, or even to bodies: the affect of technologies, winds, vegetables, minerals…’ (61). This is a version of affect that manifests itself in the form of exhaust described by Bogost and that, I think, lends itself to emerging, data-driven strategies of post-narratival analysis: tracing the exhaust of an unfolding litany of actants and interactants: of complex webs of interactions ‘too big to know,’ as David Weinberger (2011) puts it.

Bennett’s (2009) version of the endless unfolding of material detail surely expands beyond the realm of narrative containment – the ongoing chain of connections toward which her account gestures is both breathtaking and frustrating. Any outcome is the result of potentially infinite array of agentive factors. Her (inadvertently) complementary gesture to the ‘end of theory’ manifesto is a post-theoretical fascination with a kind of infinite regression: the attempt to contain everything so as to eschew the ostensible evils of abstraction. The growing reach of the big data database and the breadth of Bennett’s ambition to take into account what she describes as ‘an interstitial field of non-personal, ahuman forces, flows, tendencies, and trajectories’ (61), share a conserved impulse toward totality, although Bennett retains the model of narrative closure while frustrating it utterly. The prospect of unfolding the full list of participants in a particular event or outcome is an ongoing one. Similarly, the database in its ideal-typical form approaches the levelling, allegedly democratising ambition of Bennett’s vibrant materialism, allowing a promiscuous jumble of factors to rub shoulders.

With these affinities in mind – between correlational forms of data mining and post-narrative, post-explanatory modes of analysis, the remainder of this article sets out to explore their relevance to the topics of affective computing and sentiment analysis and the role of so-called mood reading in the process of affective modulation (Clough, 2009). Consider some examples from the realm of mood-mining as one frontier of data collection in the service of so called affective computing. Microsoft's 'MoodScope' initiative seeks to turn smart phones into mood sensors, not by adding a dedicated sensor, but by tracking usage patterns and their correlation with self-reported mood. By correcting their models over time, the researchers eventually automate the prediction process and claim to move from 63 percent accuracy to 93 percent accuracy (LiKimWa, 2012: 1). As the project's researchers put it, 'we find smart phone usage correlates well with users' moods…Users use different applications and communicate with different people depending on their moods. Using only six pieces of information, SMS email, phone call, application usage, browsing and location, we can build robust statistical models to estimate mood' (LiKimWa, 2012: 2). Of course, the goal of inferring mood is, for Microsoft a commercial one that serves the generation of recommendation algorithms and marketing strategies that monitor and influence shifting consumer preferences.

We might describe the MoodScope as partaking of drone logic (and drone experience): it envisions a network of mobile, distributed, always-on sensors that underwrite automated forms of data collection, processing and response (targeting). The invocation of 'mood' should not distract – it is a placeholder that does not refer to an underlying state but simply to pattern of correlations: the nexus of a particular set of behaviours (as monitored by smart phone sensors) and the measured probability of a particular response. The next logical step for the development of such an app is to bypass the placeholder of 'mood' altogether, simply extrapolating from patterns of activity to predict susceptibility to particular prompts and appeals during particular times within specified contexts. This modality of prediction or influence operates at a machinic level, linking flows of activity to patterns of response in order to get something done (generate a response or action of some kind). The point is not interpretation (of mood, subjective state, evidence of desire) but intervention in flows of viewing, clicking, spending, consumption. This way of thinking lends itself to the machinic imaginary of scholars such as, for example William Bogard (1998), who quoting Deleuze and Guattari, notes that, 'The social machine… is literally a machine, irrespective of any metaphor, inasmuch as it exhibits an immobile motor and undertakes a variety of interventions: flows are set apart, elements are detached from a chain, and portions of the tasks to be performed are distributed' (54). The notion of an 'immobile motor' neatly invokes the figure of the 'exhaust' of things. The process of 'sensorisation' works to codify these flows for the purpose of intervention. As Daniel Smith (2007) puts it, these networks of affect (and the information networks through which they flow) become 'infrastructural': 'They are, if I can put it this way, part of the capitalist infrastructure; they are not simply your own individual mental or psychic reality. Nothing makes this more obvious that the effects of marketing, which are directed entirely at the manipulation of the drives and affects: at the drug store, I almost automatically reach for one brand of toothpaste rather than another' (74).

The infrastructure of affect continues to be 'built out' by the growing platform of affects apps. Apple has already patented technology that relies on an embedded tactile heartbeat sensor to identify users and monitor their moods (Calorielab, 2010). The technology combines the promise of convenience with enhanced monitoring capability: the phone can be unlocked just by picking it up, but the monitor, unlike a fingerprint scanner, simultaneously gathers information to potentially serve a host of marketing, security, and medical functions. As one news account put it, 'By monitoring your heartbeats, the device will also be able to tell how you're feeling (better than you can tell yourself, presumably), what you've been eating and if you've just come back from a jog' (Calorielab, 2010). The vectors for capturing, monitoring, and intervening in the flows that link 'mood' and response are multiple and expanding alongside the various registers of interactivity: they piggyback on multiplying applications and the behaviour patterns these elicit.

Similarly, the company that developed the technology that powers Apple's Siri is working on adding voice recognition ID systems that simultaneously incorporate mood detection. Soon Siri will respond not just to what you say, but to its conception of how you feel. Once again the promise combines convenience with the prospect, at least in this case, of commercial monitoring. As an interview with the company's marketing chief put it: 'If your car thinks you sound stressed, it may SMS your office to say you're late or even automatically suggest another route that avoids traffic' (Eaton 2012). But the company is looking to monetise the technology: 'What if when you ask Siri for information about a movie, she works out that you're sad and recommends a comedy film that you otherwise wouldn't have seen, paired with an ad campaign?'.

And the litany of mood apps goes on: MIT has spun off a company called Affdex that uses facial recognition technology to gauge emotional response. It has been used by companies like Forbes to crowdsource reader's responses to ads shown on the company's website. Yes, soon not just the TV, but the ads, the music, the magazines and books will be watching, analysing, and responding in the affective register. A company called Sensum develops apps that use galvanic skin response to measure stress levels. Microsoft is building emotion recognition into its Kinect device, so that next-generation games (and, yes, ads) will be able to react to facial expressions and monitor heart rate. The anticipated result is, as a somewhat breathless account puts it, that 'games will react to your emotionality, and even your cars will route you to entirely new destinations based on how you're feeling. The next generation of advertising will determine how you're feeling… And it's not just the question of detecting your mood, it's all about how this leads the person expressing the mood to discover new information' (Eaton, 2012). It also leads to the prospect of more effectively sorting, targeting, and influencing in a variety of registers for a range of purposes.

Coming full circle, security is one of the pioneering and recurring applications of affective monitoring, thanks in no small part to department of homeland security funding. The DHS, has funded Cambridge-based Draper Labs 'to develop computerized sensors capable of detecting a person's level of "malintent" – or intention to do harm' as part of the 'Future Attribute Screening Technologies,' program (Segura, 2009). The goal is to, 'detect subjects' bad intentions by monitoring their physiological characteristics, particularly those associated with fear and anxiety,' according to the DHS (Segura, 2009).

Possible technological features of FAST include 'a remote cardiovascular and respiratory sensor' to measure 'heart rate, heart rate variability, respiration rate, and respiratory sinus arrhythmia,' a 'remote eye tracker' that 'uses a camera and processing software to track the position and gaze of the eyes (and, in some instances, the entire head),' 'thermal cameras that provide detailed information on the changes in the thermal properties of the skin in the face,' and 'a high resolution video that allows for highly detailed images of the face and body … and an audio system for analyzing human voice for pitch change' (Segura, 2009). The project is based on another DHS project called 'Hostile Intent,' which 'aims to identify facial expressions, gait, blood pressure, pulse and perspiration rates that are characteristic of hostility or the desire to deceive' (Segura, 2009).

Researchers are developing applications that claim to be able to identify a person's emotional state by listening in on mobile phone conversations. Some companies in the United States already use the system in their call centres. Researchers are testing the software's use in diagnosing medical conditions like autism, schizophrenia, heart disease and even prostate cancer (DiscoveryNews, 2013). One could continue indefinitely in this register: since emotion detection covers the gamut of securitisation applications: economic, criminal, health, social and so on. And the sensor array proliferates on the various forms of drone devices, broadly construed, that circulate amongst us, upon us, with us.

It is just one step from these examples to what might be described to the redoubling of drone logic: equipping drones with 'malintent-detection' sensors. Drones already target strikes based on mobile phone signatures, using the device to identify a particular individual. But drone logic pushes beyond strategies of identification in which a device comes to represent a particular target to strategies of pre-emption in which a device identifies potentially threatening or risky affective states with the potential to result in action.

In this regard, the invocation of terms like mood, emotion, or sentiment (or even ‘malintent’) is not meant to speak to a particular conception of subjective interiority nor even to have any definitive discernable stable referential content, but rather to mark the intent of detecting, predicting, and influencing response in a register other than that of reflexive, self-conscious communication – indeed to, in a sense, bypass this register in any respect other than as a potential source of more raw material for pattern analysis. The promise of bypassing this register is to bypass the vagaries, pathologies, deceptions, and self-deceptions of self-consciousness: to read affective response directly and thereby to develop strategies for intervening in it. In this context, speech, to take an example, is not about content, but about voice stress, or detectable word patterns that correlate with signature patterns – as in a signature strike. That is, the strategies of influence mobilised in response to detected ‘emotional’ states may take the forms of standard types of communication, but the register in which their potential effectivity is posited is other than the ideological – the narrative, the content-based. In Papoulias and Callard’s (2010) formulation, the intervention, ‘is seen as proceeding directly from the body – and indeed between bodies – broadly construed here – without the interference or limitation of consciousness, or representation’ (37).

In her critique of the turn to affect, Ruth Leys characterises the split at work here in terms of the, ‘presumed separation between the affect system on the one hand and signification or meaning or intention on the other’ (2011a: 800). It is a presumption she is concerned about not least because it smuggles in the very binaries these theorists imaged they had surpassed: ‘in spite of their explicit hostility to dualism, many of the new affect theorists succumb to a false dichotomy between mind and body’ (2011a: 801). This dualism is characteristic of ‘post-comprehension’ strategies of influence and ‘literacy’ (brain reading and body reading). The ‘mind’ (intentional, conscious, available for rational cognition) may have gotten much of the attention when it comes to information processing and communication, but the body’s language is efficacious. As Leys puts it, affect is figured as ‘prior to ideology’: ‘an inhuman, nonsignifying force that operates below the threshold of intention, consciousness, and signification’ (2011a: 802).

The turn to affect in the strands of theory outlined earlier is thus framed as a (re)turn to the body as subsumed to the status of object with particular types of experience, that take into account what Thrift describes as, ‘the way that political attitudes and statements are partly conditioned by intense autonomic bodily reactions that do not simply reproduce the trace of a political intention and cannot be wholly recuperated within an ideological regime of truth' (as quoted in Leys, 2011b: 436). This model of affective communication as immediate influence is rehabilitated not least in the strategies of neuromarketers and the sentiment analysts (as is the temporal and conceptual split between affective response and post hoc rationalisation: the attempt to narrativise the impulse that always comes after the fact). Although data mining is agnostic about this split, allegedly eschewing models of causation and explanation, in this very refusal it has already chosen sides.

Something related takes place in the development of so-called sentiment analysis: the attempt to data mine expressed sentiment on the social web in real time so as to intervene and influence an aggregate conception of the internet's 'feeling tone.' The field is popularly described as one in which, 'the vagaries of human emotion are translated into hard data' (Wright, 2009). But this description is not quite right: the goal of marketers is not to gauge personal, individual 'human' emotion, but rather to probe an affective landscape without having to pore over the individual contributions of millions of Internet users. Sentiment analysis relies on technological advances that make it possible to sift through all these forms of expression, to treat them as measurements of a capability to affect or a susceptibility to influence, without actually reading them. The goal is a kind of thin-slicing or pulse-reading of the Internet as a whole. Pioneering companies in the field develop applications that troll through twitter feeds, blogs, social networking sites, online forums, bulletin boards, and chat rooms, probing the emotional pulse of the Internet. The industry places a premium on speed and volume: processing as many posts and messages as possible in real time.

As in the case of the app examples, the model is not a descriptive, referential one (that would aim to accurately describe how individuals are feeling) but a predictive, correlational one. Applied to sentiment analysis, the goal of data mining is both pre-emptive and productive: to minimise negative sentiment and maximise emotional investment and engagement: not merely to record sentiment as a given but to modulate it as a variable and thereby to influence the forms of behaviour with which these shifts are associated. The process relies on giving 'populations over to being a probe or sensor' (to borrow Patricia Clough's formulation) to provide the raw material for tracking the emotional indicators that correlate with desired outcomes – and for developing ways of exploiting them (Clough, 2009: 53).

What is suggestive about the proliferation of apps in the affective register is the way they redouble all content in the form of post-content 'knowledge.' Recall the goal of MoodScope or the FAST program: not to read all messages, or listen to all calls, but to piggyback on content to get machine-sortable metadata: you may use your apps or your email to collect information or communicate with others, but these uses generate patterns that, without your conscious knowledge communicate a user state (and an aggregate state) that can then be correlated with your responses. We might describe this monitoring logic as the meta-datafication of everything: content becomes metadata, when it is not read (for significance), but sorted, mined, and correlated (for useful patterns). This is why no human at Google reads your email. Such applications use the placeholder of mood, or affective state, to generate correlations that underwrite more direct modes of influence – techniques for enhancing the power of acting or being acted upon. That is, the goal is to define a state of receptivity in which the broadened and flattened conception of experience allows all kinds of collected data to commingle. The result is a litany of content – in its machine-readable form – including patterns of search, typing speed, Web sites visited, patterns of communication (who one emails, how frequently), movement throughout the course of the day, barometric pressure, sunspots (why not?), magnetic fields, and on and on, limited only by the capabilities of the growing sensor array. I am using the term post-comprehension, somewhat freely here, to designate the forms of too-big-to-know knowledge that represent the displacement of causation or explanation by correlation. The descriptor 'post-comprehension,' then refers to the goal of discerning patterns that are neither conscious nor meaningful to users. The term refers also to the detection of receptivity to particular influences – whether such and such a 'mood' – or, more properly speaking, the patterns of use which the placeholder of mood is meant to designate – correlates with a heightened tendency to respond in particular ways. Additionally the notion of post-comprehension refers to the fact that the generation of these patterns is portrayed as an emergent one, and is, in this respect, unmodellable, unanticipatible, and potentially, un-reverse-engineerable. Why post-comprehension and not pre-comprehension? Because the goal of explaining is not deferred but dispensed with: there is no background assumption that in the end, the infinite database will yield total comprehension. Once everything is coded, it is not understood, but simply processed: the ongoing interventions of the (total) immobile motor.

These forms of opacity, or unmodellability characterise the emerging asymmetries of a big data divide. From a research perspective, Boyd and Crawford (2012) have characterised the divide between 'the Big Data rich' (companies and universities that can generate or purchase and store large datasets) and the 'Big Data poor' (those excluded from access to the data, expertise, and processing power), highlighting the fact that a relatively small group with defined interests threatens to dominate and control the research agenda. The notion of a 'big data divide' needs to be extended to incorporate a distinction between ways of thinking about data and putting it to use. That is, it needs to acknowledge the consequences of emerging forms of opacity and asymmetry: between those who are able to put to use the unanticipatable and inexplicable correlations generated by the data mining process and those who are subject to the forms of sorting and exclusion they license. This is also a divide between those who seek to exploit detected correlational patterns of affective response and those whose actions are subject to the forms of inferential data mining enabled by the growing sensor array and the expanding database.

Despite the rhetoric of personalisation associated with data mining, it yields predictions that are probabilistic in character, privileging decision making at this level. Moreover, it ushers in the era of what might be called emergent social sorting: the ability to discern un-anticipatable patterns that can be used to make decisions that influence the life chances of individuals and groups. Notions like that of ‘informed consent’ when it comes to online tracking and other types of digital-era data surveillance are rendered largely meaningless by the logic of data mining, which proposes to reveal unanticipated and unpredictable patterns in the data. At a deeper level, the big data paradigm proposes a post-explanatory pragmatics (available only to the few) as superior to the forms of comprehension that digital media were supposed to make more accessible to a greater portion of the populace.

In this regard, the privileging of correlation and prediction – like the figure of the drone – leads us back to issues of infrastructure. If, as Weinberger (2011) puts it, the smartest person in the room is the room, in the era of post-comprehension knowledge, it matters who owns, operates, and controls the room. It is worth emphasising that such forms of asymmetry and opacity are the specific goal of so-called affective forms of context awareness. At the moment when access to traditional forms of understanding and evidence is enhanced by the new technology, these are treated as ostensibly outdated.

Practices of data-driven affect mining anticipate a context in which only the few will have access to useful forms of ‘knowledge’ that are not just unavailable to the majority, but incomprehensible. Thus, there is no way for individual users to anticipate how information about them might prove salient for particular forms of decision-making. Isn’t this the endgame logic of the ‘unmanned’ LAW? The figure of the drone augers not simply prosthetic enhancement but displacement: the cultivation of forms of automation that result not simply in synthetic perception (Virilio, p. 58), but in synthetic action. In this regard, the figure of the drone comes to stand for a particular kind of alienation: of perception and practice that is becoming increasingly familiar in our auto-sorted, curated, algorithmically directed information environment. We come to experience the re-processing of our actions, desires, and responses in an unrecognisable form directed back upon us in the service of ends built into the infrastructure. In the contemporary theoretical climate, the familiar critique of alienation (as a critical conceptual tool) is that it introduces an outdated form of (pre-post-) humanism (and thus, of the subject). When everything is alien, alienation, of course evaporates. What if the critique of alienation invokes, rather, the spectre of what Smith refers to as ‘an ethics of immanence’ that will criticise anything that ‘separates a mode of existence from its power of acting’ (2007, 68)? Rather than proposing the alien as a starting point, in the face of the developments outlined above, why not alienation? To invoke Guy Debord’s diatribe against Jean-Marie Domenach’s dismissal of the very concept of alienation: ‘Let us speak vulgarly since we’re dealing with priests: alienation is the point of departure for everything — providing that one departs from it’ (Situationist International, 1966).

Biographical Note

Mark Andrejevic teaches and writes about popular culture, surveillance, and digital media. He is the author of *Infoglut: How Two Much Information is Changing the Ways We Think and Know*, as well as two other books and a variety of articles and book chapters. He is currently writing about the droning of the social.

References

Agence France Presse. ‘UN talks take aim at “killer robots,”’ *The Express Tribune*, May 13, 2014. http://tribune.com.pk/story/707899/un-talks-take-aim-at-killer-robots/.

Anderson, Chris. ‘The End of Theory: The data Deluge Makes the Scientific Method Obsolete.’ *Wired Magazine*, June 23, 2008, http://www.wired.com/science/discoveries/magazine/16- 07/pb_theory. Accessed 30 August 2008/

Bennett, Jane. *Vibrant Matter: A Political Ecology of Things* (Durham: Duke University Press, 2009).

Bogard, William. ’Sense and Segmentarity: Some Markers of a Deleuzian□Guattarian Sociology.’ *Sociological Theory* 16.1 (1998): 52–74.

Bogost, Ian. *Alien Phenomenology, or, What it’s Like to be a Thing* (Minneapolis: University of Minnesota Press, 2012).

Boyd, Danah, and Crawford, Kate. ’Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon.’ *Information, Communication & Society* 15.5 (2012): 662–679.

Calorielab. ‘iPhone That Can Detect Your Mood in a Heartbeat’, 2010. http://calorielab.com/labnotes/20100510/iphone-to-recognize-mood-by-detecting-heartbeat/

Clough, Patricia Ticineto. ’The New Empiricism: Affect and Sociological Method.’ *European*

Journal of Social Theory 12.1 (2009): 43–61.

Daley, Jason, Piore, Adam, Lerner, Preston, and Svoboda, Elizabeth. ‘How to Fix Our Most Vexing Problems, From Mosquitoes to Potholes to Missing Corpses,’ *Discover Magazine* (October, 2011), http://discovermagazine.com/2011/oct/21-how-to-fix-problems-mosquitoes-potholes-corpses/

Eaton, Kit. ‘Does Your Phone Know How Happy You Are?’ FastCompany.com, June 7, 2012. http://www.fastcompany.com/1839275/does-your-phone-know-how-happy-you-are-emotion-recognition-industry-comes-giddily-age/

Foust, Joshua. ‘Soon, Drones May Be Able to Make Lethal Decisions on Their Own,’ *National Journal* (October 8, 2013), http://www.nationaljournal.com/national-security/soon-drones-may-be-able-to-make-lethal-decisions-on-their-own–20131008/

Gates, Bill, Myhrvold, Nathan and Rinearson, Peter. *The Road Ahead* (New York: Penguin, 1995).

Leys, Ruth. ’Affect and Intention: A Reply to William E. Connolly.’ *Critical Inquiry* 37.4 (2011a): 799–805.

Leys, Ruth. ‘The Turn to Affect: A Critique.’ *Critical Inquiry* 37.3 (Spring 2011b):434–472.

LiKimWa, Robert. ‘MoodScope: Building a Mood Sensor from Smartphone Usage Patterns’ (Doctoral dissertation, Rice University, Houston, TX, 2012).

Papoulias, Constantina, and Callard, Felicity. ’Biology’s gift: Interrogating the turn to affect.’ *Body & Society* 16.1 (2010): 29–56.

Segura, Liliana. ‘Homeland Security Embarks on Big Brother Programs to Read Our Minds and Emotions,’ *Alternet* (December 8, 2009), http://www.alternet.org/story/144443/homeland_security_embarks_on_big_brother_programs_to_read_our_minds_and_emotions/

Smith, Daniel W. ’Deleuze and the question of desire: Toward an immanent theory of ethics.’ *Parrhesia* 2 (2007): 66–78.

Weinberger, David. *Too Big to Know: Rethinking Knowledge Now That the Facts Aren’t the Facts, Experts Are Everywhere, and the Smartest Person in the Room is the Room* (New York: Basic Books, 2011).

Wright, Alex. ‘Mining the Web for Feeling, not Facts’, *The New York Times* (August 23, 2009), http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html/

OPEN HUMANITIES PRESS

The Fibreculture Journal is published by The Fibreculture Journal Incorporated in partnership with Open Humanities Press.

AND

PHENOMENOLOGY OF THE END

Franco “Bifo” Berardi

<e>

Contents

Introduction

Concatenation, Conjunction, and Connection

> A rhizome has no beginning or end; it is always in the middle, between things, interbeing, *intermezzo*. The tree is filiation, but the rhizome is alliance, uniquely alliance. The tree imposes the verb "to be," but the fabric of the rhizome is the conjunction, "and … and … and …" This conjunction carries enough force to shake and uproot the verb "to be." [… And to] establish a logic of the AND, overthrow ontology, do away with foundations, nullify endings and beginnings.
>
> —Gilles Deleuze and Félix Guattari, *A Thousand Plateaus*

The Metaphor of the Rhizome

In a rhizome there is no beginning and no end, according to Deleuze and Guattari, who propose that we view reality as an infinite rhizome, that is, an open concatenation of ands: and ... and ... and ...

This is why I'm writing this phenomenology of the end.

There is no end. Some may take this assertion as a source of endless hope; others may take it as a source of endless despair.

Both would be on the wrong path.

Do not get me wrong. I don't pretend to know what is good or bad. I am not hopeful, but neither am I hopeless. Phenomenology is an infinite task, so the phenomenology of the end must also be an interminable task.

I decided to stop writing this book here because my life is not endless, and I am approaching the end. But even so, I know that I will not stop concatenating: and, and, and.

In 1977, in the year of the premonition, Deleuze and Guattari wrote a short text called *Rhizome*, later published as the introduction to *A Thousand Plateaus*.

That year, social movements, punk culture and the dystopian imagination of art and literature foreshadowed in many ways a mutation that we are now witnessing and living through, and that has infiltrated the technological environment, social relations, and culture.

The rhizome is simultaneously the announcement of a transformation of reality, and the premise to a new methodology of thought. It is a description of the chaotic deterritorialization that follows modern rationalism, as well as a methodology for the critical description of deterritorialized capitalism.

This short text by Deleuze and Guattari foretold both the dissolution of the political order inherited from modernity, and the vanishing of the rational foundations of Western philosophy. At the same time, it opened the way to a new methodology that adopted what I call concatenation, rather than dialectical opposition, as a model to conceptualize cultural processes and social transformations.

Decades after the publication of this text, the rhizomatic metaphor can be seen as a way of mapping the neoliberal process of globalization, and the precarization of labor that it entails. But it also refers to the interminability of the philosophical task. But does the philosopher even have a task? And what, in that case, is that task? To map the territory of the mutation, and to forge conceptual tools for orientation in its ever-changing, deterritorializing territory: such are the tasks for the philosopher of our times.

Diachronic and Synchronic Phenomenology

A rhizomatic methodology shapes my approach to the subject of this book: the phenomenology of sensibility in our present age of technocultural mutation.

I argue that the ongoing transition from the alphabetical to the digital infosphere marks a shift from the cognitive model of conjunctive concatenation to a model of connective concatenation.

This book is concerned with the effects of this shift in the fields of aesthetic sensibility and emotional sensitivity.

The shift I am referring to is diachronic. It occurs as a transition, extending over a span of several human generations, during which time it transforms cognitive patterns, social behavior, and psychological expectations. But there is also a synchronic frame in which this shift occurs. Investigating that frame will allow me to describe the composition, conflicts, and coevolution of different psychocultural regimes as they simultaneously approach each other, collide, and interweave through the process of globalization.

The first, diachronic, and temporal axis of the phenomenology of sensibility that I am introducing here is the transition from the mechanical to the digital order, and the effects of this transition in the psychosphere.

The second, synchronic, and spatial axis of this phenomenology of sensibility is the coevolution of different cultural regimes of subjectivation in a globalized world.

During the last thirty years, the shift from the mechanical to the digital technosphere has provoked a mutation in the texture of human experience, and in the fabric of the world itself. The conjunctive mode of social interaction, which was prevalent since the Neolithic revolution, has been rapidly replaced by a connective mode of interaction. The latter began to prevail when the automating interfaces of the information machine pervaded and innervated the linguistic sphere.

I will try to describe the transition from the age of industrial capitalism to the age of semiocapitalism from the point of view of a shift from conjunction to connection as the dominant mode of social interaction.

Both sensibility and sensitivity are affected by this shift, although the mutation takes different forms and intensities in different geo-cultural areas of the world. I will thus trace the general lines of its aesthetic genealogy.

Sensibility will be my main concern: in these pages, I propose to draw a phenomenological map of the global mutation, investigating the aesthetic and the emotional side of sensibility.

For this purpose, I will trace the effects of the shift from the conjunctive to the connective mode in different geo-cultures.[1]

I must add that this research does not pretend to exhaustivity, as we know from Husserl that "phenomenology is an infinite task."

Sensibility and Creation

Emotion is a concatenation of unconnected things, events, and perceptions. But, we might ask, how is a concatenation possible between things that have no connection? Are there filters and grids that make the human organism sensitive to the color of autumn leaves, to the tenderness of a gesture, or to the sound of a song? Are the parts that enter into a concatenation fragments of a mosaic whose unity has been lost? Should we perhaps reconstruct the design to which the fragments once belonged? Or should we instead avoid presupposing a pre-existing design wherein segments are integrated and meaningful?

A conjunctive concatenation does not imply an original design that must be restored. A conjunction is a creative act; it creates an infinite number of constellations that do not follow the lines of a pre-conceived pattern, or an embedded program.

There is no design to fulfill at the beginning of the act of conjunction. Neither is there a model at the origin of the process of the emergence of form. Beauty does not correspond to a hidden harmony embedded in a universal spirit or in the mind of god. There is no code to comply with.

On the contrary, conjunctive concatenation is a source of singularity: it is an event, not a structure, and it is unrepeatable because it happens in a unique point in the network of space and time.

> The more we study the nature of time, the more we shall comprehend that duration means invention, the creation of forms, the continual elaboration of the absolutely new.[2]

According to Bergson, "we perceive duration as a stream against which we cannot go," a stream whose current we cannot move back up, and in this stream, new configurations of being arise out of nothing at every instant.

Sensibility is the faculty that makes it possible to find a path that does not yet exist, a link between things that have no intrinsic or logical implication. Sensibility is the sense-driven creation of conjunctions, and the ability to perceive the meaning of shapes once they have emerged from chaos. This does not happen by way of recognition, in the sense that such forms would be compatible with others that we would have seen before. It occurs because we perceive their aesthetic correspondence, their accordance, and conformity with the expectations of the conscious, sensitive, and sensible organism.

Expectations are crucial for the act of aesthetic conjunction, and for both the perception and the projection of forms. Such expectations are formed in the sphere of culture, which has a temporal history and a geographic location: what I call geo-cultures anchored in the flow of time. There is no implicit logic bringing together one sign with another, and their composition does not aim to arrive at an isomorphism with the world. The part is not completed through a conjunction with another part, nor do parts put side by side give life to a totality.

The only criterion of truth is the pleasure of the conjunction: you and I, this and that, the wasp and the orchid.

The conjunction is the pleasure of becoming other, and the adventure of knowledge is born out of that pleasure.

The problem is: How does it happen that under certain circumstances conjoined signs give birth to meaning? How does it happen that under certain circumstances conjoined events become history? And conjoined percepts become reality? Witold Gombrowicz suggests that reality is the effect of obsession.[3]

Gregory Bateson suggests that the skin is the line of conjunction and the sensible interface par excellence.[4] Forms are evoked and conjured within the aesthetic sphere. But what does *aesthetic* mean? By the word aesthetic, Bateson refers to everything that belongs to the sphere of sensibility. The latter is not the space where conjunction is recorded; instead, it is the factory of conjunctions. These do not happen somewhere in the world, they happen in a sensible mind.

For Bateson, the question of truth must shift from the realm of metaphysics and history to the realm of biology and sensibility. The mind is able to think life because it belongs to the living world. It's a matter of co-extensivity, not of representation. There is no ontological correspondence between the mind and the world, as the metaphysicians would like to believe. There is no historical totalization in which mind and world would coincide. There is no correspondence, adjustment, or *aufhebung*-realization. There are only conjunctions.

(And connections, as we'll see. But this is another story.)

Reality could be described as the point of conjunction of innumerable psycho-cognitive projections. If the mind can process the world as an infinite set of co-evolving realities that act on one another, this is only because the mind is in the world. Language is the realm where man brings forth being, and language is the conjunction of artificial fragments (signs) that produce a meaningful whole. But meaning does not take place in a preexisting nature or reality that exists as such, independently, it only occurs in the concatenation of minds.

Mirror Neurons, Language, and Connective Abstraction

When it comes to connection, the conceptual frame changes completely. When I use the word *connection*, I mean the logical and necessary implication, or inter-functionality, between segments. Connection does not belong to the kingdom of nature; it is a product of the logical mind, and of the logical technology of mind.

Since this text is essentially concerned with the anthropological and aesthetic effects of the shift from the sphere of conjunction to that of connection, I will return later to the distinction between conjunction and connection.

In his book *Saggio sulla negazione*, Paolo Virno argues that language, far from easing human contact, is, in reality, the basic source of conflict, misunderstanding, and violence.[5]

Only language establishes the possibility of negating what our senses are experiencing. Negation is like a switch that breaks the natural link between sensorial experience and its conscious elaboration. If immediate experience acknowledges a state of being, language can deny the state of being that is experienced. In this sense we can say that negation is the beginning of any mediation.

In the first pages of the book, Virno refers to biologist Vittorio Gallese's research on mirror neurons. According to Gallese and his colleagues, mirror neurons are what enable human beings to understand each other. They establish a net of inter-individual threads that trigger the process of understanding well before the individual becomes conscious of it. This implies that understanding is in fact a physical and affective phenomenon, before being an intellectual act.

According to Gallese, we understand the emotions and the actions of another person because, by looking at that person, we activate the same neurons that we would activate if we were feeling those same emotions, and performing those same actions.

We can call this mirror-like understanding *empathy*.

The development of linguistic competence, far from strengthening or confirming empathy, can be viewed as the beginning of the process of mediation that gradually erodes empathy, transforming understanding into a purely intellectual act of syntactic adaptation rather than a process of semantico-pragmatic osmosis.

According to Virno, language creates the un-natural possibility of reducing the light of immediate patency that surrounds perceptual experience. The order of language is syntactical: conventional rules open and close access to signification. In the course of human evolution, the syntactical order of language has invaded and reframed the immediacy of empathy, and in many ways it has perverted or destroyed its very possibility.

In his book *Ah Pook Is Here*, William Burroughs conceived of language as a virus that spread as a mutation in the human environment.[6] Virno adds that the content of this virus is *negation*, a laceration in the canvas of the shared perceptions and projections that we call reality.

Empathy is the source of conjunction. Over the course of the history of civilization and of techno-evolution, it seems that the syntactization of the world, that is, the reduction of the common world to the syntax of linguistic exchange, has slowly eroded traces of empathic understanding, and instead, enhanced the space of syntactic conventions. Linguistic mediation has developed technologies that in turn shape the *umwelt*, that is, the surrounding environment.

With the digital, we have reached the end-point of this process of increasing abstraction, and an apex in the increasing dissociation of understanding from empathy.

In *Zero Degrees of Empathy*, the British psychologist Simon Baron-Cohen evokes *empathy erosion* to explain cruelty and violence between human beings. For Baron-Cohen, empathy consists of two causally-related steps: the first is the interpretation of the signs that proceed from the other, and thus the extrapolation of the other's feelings, desires, and emotions; the second is the ability to respond accordingly.[7]

I call conjunction this form of empathic comprehension. I call connection, on the other hand, the kind of understanding that is not based on an empathic interpretation of the meaningful signs and intentions coming from the other, but rather on compliance and adaptation to a syntactic structure. The best explanation of the difference between conjunction and connection occurs in the third book of Tolstoy's *War and Peace*, when Prince Andrey Bolkonski compares the game of chess with the game of war.[8]

The opposition between conjunction and connection is not a dialectical opposition. The body and the mind are not reducible in an oppositional way to either conjunction or connection. There is always some connective sensibility in a conjunctive body, and there is always some conjunctive sensibility in a human body formatted in connective conditions. It's a question of gradients, shades, and undertones, not one of antithetical opposition between poles.

Recomposition and A-Signifying Recombination

In the midst of infinite births and deaths, in the midst of decay, leaves falling from trees, and waves on the sea—all the infinite chaotic events that randomly occur in the universe—the only stunning and unexpected thing is our inexhaustible craving for sense, harmony, and order.

Metaphysical and dialectic philosophy focused on the idea of totality, a concept that was based on the assumption either of a pre-existing order, or of a final order that it would restore or bring into being. According to the principles of totalitarian philosophy, each fragment would find its pre-established place, and all parts were arranged so as to compose an original or final totality, code, or destiny.

The phenomenological approach takes leave of the assumption that knowledge can lead to the perfect totality, and abandons the project of a totalitarian identification of thought and world. It thus opens the way to the possibility of different theoretical constructions, based on different *erlebnisses*, or forms of life. A rhizomatic methodology is just one among a multiplicity of possible phenomenological approaches.

According to a rhizomatic methodology, meaning emerges from a vibration that is singular in its genealogy, and can proliferate and be shared. Meaning is therefore an event, not a necessity—and we can share it with other singularities that enter into vibrational syntony, or sympathy, with our meaningful intentions.

A rhizomatic methodology does not presuppose or imply any totality that it would establish or restore. It is based on the principle of non-necessary conjunctions, and on the continuous molecular recomposition of cells, to borrow from a scientific vocabulary, whose destination is not implied in their program or genetic code.

Recomposition is a process of uncertain and autonomous subjectivation, where flows of enunciation interweave and create a common space of subjectivity. This collective subjectivity can be the result of an imagined form of belonging, such as a tribe, a nation, or a common faith. In this kind of collective existence, enunciation pretends to bring about truth, and divergence is seen as betrayal.

But collective subjectivity can also be the expression of an attraction: for example, desire as the singular creation of the other as singularity. In this case we can speak of a collective singularity, a singularity that is the living experience of a pathway from nowhere to nowhere. As Antonio Machado writes, and the Zapatistas repeat: "*Caminante no hay camino, el camino se hace al andar.*"

In this case, desire, as an attraction to singularity, generates the pathway and is the reason for collective existence (its *raison d'être*).

Rather than the homeland, the family, or ideological dogma, the collective subjectivity that I am trying to trace here is based on nomadic desire, not on belonging, or code.

I use the term recomposition to describe this process of social conjunction, that is, the opening and conjoining of individuals into a collective singularity, through which they express an affective and political solidarity that does not rely on identification, conventional codes, or marks of belonging.

Recomposition is the meeting, converging, and conjoining of singular bodies on a path that they share, provisionally, for a time. That common path is not inscribed in a genetic code, or in a cultural belonging—rather, it is the discovery of a common possibility that is the meeting point of singular drifts of desire. The community that results from the process of recomposition is a community of desire, not one of necessity. This is quite different from the process of recombination, where a-signifying segments are connected in accordance with coded rules of generation.

Conjunction versus Connection: The Ongoing Mutation

I call conjunction a concatenation of bodies and machines that can generate meaning without following a pre-ordained design, and without obeying any inner law or finality.

Connection, on the other hand, is a concatenation of bodies and machines that can only generate meaning by following an intrinsic, human-generated design through obeying precise rules of behavior and functioning.

Connection is not singular, intentional, or vibrational. Rather, it is an operative concatenation between previously formatted agents of meaning (bodies or machines) that have been codified, or formatted according to a code.

Connection generates messages whose meaning can only be deciphered by an agent (a body, a machine) that shares the same syntactic code that generated the message.

In the sphere of conjunction, the agent of meaning is a vibrating organism, where vibration refers to the uncertain and unresolved oscillation around an asymptotic point of isomorphism.

The production of meaning is the effect of the singularization of a series of signs (traces, memories, images, or words…).

Conjunction is the provisional and precarious syntony of vibratory organisms that exchange meaning.

The exchange of meaning is based on sympathy, the sharing of *pathos*.

Conjunction, therefore, can be viewed as a way of becoming other. Singularities change when they conjoin, they become something other than what they were before, in the same way as love changes the lover, or the conjunctive composition of a-signifying signs gives rise to the emergence of previously inexistent meaning.

By contrast, in the connective mode of concatenation, each element remains distinct and only interacts in a functional way. Rather than a fusion of segments, connection entails a simple effect of machine functionality.

In order for the connection to be possible, segments must be linguistically compatible. Connection thus presupposes a process whereby the elements that need to connect are rendered compatible.

The digital web, for example, extends through the progressive reduction of an increasing number of elements into a format, a standard, and a code that renders different elements compatible.

The considerations above are meant to introduce what I take to be the anthropological mutation that is underway in our times, essentially, a transition from the predominance of the conjunctive mode to the predominance of the connective mode in the sphere of human communication.

From the anthropological point of view, this technocultural change is centered on the shift from conjunction to connection in the paradigms of exchange between conscious organisms.

The leading factor of this change is the insertion of electronic segments into the organic continuum, the proliferation of digital devices in the organic universe of communication and in the body itself.

This leads to a transformed relation between consciousness and sensibility, and an increasingly desensitized exchange of signs.

Conjunction is the meeting and fusion of round or irregular bodies that are continuously weaseling their way about without precision, repetition, or perfection. Connection is the punctual and repeatable interaction of algorithmic functions, straight lines and points that overlap perfectly, and that plug in or out according to discrete modes of interaction that render the different parts compatible to a pre-established standard.

Passing from conjunction to connection as the predominant mode of conscious interaction between organisms is a consequence of the digitalization of signs, and of increasingly mediatized relations.

This digitization of communicative processes induces a desensitization to the curve, and to the continuous process of slow becoming, along with a concurrent sensitization to the code, or to sudden changes of state.

Conjunction entails a semantic criterion of interpretation. In order to enter into conjunction with another organism, the first organism sends signs to the other, signs whose meaning can only be interpreted in the pragmatic context of their interaction by tracing an intention, a shade of what remains unsaid, conscious and unconscious implications, and so on.

Connection instead requires a purely syntactic criterion of interpretation. The interpreter must recognize a sequence and be able to carry out the operation that is foreseen by the *general syntax* (or operating system); there is no margin for ambiguity in the exchange of messages, nor can intention be manifest though nuances.

The process of this gradual translation of semantic interpretations into syntactic differences runs from modern scientific rationalism, to cybernetics, and artificial intelligence programs.

Connective Logic

The debate on artificial intelligence began in the 1960s.

To outline the problem that lies at the core of artificial intelligence, Hubert Dreyfus distinguished between "areas in which relevance has been decided beforehand [...], and areas in which determining what is relevant is precisely the problem."[9]

When we exchange messages in the conjunctive sphere, we are trying to find out what is relevant for those who are participating in the communication. We don't know what our common object of interest and attention is: communication is about shedding light on that point. In the connective sphere, on the contrary, we start from a common ground of conventional knowledge, translated into technological standards and formats that make connection possible.

Concerning the genesis of connective methodology in the history of modern philosophy, Hubert Dreyfus writes:

> As Galileo discovered that one could find a pure formalism for describing physical motion by ignoring secondary qualities and teleological considerations, so, one might suppose, a Galileo of human behavior might succeed in reducing all semantic considerations (appeal to meanings) to the techniques of syntactic (formal) manipulation.
>
> The belief that such a total formalization of knowledge must be possible soon came to dominate Western thought. […] Hobbes was the first to make explicit the syntactic conception of thought as calculation. [...] Leibniz thought he had found a universal and exact system of notation, an algebra, a symbolic language, a "universal characteristic" by means of which "we can assign to every object its determined characteristic number."[10]

Dreyfus then retraces the steps that led to the formation of the contemporary digital mind-set.

> An important feature of Babbage's machine was that it was digital. There are two fundamental types of computing machines: analogue and digital. Analogue computers do not compute in the strict sense of the word. They operate by measuring the magnitude of physical quantities. Using physical quantities such as voltage, duration, angle of rotation of a disk, and so forth, proportional to the quantity to be manipulated, they combine these quantities in a physical way and *measure* the results. A digital

> computer [...] represents all quantities by discrete states, for example, relays which are open or closed, a dial which can assume any one of ten positions and so on, and then literally *counts* in order to get results. [...] since a digital computer operates with abstract symbols which can stand for anything, and logical operations which can relate anything to anything, any digital computer [...] is a universal machine.[11]

The universal digital machine is the logical and technological condition of our contemporary anthropological mutation.

Conjunction is the opening of bodies to the understanding of signs and events, and their ability to form organic rhizomes, that is, concrete, carnal concatenations of pulsating vibratory bodily fragments with other pulsating vibratory bodily fragments.

On the contrary, in a digital environment, only what fulfills the standard of compatibility can connect, meaning that certain elements will be unable to connect to others. In order for distant communicative agents to be able to connect, we must provide them with tools enabling them to access the flow of digital information.

When connection replaces conjunction in the process of communication between living and conscious organisms, a mutation takes place in the field of sensibility, emotion, and affect.

As I have noted before, this mutation occurs in time, in the diachronic dimension of the transition from the modern mechanical environment of indust-reality to the postmodern environment of semio-economy. But it is not homogeneous, as it depends on the particular features of the cultural context, geo-cultural and synchronic, in which it takes place.

I will thus turn to selected, synchronic cultural contexts to investigate the different forms of this diachronic, connective mutation, with a special attention to the relation between aesthetic sensibility and forms of emotional life.

Evolution and Sensibility

The expression *cognitive wiring* refers to the capture and submission of life and mental activity into the sphere of calculation. This capture occurs on two different levels: on the epistemic level it implies the formatting of mental activity, on the biological one it implies the technical transformation of the processes by which life is generated.

In the modern age, the modeling of the body was essentially macro-social and anatomical—as Michel Foucault has extensively shown in his works about the genealogy of modernity. The subjection of the social body to industrial discipline was linked to the macro-social action of repressive machines acting on the individual body.

Today, digital technology is based on the insertion of neurolinguistic memes and automatic devices into the sphere of cognition, into the social psyche, and into forms of life. Both metaphorically and non-metaphorically, we can say that the social brain is undergoing a process of wiring, mediated by immaterial linguistic protocols as well as by electronic devices.

As generative algorithms become crucial in the formation of the social body, the construction of social power shifts from the political level of consciousness and will to the technical level of automatisms located within the process of generating linguistic exchange and forming psychic and organic bodies.

My attention here will be focused on the biosocial modeling of sensibility, that is, on the embedding of cognitive automatisms at the deep levels of perception, imagination, and desire. This implies that social becoming is no longer understandable in the framework of history but in the framework of evolution.

History is the conceptual sphere where conscious voluntary actors transform the conditions and social structures that surround them. In the sphere of evolution, on the other hand, human beings cannot be considered actors because evolution refers to the natural becoming of organisms in their interaction with the environment.

From the point of view of intentionality, the concepts of history and evolution can be distinguished, and opposed. The concept of history, emphasized by the romantic tradition, was particularly important to the Hegelian dialectical tradition, including Marx and the Marxist movement. The concept of evolution, on the other hand, was elaborated in a cultural space more akin to the positivist school of thought.

Historical action takes place when political intentionality is effective in modeling the environment. Evolution, on the contrary, occurs when the exchange between humans and nature, and the reciprocal transformation of these terms cannot be controlled by intentional political action.

In today's conditions of hyper-complexity and technological acceleration, the social sphere can no longer be properly understood in terms of political transformation. It is better explained through evolution, particularly neural evolution. Indeed, the evolution of the brain resulting from environmental action on cognition and society, and the subjective adaptation of the human mind are today the main factors of social transformation, and can hardly be subjected to political will.

In the context of history as outlined above, political action was driven by will, rational understanding, and prediction—while in the context of evolution, the organism is understood to become attuned to its environment, with sensibility being the faculty that makes this syntonization possible. Consequently, the relevance and effectiveness of human action no longer occurs at the level of rational knowledge, political decision, and will, but instead at the level of intuition, imagination, and sensibility.

Clearly, the conceptual and practical sphere of modern politics has lost its ground.

In the age that began with Machiavelli and culminated with Lenin, human will (the prince, the state, the party) was able to reign on the infinite chaotic variation of events and projects, and subject individual interests and passions to the common goal of social order, economic growth, and civil progress.

The technical transformation that we witnessed in the last decades of the twentieth century, the infinite proliferation of sources and flows of information, unleashed by the acceleration of network technology, has rendered impossible the conscious elaboration of information by the individual mind, and the conscious coordination of willful individual agents.

The loss of effectiveness of political action is essentially due to a change in temporality: with the acceleration and complexification of the infosphere, reason and will—those essential tools of political action—can no longer process and decide in time. Technical transformation has radically altered the conditions of mental activity and the forms of interaction between the individual and the collective spheres.

In the age of voluntary action that was called modernity, these two spheres—the individual and the collective—could be seen as distinct, externally linked, and interacting on the basis of an effective intentionality.

Today, the distinction between the individual and the collective has been blurred. Crowds and multitudes are involved in automatic chains of behavior, driven by techno-linguistic dispositives. The automation of individual behavior—since individuals have been integrally penetrated and concatenated by techno-linguistic interfaces—results in a swarm effect. If the human is the animal who shapes the environment that shapes his/her own brain, the swarm effect is thus the outcome of the human transformation of its technical environment, leading to the subjugation of mental behavior.

Tega Brain
The Environment is not a system

http://www.aprja.net/the-environment-is-not-a-system/

Tega Brain, *Artist and Assistant Professor of Integrated Digital Media, New York University*

Figure 1: Seagrass in Tasmania, Australia. Credit: Tega Brain.

In late 2017, Microsoft's chief environmental scientist, Lucas Joppa announced AI for Earth, a new initiative to put artificial intelligence in the hands of those who are trying to "monitor, model and manage the earth's natural systems". AI for Earth gives environmental researchers access to Microsoft's cloud platform and AI technologies, and similar to recent initiatives by companies like Google and Planet Labs, it aims to integrate AI into environmental research and management.
It is obvious that Silicon Valley stands to profit handsomely from the uptake of AI in environmental research and management, as it has from the application of these methods in a diverse range of other fields. From urban design to the justice system, decision making processes are being automated by data-driven systems. And in spite of a growing body of criticism on the limitations of these technologies,[1] the tech industry continues to promote them with the mix of solutionism and teleology that imbues Joppa's words. He urges: "for every environmental problem, governments, nonprofits, academia and the technology industry need to ask two questions: 'how can AI help solve this?' and 'how can we facilitate the application of AI?'" (Joppa)

This paper considers some of the limitations and possibilities of computational models in the context of environmental inquiry, specifically exploring the modes of knowledge production that it mobilizes. As has been argued by authors like Katherine Hayles and Jennifer Gabrys, computation goes beyond just reading and representing the world. As a mode of inquiry it has a powerful world-making capacity, generating new pathways for action and therefore new conditions. "Computing computes."[2] Computational metaphors are also pervasive as framing devices for complex realities, particularly in the context of research on the city, the human brain or human behavior.[3]

Historic computational attempts to model, simulate and make predictions about environmental assemblages, both emerge from and reinforce a systems view on the world. The word eco-*system* itself stands as a reminder that the history of ecology is enmeshed with systems theory and presupposes that species entanglements *are operational* or *functional*. More surreptitiously, a systematic view of the environment connotes it as bounded, knowable and made up of components operating in chains of

cause and effect. This framing strongly invokes possibilities of manipulation and control and implicitly asks: *what should an ecosystem be optimized for?*
[4]

This question is particularly relevant at a time of rapid climate change, mass extinction and, conveniently, an unprecedented surplus of computing. As many have pointed out, these conditions make it tempting (and lucrative) to claim that neat technological fixes can address thorny existential problems.[5] This modernist fantasy is well and truly alive for proponents of the smart city, and even more dramatically in proposals for environmental interventions that threaten to commodify earth's climate conditions, such as atmospheric engineering.[6]

What else does a systems view of the environment amplify or edit out? This discussion revisits several historic missteps in environmental measurement and modeling in order to pull focus on the epistemological assumptions embedded into a systems perspective. It then asks, what are other possibilities for ecological thought? Does AI have any potential to reveal environments in ways that escape the trapping of systems? Critical to my inquiry is the recent work of Anna Tsing and what she calls, "the arts of noticing". Tsing's work offers a starting point for thinking outside of both a systems framework and assumptions of progress (17). Her perspective on ecology and the lifeworlds it describes unfolds and emerges through "encounters" (20) which bring together entities, transforming them in indeterminate ways. Might AI operate through modes of environmental encounters or will it simply amplify "an informatics of domination" (Haraway 162)?

The Poverty of Numbers

A systems view of the environment reinforced by computation, has numerous precedents, including 18th and 19th century attempts at scientific forest management. This early attempt at centralized ecosystem management through numerical modeling foreshadows the contemporary use of these approaches in the context of computation. James C. Scott traces how the introduction of centralized forestry required forests to be made legible in new ways.[7] Trees in forests were measured, quantified and modeled to optimize harvest and replanting for timber yield. Thus the fastest growing species were replanted in felled areas, and trees became viewed as autonomous machines for producing wood. Those species not harvestable for timber – low lying bushes, fungi and plants (Scott 13), as well as traditional 'unofficial' use of forests by local communities – were edited out of the system (Hölzl 436). These scientific or fiscal forests, were managed with the assumption that complex species entanglements were irrelevant and could be treated as external to a system designed to efficiently transform trees into commodities. Yet after a couple of generations of felling and replanting, yields began to drop and the health of managed forests deteriorated (Scott 20). Viewing the forest as a factory oversimplified the reality of the relations and interdependencies of its species.

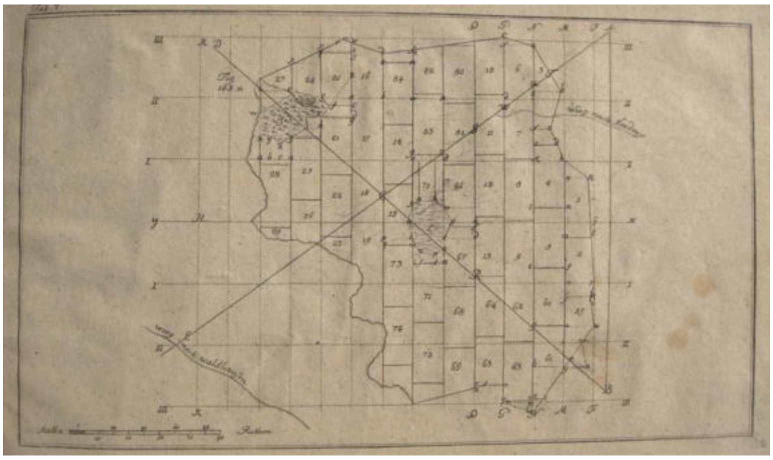


Figure 2: Imaginary forest patch partitioned in 84 sections. Credit: Grünberger, G. (1788) Lehrbuch

für den pfalzbaierischen Förster, Vol. 1 (München: Strobl), Figure 163 from Historicizing Sustainability: German Scientific Forestry in the Eighteenth and Nineteenth Centuries (Hölzl).

The scientific forest failed by its own criteria: timber yield. However it is worth acknowledging that if yield had remained high while biodiversity declined, this history of sustainable environmental management would be remembered as a success, analogous to industrial agriculture. Tsing calls environments that are simplified and optimized to produce commodities “plantations” (435). The economic drivers of capitalism make crop yields the ultimate goal of agricultural landscapes, and shape how they are measured, modeled and manipulated. When a landscape is managed as a factory, its species become assets alienated from their lifeworlds[8] like workers who fulfill HITs on Mechanical Turk with no bearing on each other or what they produce. When the asset can no longer be extracted, the landscape becomes a ruin and disappears from view, deemed worthless (Tsing 31). Both the plantation and the scientific forest are the results of numerical approaches to landscape management applied in the name of economics. They highlight that data collection and modeling practices are never neutral. Rather, they are contingent on decisions of what is deemed important or trivial in the eyes of the manager and therefore are profoundly driven by culture and economics, class and race.

The Fantasy of Stability

In the twentieth century, the science of ecology emerged in dialogue with cybernetics and systems theory. There is a rich body of literature critiquing how these conditions influenced environmental research.[9] Cybernetics, first defined in the 19th century by André-Marie Ampère as “the science of governance” was catalyzed as an interdisciplinary field by proponents like Norbert Wiener in the post war decades.[10] It inspired ecologists to pursue questions of control and self regulation in the context of species lifeworlds. Some early ecosystem diagrams were even realized in the style of circuitry.

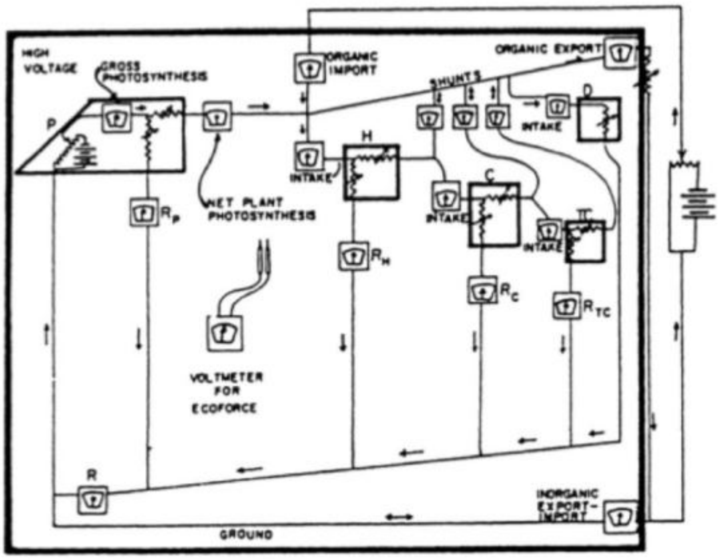


Figure 3: Prominent biologist of the 1960s, Howard Odum’s first presentation of an ecosystem using the symbolism and aesthetic of electric circuit diagrams. Image by Howard Odum, 1960 cited in Madison (218).

Botanist Michael Tansley was among the first to use the term “ecosystem” in 1935 to describe the “systematic” functioning of forests, grasslands and wetlands environments. He saw ecosystems as “the whole system (in the physical sense), including not only the organism-complex, but also the whole complex of physical factors forming what we call the environment of the biome [… these] are the basic units of nature” (299). Like the scientific foresters, Tansley proposed that ecosystems were made of discrete stable units, interacting in ways that tend towards a state of dynamic equilibrium. He also assumed that natural selection favors stability, that “systems that can attain the most stable equilibrium, can survive the longest" (Tansley 299). This idea of ecological equilibrium remains stubbornly influential, as does the idea of the environment as a unified "whole". As philosophers like Bruno Latour and Timothy Morton discuss, the idea that the "natural world" exists in a balanced harmonious state that is then disrupted by humans reiterates the misconception that humans and environment are separate.[11]

Towards the late 1960s, Tansy's assumption of ecosystem homeostasis was proving difficult to verify, even in ambitious large-scale ecosystem modeling projects enabled by the availability of computation. One such project was the Grasslands Biome, started in 1968 at Colorado State University. It was an unprecedented attempt to comprehensively model a grasslands ecosystem with a computational model and aimed to uncover new ecological principles (Kwa 1). Employing hundreds of full time researchers, the project involved extraordinary methods of data collection as researchers tried to account for all forms of energy entering and leaving the system, attempting to quantify everything eaten and excreted by all organisms in the biome and then inputting this data into a mathematical model. Students and researchers would follow animals around the grasslands whispering into tape recorders. They would 'collect' animals and analyze their stomach content by inserting probes into their digestion systems (Coupland). Soil microbiology was also studied, yet soil invertebrates and highly mobile species such as insects and birds remained frustratingly uncooperative in yielding information to researchers (Coupland 35).

Despite this labor, the Grasslands model, like similar large-scale ecological modeling programs of the time, revealed very few new ecological principles. Deemed "too simplified biologically" despite implementing an unprecedented number of variables (Coupland 154), the model was built with an assumption of default equilibrium. Coupland argues that the Biome Model was simply "a sophisticated version of a cybernetic system […] and cast […] the ecologist in the role of systems engineer" (146). The project disproved its foundational hypothesis – that complex ecological realities can be reconciled with mathematical models and be described as abstracted structures of inputs and outputs. "The grandiose ideal of achieving total control over ecosystems, which around 1966 appealed so much to systems ecologists as well as politicians, was dismissed as a hyperbole" (Coupland 155).

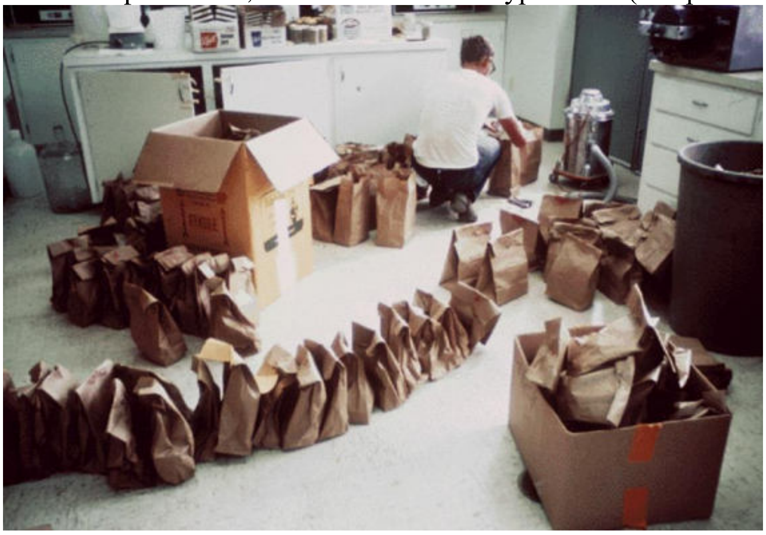


Figure 4: Processing of replicate biomass samples, ready for drying and weighing, in the field laboratory at the CPER/Pawnee grassland site, Colorado, USA. Credit: Larry Nell, Colorado State University, July 1971.

Data collection and modeling practices remain shaped by what is considered typical or atypical, important and peripheral – summations of the boundary conditions of reality. However making these assumptions is difficult. Even with the growing capacity of contemporary computing, it is dangerous to simply assume that more data equals more reality. An example of this is the story of how Joe Farman, a British geophysicist working for the British Antarctic Survey, first observed the destruction of the ozone layer. Farman maintained a single ground based ozone sensor in the Antarctic throughout the 1960s and 1970s, and continued to do so in spite of the launch of NASA atmospheric monitoring satellites that collected vastly larger quantities of data (Vitello). When Farman's sensor began to show a 40% drop in ozone levels in the early 1980s, he assumed it was damaged and replaced it as NASA's atmospheric models had reported no such change. After years carefully checking, Farman published this alarming result in *Nature* as the first observation of the destruction of the ozone layer due to human pollutants. Until then, this had been only a theoretical hypothesis.[12] How had NASA's satellites missed such a marked change in ozone composition? One response from NASA suggests that their data processing software was programmed to discard readings that appeared to be outliers, thus ignoring the drastic changes that were occurring in ozone concentration (Farman). In this case, reality itself was an outlier and assumed to be an error.

The Limits of Machine Learning

What if there was no cap on the amount of data produced from an environment for analysis? Could models be derived from from datasets rather than built from theory to avoid erroneous assumptions like those made in the Grasslands model? Could machine learning be adopted to deal with quantities of data beyond human comprehension and prevent any need for discarding outliers? Can these techniques produce a more robust representation of reality, free of human judgement?

These are the arguments made for machine learning. In 1959 Arthur Samuel defined machine learning as "the ability to learn without being explicitly programmed" (McCarthy). Rules are derived from patterns in large data sets, rather than programmed based on theoretical knowledge of underlying structures. "Correlation is enough. We can stop looking for models" proclaimed *Wired* editor Chris Anderson in 2008, in an article titled "End of Theory". In other words, had the Grasslands model been derived through machine learning, energy flows through the ecosystem could have been estimated based on correlations the data, rather than estimated from inputting data into a theoretical model, hardcoded from hypothesis of ecosystem dynamics. Although this would have prevented erroneous assumptions like default homeostasis, it is important to acknowledge that machine learning substitutes one set of assumptions for another.

Machine learning assumes that enough data can be collected to adequately represent and make predictions about reality. In the context of the environment, this is an enormous challenge given the very limited size of our existing datasets. Another significant assumption is that the past is indicative of the future. Yet as the sudden unprecedented depletion of atmospheric ozone in the 1980s shows, this to not always be the case. Similarly, climate change means our ability to make accurate predictions from our existing data is diminished. Many environmental datasets like precipitation records span 250 years at best, with the majority spanning a much shorter period.[13] From a geological point of view this is an absurdly small slice of time, and one in which the earth's climate has been relatively stable. As the patterns, rhythms and cycles in both climatic and biological phenomena are drastically disrupted, it becomes increasingly difficult to make predictions based on this short, stable interval of climate data. William B. Gail calls this the coming of "a new dark age", where the accumulated observations of Earth's irreducibly complex conditions are increasingly rendered obsolete. If machine learning approaches are to be adopted, it is important to recognize the limits of these methods.

Dreams of Objectivity

Another prominent argument made for the use of AI methods is that data-driven approaches neutralize human decision making by simply representing the world as it is. The proponents of AI for Earth also make these claims to objectivity: "Decisions about what actions to take will be easier to make — and less vulnerable to politicization — if we know what is happening on Earth, when and where. AI can help to provide that information." (Joppa) However in other realms, AI systems continue to reveal and confirm biases and structural inequalities rather than offering an easy pathway to their neutralization.

For example, defendant risk scoring systems designed to help judges make decisions to “deliver better outcomes to all who touch the justice system” (Equivalent) have been shown to score black defendants at significantly higher risk for reoffense than white defendants with similar or worse criminal records (Angwin et al.). Systems like these should serve as warnings to other industries implementing automating decisions making, even in the name of environmental management. As theorist Françoise Vergès argues, “adaptation through technology or the development of green capitalism […] does not thoroughly address the long history and memory of environmental destruction […], nor the asymmetry of power.” Contemporary environmental challenges directly emerge from violent histories of colonialism, imperialism and the ongoing exploitation of marginalized communities or those living in the global South (Vergès). As such, there is no reason to suggest that AI technologies built and implemented by a cohort of wealthy white men in the US will in any way manage or distribute environmental resources in ways that are equitable for everyone.

Technologies will only ever provide partial fixes if they are not accompanied by shifts in perception and values, along with regulatory change that addresses histories of injustice and “the tradition of belief in progress” (Vergès). More efficient resource use in a system of deregulated capitalism is most likely to beget further resource use rather than net reduction. Microsoft seems to have it backwards in its mission statement “to empower every person and organization on the planet to achieve more”. Wasn’t the idea behind technologies of automation to empower us to achieve less? Or at least prompt a radical rethinking of what ‘more’ is? As Vergès argues, if these logics go unquestioned, mounting environmental challenges will not only continue to accelerate change in an already stressed biosphere, but also further augment environmental injustices.

If the Environment is Not a System, Then What is it?

How else might we think of environments in lieu of the systems metaphor? Tsing offers the concept of assemblage and here I build on her work, understanding environments as open ended assemblages of non-humans, living and nonliving, entangled in ways of life.

“Ecologists turned to assemblages to get around the sometimes fixed and bounded connotations of ecological ‘community.’ The question of how the varied species in a species assemblage influence each other — if at all — is never settled: some thwart (or eat) each other; others work together to make life possible; still others just happen to find themselves in the same place. Assemblages are open-ended gatherings. They allow us to ask about communal effects without assuming them.” (Tsing 54)

Like Tsing, many authors have taken up the concept of assemblage to round out the simplification and abstraction connotated through use of technological metaphors. Following Latour, to assume a system is also to surreptitiously assume “the hidden presence of an engineer at work”, a presence that suggests intention and that what we can see are parts of a unified whole (*Some Advantages of the Notion of “Critical Zone” for Geopolitics*, 3). Assemblage relieves us of this view, instead suggesting a collection of entities that may or may not exhibit systematic characteristics. The edges of an assemblage are fuzzy – modes of interaction are always shifting and agencies within them are heterogeneous. Katherine Hayles also invokes the term in her inquiry on cognition in complex human technological entanglements, what she calls “cognitive assemblages” (*Unthought 3*). Hayles chooses assemblage over network arguing that network conveys “a sense of sparse, clean materiality”, whilst assemblage offers “continuity in a fleshy sense, touching, incorporating, repelling, mutating” (118). She continues: “I want to convey the sense of a provisional collection of parts in constant flux as some are added and others lost. The parts are not so tightly bound that transformations are inhibited and not so loosely connected that information cannot flow between parts” (118). Similarly, I take up assemblage as an imperfect descriptor to avoid the hubristic assumptions of a systems view. Stating “I am studying a grasslands assemblage” instead of “I am studying a grasslands system” produces a remarkable shift in expectations and assumptions. This simple substitution dismantles subtle assumptions of fixed categories of knowledge, as well as assumptions that engineering and control are always possible. Instead it foregrounds uncertainty and acknowledges the unknowability of the world.

Rather than describing ecology through interactions or exchanges between entities, Tsing proposes that it emerges through encounters. For Tsing, encounters open new possibilities for thinking. They produce transformation and are therefore indeterminate (63). They are also non-human centered. There can be encounters between different species – say a mushroom and a pine tree – or between lifeforms and non-human materials. Components of a system are implied to be static discrete units, leaving out processes of contamination and transformation. For example when predator-prey relations are described as transfers of energy between components in a system, say a walrus eats a mollusc, it is inferred that the walrus remains unchanged by the encounter. Seeing the world as made up of individuals sealed off from one another, allows for the assumption of stable categories, and makes the world easier to quantify through data, interpreted as pattern and codified as algorithm. The yield from a data-driven mode of knowledge production is obviously rich and wide reaching, providing new insight into phenomena like climate change. And yet, as the story of Farman's attention to the atmosphere shows, scaling and automating data collection processes can risk overpresuming the stability of the world and blind us to transformations outside of assumed possibility spaces.

In this way "smartness", in its current form, produces a kind of myopia. A smart city, home or environment contains networks of sensors automatically pinging data back to servers to train machine learning models of the world. Indeed this is also Joppa's pitch for AI for Earth: "AI systems can now be trained to classify raw data from sensors on the ground, in the sky or in space, using categories that both humans and computers understand, and at appropriate spatial and temporal resolution." This statement is worthy of carefully consideration. Firstly, how does one decide on an appropriate temporal resolution? In the case of the German forests, it took nearly a century to see that management methods were unsustainable because the life rhythms of a tree are at a vastly slower tempo than those of human economies. Joppa also infers that the world can be revealed by how it appears through "raw sensor data". Yet this implies the sensors themselves as somehow neutral and overlooks the layers of human decision making that has occurred in their production and installation.[14]

It can also be surprisingly difficult to resolve the world into clearly defined categories. And are these categories stable? Tsing's argument that encounters produce transformation suggests that neat taxonomies will never fully accommodate the fluidity and uncertainty of the world. This is particularly apparent in plant systematics where even the definition of species is contested and ever changing (Ernst). In trying to categorize plant specimens, a tension can emerge between how the specimen appears – its phenotype, and how it appears on a genetic level – its genotype. As genetic sequencing techniques have become cheaper and therefore more widely available, plant scientists sometimes find that the species indicated by phenotype does not always match up to the genotype – a discovery that has caused many herbaria to be reorganized. However even when identifying specimen on a purely genetic level, there is still dispute over how species are interpreted.[15]

Data-driven research methods necessitate the collection of huge quantities of data and in doing so, they dismantle opportunities for paying close specific attention to the world. These methods also tend to obscure the many other ways of building understanding. Also, perhaps intentionally, data collection increasingly acts to maintain the status quo. We use data to study problems that would be more effectively addressed through simple political action. The impetus to "study the problem" ad nauseam gives the appearance of addressing an issue while perfectly maintaining the present state of affairs.[16]

Amplifying Encounters

How might we reciprocally illuminate the environment and balance our well oiled capacity for imagining it from an all-conquering systems worldview? How might we elevate engagement through the specifics of encounter and narrative?

Ethnography is one possibility. Tsing's study of the matsutake mushroom explores what can be learnt from a Japanese mushroom, a lifeform that cannot be cultivated and that thrives in highly disturbed forests. Through her ethnography she shows how close attention inevitably facilitates transformation. Tsing calls this "the arts of noticing", tactics for thinking without either the abstraction produced by quantification or deeply held assumptions of progress. If we are "agnostic about where we are going, we might look for what has been ignored" (51). As Farman's ozone research showed, paying close attention rather than outsourcing observation and interpretive capacities can reveal the world in different ways. In particular, attention can emphasize the indeterminacy and messiness of encounters outside of an engineering agenda. It can transform the observer, directly involving us in the weirdness of the world.

Could technologies like machine vision and remote sensing be used to amplify environmental encounters and the arts of noticing our ecological entanglements? The rise of digital naturalism sees the development of apps and initiatives that focus attention on the lifeforms in our various bioregions. Initiatives such as *iNaturalist*, *Merlin Bird ID* and *eBird* invite non-scientists to contribute environmental observations and use either crowd-sourced or "assisted identification" to identify species and build biodiversity databases. Assisted identification utilizes computer vision techniques to guide species identification from images by identifying broad categories and making suggestions. Through this process, the system is also gradually being trained, and over time will therefore make better suggestions. Many scientific institutions also hope that data-driven species identification can help to reduce the bottlenecks in identification processes as human taxonomists are in short supply (Kim).

It is also worth emphasizing that these apps do not purport to replace human identification but rather facilitate human computer collaboration to reach conclusions quicker. This is significant, as it shows a way that AI can produce more meaningful environmental encounters rather than automate them away. This use case for AI also serves as a reminder that data can be much more than a material for building a simulation or instrumentalizing whatever is being measured. The act of data collection and collaborative identification can amplify encounters and, by extension, yield transformation or what artist Jenny Odell calls "a certain dismantling in the mind." In observing a local bird, and being assisted to identify it as a magpie, I'm learning and tuning my perception to the lifeworlds I inhabit: I'm subject to transformation.

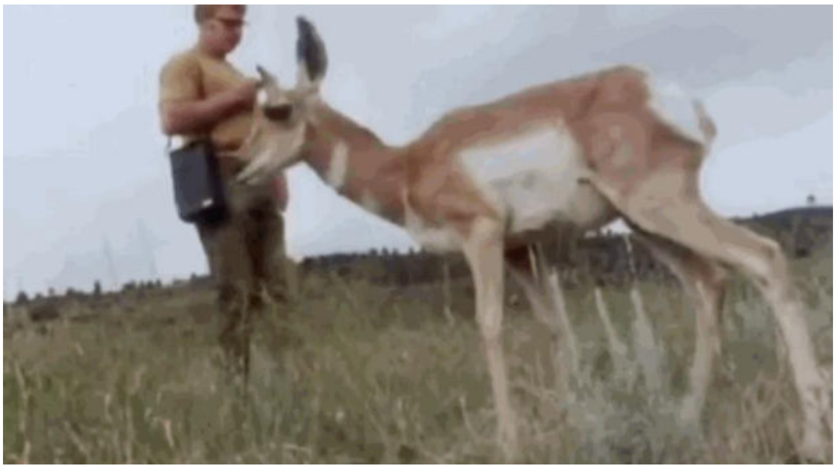


Figure 5: Deer observations made at the CPER/Pawnee grassland site, Colorado, USA. Credit: Animated GIF made from Adam Curtis' documentary All Watched over by Machines of Loving Grace.

Accounts of the scientific forest, the Grasslands Biome and Farman's ozone observations, mostly focus on the success or failure of the science – on whether these projects of observation or modeling succeeded or failed in revealing new patterns, on whether the resultant environmental models proved accurate, and, by extension, on whether they produced new possibilities for environmental management and manipulation. But telling these stories like this, is telling them from a systems point of view. And what tends to get overlooked is how these are actually stories of environmental encounter though data collection. As encounters, they are also stories of transformation of both the environments and the humans involved. How did the meticulous observation of the environmental assemblages in question shift and transform the people studying them? In itself, this question rejects a false binary between human and environment. It acknowledges the instability of the observer and the tendencies of Western science to edit out intuition, emotion and philosophical recalibrations. The reciprocal transformation that occurs with attention and encounter, what Nobel prize winning geneticist Barbara McClintock called "getting a feeling for the organism", is not only critical for formulating original scientific hypothesis, but more deeply, for questioning foundational assumptions of what is counted as knowledge and what we then expect knowledge to do.[17] Looking back on the early scientific forests and even on the more recent Grasslands Biome, it is difficult to speculate on how these projects changed the people involved. However, their stories remind us of the irreducibility of an unruly and complex environment. That as hard as we try to contain the world in neat technological metaphors, it will always leak out and transform us.

Notes

[1] See recent books *Weapons of Math Destruction* by Cathy O'Neil, *Automating Inequality* by Virgina Eubanks, *Code and Clay, Data and Dirt: Five Thousand Years of Urban Media* by Shannon Mattern, and the *Machine Bias* Series published by Propublica and written by Julia Anguin et al.

[2] See Katherine Hayles (*My mother was a computer*, 7-31) and Jennifer Gabrys' discussion in *Program Earth* (11).

[3] Sociologist Shannon Mattern warns of the "the city as computer model" arguing that it often hinders "the development of healthy, just, and resilient cities" (*The City is Not a Computer*). Psychologist Robert Epstein highlights similar issues in the context of brain research observing that historically, metaphors for cognition have always been drawn from the dominant technology of the time – hydraulics, springs and mechanics, electrical circuits and now computation. Epstein argues that the ubiquity of information processing metaphors in brain research may well be constraining the field by confining hypotheses and explanations to those that align with computational processes. These metaphors equally constrain approaches to environment inquiry.

[4] This question is inspired by Shannon Mattern's discussion of the city as a computer metaphor (*The City is Not a Computer*).

[5] See Bratton et al. (9); Gabrys (230); Stengers (1000), and Szerszynski et al (2818).

[6] See Temple on the planned atmospheric tests scheduled to occur in the US in 2018.

[7] See James C. Scott's well known account of scientific forestry in *Seeing Like a State*.

[8] I use the word 'lifeworlds' following Anna Tsing who describes objects in capitalist exchange as being alienated and "torn from their lifeworlds" (121).

[9] Many authors discuss the influence of systems theory on ecology, such as Elichirigoity, *Planet Management*, and Latour, *Some Advantages of the Notion of "Critical Zone" for Geopolitics*. Some also consider the influence of cybernetics such as Haraway, *The High Cost of Information*, and Jennifer Gabrys, *Program Earth*.

[10] See Wiener's landmark 1948 book, *Cybernetics.*

[11] Latour's concept of "naturecultures" introduced in the *Politics of Nature* is an attempt to collapse a false binary between the human concerns and nature. Morton, builds on this in *The Ecological Thought* that also rejects this bifurcation.

[12] The theory of ozone destruction was published by Molina et al.

[13] See Simpson.

[14] See Gabrys; Bratton et al.

[15] See Fazekas for discussion of differences in species definitions. Hull discusses how these uncertainties have led to the concept of reciprocal illumination in plant systematics. This concept acknowledges the multiple methods for classifying and naming species.

[16] Now discontinued, *The Human Project* was an example of data collection in lieu of political action. The project planned to address issues of health, urban design and inequality by collecting huge volumes of data from 10000 New Yorkers over 20 years.

[17] See Keller's biography of McClintock's life.

Works cited

AI for Earth Grant, Microsoft, 2017. Web. https://www.microsoft.com/en-us/research/academic-program/azure-research-award-ai-earth/. Accessed 10 Jan. 2018.

Ampère, André-Marie, Charles Augustin Sainte-Beuve, and Émile Littré. *Essai sur la philosophie des sciences*. Vol. 1. Paris: Bachelier, 1834. Print.

Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired* magazine, 16 July 2008. Web. https://www.wired.com/2008/06/pb-theory/. Accessed 10 Feb. 2018.

Angwin, Julia, et al. "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased against Blacks." *ProPublica, May* 23 (2016). Print.

Bratton, Benjamin. H., and Natalie Jeremijenko. "Latent Interfaces, Suspicious Images." *Situated Technologies pamphlets* 3 (2008). Print.

Curtis, Adam. *All Watched Over by Machines of Loving Grace*. London: BBC, 2011. Film Series.

Coupland, Robert T., ed. *Grassland Ecosystems of the World: Analysis of Grasslands and their Uses*. Vol. 18. Cambridge: Cambridge University Press, 1979. Print.

eBird, *The Cornell Lab for Ornithology*. Web. https://ebird.org. Accessed 10 Feb. 2018.

Elichirigoity, Fernando. *Planet Management: Limits to Growth, Computer Simulation, and the Emergence of Global Spaces*. Evanston, IL: Northwestern University Press, 1999. Print.

Epstein, Robert. "The Empty Brain." *Aeon* (2016). Web. https://aeon.co/essays/your-brain-does-not-process-information-and-it-is-not-a-computer. Accessed 10 Feb. 2018.

Equivalent, 2018. Web. http://www.equivant.com/about-us. Accessed 10 Feb. 2018.

Ernst, Mayr, "Species Concepts and Definitions." *Topics in the Philosophy of Biology*. Dordrecht: Springer, 1976, pp. 353-371. Print.

Eubanks, Virginia. *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor*. Basingstoke: Macmillan, 2018. Print.

Farman, Joseph C., Brian G. Gardiner, and Jonathan D. Shanklin. "Large Losses of Total Ozone in Antarctica Reveal Seasonal ClOx/NOx Interaction." *Nature* 315.6016 (1985): 207. Print.

Fazekas, Aron J., et al. "Are Plant Species Inherently Harder to Discriminate than Animal Species using DNA Barcoding Markers?." *Molecular Ecology Resources* 9.s1 (2009): 130-139. Print.

Gail, William B. "A New Dark Age Looms." *New York Times*, 19 April 2016. Web. https://www.nytimes.com/2016/04/19/opinion/a-new-dark-age-looms.html. Accessed 10 Feb. 2018.

Gabrys, Jennifer. *Program Earth: Environmental Sensing Technology and the Making of a Computational Planet*. Minneapolis: University of Minnesota Press, 2016. Print.

Hayles, N. Katherine. *Unthought: The Power of the Cognitive Nonconscious*. Chicago: University of Chicago Press, 2017. Print.

Hayles, N. Katherine. *My Mother was a Computer: Digital Subjects and Literary Texts*. Chicago: University of Chicago Press, 2010. Print.

Haraway, Donna. "The High Cost of Information in Post-World War II Evolutionary Biology: Ergonomics, Semiotics, and the Sociobiology of Communication Systems." *Philosophical Forum*. Vol. 13. No. 2-3 (1981). Print.

Haraway, Donna. *Simians, Cyborgs and Women: The Reinvention of Nature*. London: Free Association Books, 1991.

Hölzl, Richard. "Historicizing Sustainability: German Scientific Forestry in the Eighteenth and Nineteenth Centuries." *Science as Culture* 19.4 (2010): 431-460. Print.

Hull, David L. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press, 2010. Print.

iNaturalist, *California Academy of Sciences*. Web. www.inaturalist.org. Accessed 10 Feb. 2018.

Joppa, Lucas N. "The Case for Technology Investments in the Environment" *Nature Comment*, 19 December. 2017. Web. https://www.nature.com/articles/d41586-017-08675-7. Accessed 10 Feb. 2018.

Keller, Evelyn Fox. *A Feeling for the Organism, 10th Anniversary Edition: The Life and Work of Barbara McClintock*. Basingstoke: Macmillan, 1984. Print.

Kim, Ke Chung, and Loren B. Byrne. "Biodiversity Loss and the Taxonomic Bottleneck: Emerging Biodiversity Science." *Ecological Research*, 21.6 (2006): 794. Print.

Kwa, C. "Modeling the Grasslands." *Historical Studies in the Physical and Biological Sciences*, vol. 24, no. 1 (1993): 125-155. Print.

Latour, Bruno. "Some Advantages of the Notion of "Critical Zone" for Geopolitics." *Procedia Earth and Planetary Science*, vol. 10 (2014): 3-6. Print.

Latour, Bruno. *Politics of Nature*. Harvard University Press, 2004.

Machine Bias Series, Propublica, 2018. Web. https://www.propublica.org/series/machine-bias. Accessed 15 Jan. 2018.

Madison, Mark Glen. "'Potatoes Made of Oil': Eugene and Howard Odum and the Origins and Limits of American Agroecology." *Environment and History* vol. 3, no. 2 (1997): 209-238. Print.

Mattern, Shannon. "A City Is Not a Computer." *Places Journal* (2017). Web. https://placesjournal.org/article/a-city-is-not-a-computer/. Accessed 1 Mar 2018.

Mattern, Shannon. *Code and Clay, Data and Dirt: Five Thousand Years of Urban Media*. Minneapolis: University of Minnesota Press, 2017. Print.

McCarthy, John and Feigenbaum, Edward A., "Arthur Samuel: Pioneer in Machine Learning," *AI Magazine*, vol. 11, no. 3 (1990): 10-11. Print.

Merlin, *The Cornell Lab*. Web. http://merlin.allaboutbirds.org/. Accessed 10 Mar. 2018.

Molina, Mario J., and F. Sherwood Rowland. "Stratospheric Sink for Chlorofluoromethanes: Chlorine Atom-catalysed Destruction of Ozone." *Nature* 249.5460 (1974): 810. Print.

Morton, Timothy. *The Ecological Thought*. Cambridge, Mass.: Harvard University Press, 2010. Print.

Odell, Jenny. "Notes of a Bioregional Interloper." *Open Space*, San Francisco Museum of Modern Art (2017). Web. https://openspace.sfmoma.org/2017/10/notes-of-a-bioregional-interloper/. Accessed 10 Feb. 2018.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books, 2017. Print.

Scott, James C. *Seeing like a State: How Certain Schemes to Improve the Human Condition have Failed*. New Haven, Connecticut: Yale University Press, 1998. Print.

Simpson, I. R., and P. D. Jones. "Analysis of UK Precipitation Extremes Derived from Met Office Gridded Data." *International Journal of Climatology* vol. 34 no. 7 (2014): 2438-2449. Print.

Stengers, Isabelle. "The Cosmopolitical Proposal." *Making Things Public: Atmospheres of Democracy* (2005): 994. Print.

Szerszynski, Bronislaw, and Maialen Galarraga. "Geoengineering Knowledge: Interdisciplinarity and the Shaping of Climate Engineering Research." *Environment and Planning A* 45.12 (2013): 2817-2824. Print.

Tansley, Arthur G. "The Use and Abuse of Vegetational Concepts and Terms." *Ecology* vol. 16, no. 3, 1935: 284-307. Print.

Temple, James. "Harvard Scientists Moving Ahead on Plans for Atmospheric Geoengineering Experiments." *Technol Rev* vol. 24 (2017).

The Human Project. Web. https://www.thehumanproject.org/. Accessed 10 Feb. 2018.

Tsing, Anna Lowenhaupt. *The Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins*. Princeton, NJ: Princeton University Press, 2015.

Vergès, Françoise. "Racial Capitalocene: Is the Anthropocene Racial?" Verso Blog, 30 August. 2017. Web. https://www.versobooks.com/blogs/3376-racial-capitalocene. Accessed 10 Feb. 2018.

Vitello, Paul. "Joseph Farman, 82, Is Dead; Discovered Ozone Hole." *New York Times*, May 18, 2013. Web. https://www.nytimes.com/2013/05/19/science/earth/joseph-farman-82-is-dead-discovered-ozone-hole.html. Accessed 10 Feb. 2018.

Wiener, Norbert. *Cybernetics: Control and Communication in the Animal and the Machine*. New York: Wiley, 1948. Print.

Posted in Journal Issues, Research Values

“Raw Data” Is an Oxymoron

Edited by Lisa Gitelman

The MIT Press
Cambridge, Massachusetts
London, England

Contents

Introduction

Lisa Gitelman and Virginia Jackson

"Raw data" is both an oxymoron and a bad idea.
—Geoffrey C. Bowker, *Memory Practices in the Sciences*

Data are everywhere and piling up in dizzying amounts. Not too long ago storage and transmission media helped people grapple with kilobytes and megabytes, but today's databases and data backbones daily handle not just terabytes but petabytes of information, where *peta-* is a prefix which denotes the unfathomable quantity of a quadrillion, or a thousand trillion. Data are units or morsels of information that in aggregate form the bedrock of modern policy decisions by government and nongovernmental authorities. Data underlie the protocols of public health and medical practice, and data undergird the investment strategies and derivative instruments of finance capital. Data inform what we know about the universe, and they help indicate what is happening to the earth's climate. "Our data isn't just telling us what's going on in the world," IBM advertises; "it's actually telling us where the world is going." The more data the better, by these lights, as long as we can process the accumulating mass. Statisticians are on track to be the next sexy profession in the digital economy, reports the front page of the *New York Times*. "Math majors, rejoice," the newspaper urges in another instance, because businesses are going to need an army of mathematicians as they grapple with increasing mountains of data.[1]

What about the rest of us? What are we to data and data to us? As consumers we tend to celebrate our ability to handle data in association with sophisticated technology. My iPad has 64 gig! My phone is 4G! We don't always know what this means and typically don't know how these devices actually function, but they are "friendly" to users in part according to the ways they empower us to store, manipulate, and transmit data.

Yet if data are somehow subject to us, we are also subject to data, because Google collects so much information on users' interests and behaviors, for instance, and the U.S. National Security Agency mines fiber-optic transmissions for clues about terrorists. Not too long ago it was easier to understand the ways that data was collected about us, first through the institutions and practices of governmentality—the census, the department of motor vehicles, voter registration—and then through the institutions and practices of consumer culture, such as the surveys which told us who we were, the polls which predicted who we'd elect, and the ratings which measured how our attention was being directed. But today things seem different—in degree if not always in kind—now that every click, every move has the potential to count for something, for someone somewhere somehow. Is data about you *yours*, or should it be, now that data collection has become an always-everywhere proposition? Try to spend a day "off the grid" and you'd better leave your credit and debit cards, transit pass, school or work ID, passport, and cell phone at home—basically, anything with a barcode, magnetic strip, RFID, or GPS receiver.[2]

In short, if World War II helped to usher in the era of so-called Big Science, the new millennium has arrived as the era of Big Data.[3] For this reason, we think a book like *"Raw Data" Is an Oxymoron* is particularly timely. Its title may sound like an argument or a thesis, but we want it to work instead as a friendly reminder and a prompt. Despite the ubiquity of the phrase *raw data*—over seventeen million hits on Google as of this writing—we think a few moments of reflection will be enough to see its self-contradiction, to see, as Bowker suggests, that data are always already "cooked" and never entirely "raw." It is unlikely that anyone could disagree, but the truism no more keeps us from valuing data than a similar acknowledgment keeps up from buying jumbo shrimp. The analogy may sound silly, but not as silly as it first appears: just as the economy of shrimp and shrimping has shifted radically in the decades since the birth of industrial aquaculture in the 1970s, so the economy of data has an accelerated recent history. The essays in this volume do not present one argument about that economy, but they do begin to supply a little heretofore-unwritten history for the seismic shift in the contemporary conception and use—the sheer existence—of so much data.

However self-contradicting it may be, the phrase *raw data*—like *jumbo shrimp*—has understandable appeal. At first glance data are apparently before the fact: they are the starting point for what we know, who we are, and how we communicate. This shared sense of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself. If we're not careful, in other words, our zeal for more and more data can become a faith in their neutrality and autonomy, their objectivity. Think of the ways people talk and write about data. Data are familiarly "collected," "entered," "compiled," "stored," "processed," "mined," and "interpreted." Less obvious are the ways in which the final term in this sequence—interpretation—haunts its predecessors. At a certain level the collection and management of data may be said to presuppose interpretation. "Data [do] not just exist," Lev Manovich explains, they have to be "generated."[4] Data need to be imagined *as* data to exist and function as such, and the imagination of data entails an interpretive base.

Here another analogy may be helpful. Like *events* imagined and enunciated against the continuity of time, *data* are imagined and enunciated against the seamlessness of phenomena. We call them up out of an otherwise undifferentiated blur. If events garner a kind of immanence by dint of their collected enunciation, as Hayden White has suggested, so data garner immanence in the circumstances of their imagination.[5] Events produce and are produced by a sense of history, while data produce and are produced by the operations of knowledge production more broadly. Every discipline and disciplinary institution has its own norms and standards for the imagination of data, just as every field has its accepted methodologies and its evolved structures of practice. Together the essays that comprise *"Raw Data" Is an Oxymoron* pursue the imagination of data. They ask how different disciplines have imagined their objects and how different data sets harbor the interpretive structures of their own imagining. What are the histories of data within and across disciplines? How are data variously "cooked" within the varied circumstances of their collection, storage, and transmission? What sorts of conflicts have occurred about the kinds of phenomena that can effectively—can ethically—be "reduced" to data?

Treating data as a matter of disciplines—rather than of computers, for instance—may seem curious at first. The subject of data is bound to alienate students and scholars in disciplines within the humanities particularly. Few literary critics want to think of the poems or novels they read as "data," and for good reason. The skepticism within literary studies about Franco Moretti's "distant reading" approach, which in part reduces literary objects to graphs, maps, and other data visualizations, testifies to the resistance the notion of literature as data might provoke. Similarly, many historians would not like to reduce their subjects to abstract objects useful in the production of knowledge about the past. Their reluctance was evidenced by the hostile reception accorded to cliometrics in the 1960s and it persists today. In some sense, data are precisely *not* the domain of humanistic inquiry. Yet we propose that students and scholars in the humanities do worry about data, broadly speaking, to the extent that they worry about how their objects of study have been assumed as well as discerned. Don't all questions presuppose or delimit their answers to some degree? Recent work in historical epistemology has challenged the status of the research object, or as Michel Foucault would have it, has raised questions about the boundaries of the archive, about the form, appearance, and regularity of the statements and practices available to us in knowing what we know.[6] When we put our own critical perspectives into historical perspective, we quickly find that there is no stance detached from history, which is to say that there is no persistently objective view.

The conditions of evolving, possessing, and assessing knowledge turn out to be remarkably available to cultural and historical change. The field of science studies has pursued this observation in the greatest detail, and *"Raw Data" Is an Oxymoron* is inspired by science studies while directed beyond it to a broader audience. Evolved over the same decades as other "studies"—like area studies, ethnic studies, cultural and media studies—science studies takes as its object the work of scientists and engineers.[7] The field has helped to confound simplistic dichotomies like theory/practice and science/society in a rich, diverse body of work that, among other things, has explored the situated, material conditions of knowledge production. Looking at the ways scientific knowledge is produced—rather than innocently "discovered," for instance—resembles our project of looking into data or, better, looking *under* data to consider their root assumptions.[8] Inquiries such as these may be seen as contributions toward a critique of objectivity. The point of such a critique—we must quickly emphasize—is not that objectivity is *bad* or that objectivity is mythical. Any such claim must depend, as Lorraine Daston and Peter Galison note, on first achieving a careful understanding of "what objectivity *is*."[9] The point is not how to judge whether objectivity is possible—thumbs up or thumbs down—but how to describe objectivity in the first place. Objectivity is situated and historically specific; it comes from somewhere and is the result of ongoing changes to the conditions of inquiry, conditions that are at once material, social, and ethical.

The very idea of objectivity as the abnegation, neutrality, or irrelevance of the observing self turns out to be of relatively recent vintage. Joanna Picciotto has recently suggested that "the question raised by objectivity is how innocence, traditionally understood to be a state of ignorance, ever came to be associated with epistemological privilege."[10] As a moment in which we can see the emergence of a modern privileging of objectivity, Picciotto nominates "the seventeenth century's conversion of the original subject of innocence, Adam, into a specifically intellectual exemplar. Used to justify experimental science, an emergent public sphere, and the concept of intellectual labor itself," Adam became emblematic of "a new ideal of estranged and productive observation."[11] This means that Milton's *Paradise Lost* and *Paradise Regain'd* may be as important to the development of experimental science as the invention of the microscope.

The innocent observer has had a long, diverse career. Looking at scientific atlases, not Milton poems, Daston and Galison discern the arrival of a version of objectivity that is mechanical: characterized by the observer's restraint and distinguishable from other versions in which the skill and discernment of the observing self counts for something, such as cases in which knowledgeable observers idealize multiple, idiosyncratic specimens into a single type, or in which practiced diagnosticians exert trained judgment in order to make sense of blurry scans. Mechanical objectivity emerged as a dominant ideal in the sciences only in the middle of the nineteenth century, and it is perhaps simplest to describe it contextually with reference to the development of photography during those same years. When Louis Daguerre, Henry Fox Talbot, and others developed and then popularized the first photographic processes, observers were struck by the apparent displacement of human agency in the production of lifelike images. Fox Talbot's lavish account of his calotype process captures this displacement in its title, *The Pencil of Nature*. No artist necessary. Light itself is enough. Photography is objective.

David Ribes and Steven Jackson (chapter 8) direct attention toward some of the difficulties that mechanical objectivity presents today in scientific practice, when biologists rely upon data collected by remote sensors. But mechanical objectivity was something of a conundrum even in Fox Talbot's day. From the very first, the mechanical objectivity of photography was framed by a counter discourse in which photographers were praised for their ability to capture "inner" or "higher" truths on film. The pencil of nature is not enough. Artists are necessary. Photography is subjective. This isn't a question of *either/or* as much as a matter of *and yes*: mechanical objectivity is an "epistemic virtue" among other competing virtues.[12] The presumptive objectivity of the photographic image, like the presumptive rawness of data, seems necessary somehow—resilient in common parlance, utile in commonsense—but it is not sufficient to the epistemic conditions that attend the uses and potential uses of photography. At the very least the photographic image is always framed, selected out of the profilmic experience in which the photographer stands, points, shoots. Data too need to be understood as framed and framing, understood, that is, according to the uses to which they are and can be put. Indeed, the seemingly indispensable misperception that data are ever raw seems to be one way in which data are forever contextualized—that is, framed—according to a mythology of their own supposed decontextualization.

Thus the history of objectivity turns out to be inescapably the history of subjectivity, of the self,[13] and something of the same thing must hold for the concept of data. Data require our participation. Data need us. Yet for all of the suggestive parallels, the history of objectivity is not the history of data. Where did the modern concept of data come from? The first two chapters in this volume tackle this question in different ways. In "Data before the Fact" (chapter 1), Daniel Rosenberg plumbs the derivation and use of *datum* (the singular form) and *data*, offering an intellectual history of the concept that stretches back to the Enlightenment, before the virtue of mechanical objectivity had fully taken shape. Rosenberg is aided in his study—if also provoked—by a new set of tools that offer ways to find and visualize patterns within the digitized corpus of Western printed thought. He gives us the data on data, as it were. Travis D. Williams heads even further back in time, to the Renaissance, in order to consider the history behind one of the strongest epistemic conditions shaping the contemporary data imaginary: the self-evidence of numbers and arithmetic fact as such. Previous scholars have rendered the history of math as or relating to a pre-history of capitalism, and Williams's "Procrustean Marxism and Subjective Rigor" (chapter 2) seeks an additional path, giving an account of English math books with their hilariously prosaic story problems. Like Rosenberg's self-conscious use of present tools in rendering the past, Williams is at pains to take early modern math on its own terms while also considering just what such an endeavor means, since the terms of math are supposed to be universal in time and space. Two plus two equals four, always and everywhere, and "Numbers never lie."

No two chapters could exhaust the multiple origins of data as a concept; Rosenberg and Williams only open the question in different ways. The association of data with diagrams and graphs, in the first instance, and with numbers and mathematical functions, in the second, leads us to the general precept that *data are abstract*. While this quality can make it hard to think or write about data in general—that is, in the abstract—it follows from their abstraction that data ironically require material expression. The retention and manipulation of abstractions require stuff, material things. Just as Cambridge University could become a training ground for mathematical physics only after the introduction of written exams at the end of the eighteenth century (paper and pencil are the things of things where modern abstractions are concerned), so the contemporary era of Big Data has been enabled by the widespread availability of electronic storage media, specifically mainframe computers, servers and server farms, and storage area networks.[14] Both the scale and ontology of electronic storage pose an interesting challenge across the humanities, where lately there has been a renewed interest in *things*.[15] Indeed, as Wendy Hui Kyong Chun has observed, this current scholarly interest in things or "thing theory" needs to be seen against the context of digital media within which things "always seem to be disappearing" in such crucial ways.[16] What sort of things are electronic data, after all?

As we suggested earlier, one productive way to think about data is to ask how different disciplines conceive their objects, or, better, how disciplines and their objects are mutually conceived. The second pair of chapters in this volume takes that tack. In "From Measuring Desire to Quantifying Expectations" (chapter 3), Kevin R. Brine and Mary Poovey address the discipline of economics, and in "Where Is That Moon, Anyway?" (chapter 4), Matthew Stanley considers astronomy. Brine and Poovey follow the work of Irving Fisher, the twentieth-century economist who created the scaffolding for today's financial modeling by linking capital to the concept of present value, which calculates value by taking into account expectations about future yields or benefits. Although the data he used needed to be "scrubbed" to be usable, models like those that Fisher created continue to be influential because they claim a basis that is situated as the objective source of information it can never actually be. As Rosenberg's history helps us understand, this fundamental contradiction may actually be intrinsic to the concept of data, since "the semantic function of data is specificall*y rhetorical*." Data by definition are "that which is given prior to argument," given in order to provide a rhetorical basis. (Facts are facts—that is, they are true by dint of being factual—but data can be good or bad, better or worse, incomplete and insufficient.) Yet precisely because data stand as a given, they can be taken to construct a model sufficient unto itself: given certain data, certain conclusions may be proven or argued to follow. Given other data, one would come to different arguments and conclusions.

Disciplines operate according to shared norms, and data scrubbing is an accepted and unexceptional necessity in economics and finance. Disciplines also operate by dint of "data friction"—Paul Edwards's term—friction consisting of worries, questions, and contests that assert or affirm what should count as data, or which data are good and which less reliable, or how big data sets need to be.[17] Stanley's chapter offers a fascinating example of data friction in the field of astronomy. In efforts to derive a particular lunar constant—called the secular acceleration—astronomers have repeatedly engaged in research that on its face seems a lot less like astronomy than it does textual analysis, history, and psychology: poring over the works of classical authors to evaluate their accounts of solar eclipse. The apparent intrusion of psychology into astronomy, or history into climate science, or bibliography into botany—to mention additional examples recently documented—serves as a reminder of just how diverse and dynamic disciplines are.[18] Disciplines aren't just separate subjects you pick out of a course catalogue. They involve infrastructures comprised of "people, artifacts, and institutions that generate, share, and maintain specific knowledge" in complex and interconnected ways.[19] The bodies of knowledge made and maintained by the professions can be more or less specific than those of academic disciplines, but they involve related infrastructures and a similarly evolved and evolving "trust in numbers."[20]

Data aren't only or always numerical, of course, but they do always exist in number in the sense that data are particulate or "corpuscular, like sand or succotash." Something like information, that is, data exist in little bits.[21] This leads us to a second general precept, that *data are aggregative*. They pile up. They are collected in assortments of individual, homologous data *entries* and are accumulated into larger or smaller data *sets*. This aggregative quality of data helps to lend them their potential power, their rhetorical weight. (More is better, isn't it?) Indeed, data are so aggregative that English usage increasingly makes many into one. The word *data* has become what is called a mass noun, so it can take a singular verb. Sentences that include the phrase "data is . . ." are now roughly four times as common (on the web, at least, and according to Google) as those including "data are . . ." despite countless grammarians out there who will insist that *data* is a plural. So far in this introduction we have been assiduous in using the word *data* with plural verbs, and some readers may already have sensed the strain. Data's odd suspension between the singular and the plural reminds us of what aggregation means. If a central philosophical paradox of the Enlightenment was the relation between the particular and the universal, then the imagination of data marks a way of thinking in which those principles of logic are either deferred or held at bay. The singular *datum* is not the particular in relation to any universal (the elected individual in representative democracy, for example) and the plural *data* is not universal, not generalizable from the singular; it is an aggregation. The power within aggregation is relational, based on potential connections: network, not hierarchy.

To be sure, data also depend upon hierarchy. Part of what distinguishes data from the more general category, information, is their discreetness. Each datum is individual, separate and separable, while still alike in kind to others in its set. It follows that the imagination of data is in some measure always an act of classification, of lumping and splitting, nesting and ranking, though the underlying principles at work can be hard to recover. Once in place, classification schemes are notoriously difficult to discern and analyze, since "Good, usable systems disappear almost by definition. The easier they are to use, the harder they are to see."[22] This is the provocation animating an important book by Bowker and Susan Leigh Star entitled *Sorting Things Out*. Working with a group of examples—such as classifying causes of death; classifying the labor of healthcare workers; and classifying race in apartheid-era South Africa—Bowker and Star illuminate the ways that classifications function, for good and ill, to underpin the social order. When phenomena are variously reduced to data, they are divided and classified, processes that work to obscure—or *as if* to obscure—ambiguity, conflict, and contradiction.

Today the ubiquitous structures of data aggregation are computational forms called relational databases. Described and developed since 1970, relational databases organize data into separate tables ("relational variables") in such a way that new data and new kinds of data can be added or subtracted without making the earlier arrangement obsolete. Data are effectively made independent of their organization, and users who perform logical operations on the data are thus "protected" from having to know how the data have been organized.[23] The technical and mathematical details are not important here, but imagine sorting a giant stack of paperwork into separate bins. Establishing which and how many bins are appropriate would be your first important task, but it is likely that as you proceed to sort your papers, you will begin to have a nagging sense that different bins are needed, or that some bins should be combined, or that some papers impossibly belong in multiple bins. You may even wind up with an extra bin or two marked "miscellaneous" or "special problems." It is just this sort of tangle that database architecture seeks to obviate while making relational variables (bins) and their data (papers) available to a multiplicity of desirable logical operations, like queries.

The third pair of chapters in this volume, "facts and FACTS" by Ellen Gruber Garvey (chapter 5) and "Paper as Passion" by Markus Krajewski (chapter 6), takes our paperwork metaphor at face value. Each imagines a different prehistory of the database by considering a specific trove of paper. Garvey describes a giant mass of clippings taken from Southern newspapers to document the horrors of slavery in the antebellum United States, while Krajewski describes the enormous file amassed in the twentieth century by the German sociologist and prolific theorist Niklas Luhmann. Two examples could hardly exhaust the possible prehistories of databases—papery and not—which reach at least as far back as early modern note-taking practices and the accompanying sense of what can anachronistically be called "information overload" that together led to giant compendia with elaborate finding aids.[24] Yet Garvey's example comes from that important moment when the concept of information—close relative of data—finally emerged in something like its present form, as the alienable, abstract contents of an *inform*ative press,[25] while Krajewski's example comes from the equally important moment of systems theory and cybernetics in the second half of the twentieth century.

Garvey's trick, or rather, the trick of the Grimké sisters she writes about, is to fix on an instance where information collected in one locale can take on wholly different meanings in another, as advertisements for runaway slaves become data in the argument against slavery. This is fully remaking the power of the press in the user-dimension, where users may differ in locale if also in their gender, race, and politics. Krajewski by contrast addresses a single user, Niklas Luhmann, who is famous in some quarters for working from his own huge and all-encompassing card index. Author of more than forty books—not a few of them considered "difficult"—Luhmann developed his systems theory, Krajewski suggests, because of, out of, and in collaboration with his card index, a sort of paper machine—a system—for remembering and for generating thought. Papery databases are only metaphorically databases, of course, yet the example of Luhmann's card index helps to clarify the extraordinary generative power that data aggregation can possess while also raising the question of the human or—one must wonder—the posthuman, the human-plus-machine/machine-plus-human hybrids that living with computers make increasingly integral to our understanding.

The final pair of chapters, "Dataveillance and Countervailance" by Rita Raley (chapter 7) and "Data Bite Man" by David Ribes and Steven J. Jackson (chapter 8), pursues the question of data in the present day. Readers will be challenged to think in some detail about the kinds of data being collected about them today, and they will be challenged to consider the difficulties that scientists and policy makers confront when they try to make data useful today and also reusable potentially by others in the future. What are the logics and the ethics of "dataveillance," now that we appear to be moving so rapidly from an era of expanding data resources into an era in which we have become the resource for data collection that vampirically feeds off of our identities, our "likes," and our everyday habits? If while using the Internet we click on a book or a pair of shoes at Amazon.com, or in a box to sign a petition to stop a Congressional bill, or on a link to a porn website, or on a Google Books page or on an online map to find directions, are we making a choice or are we giving Amazon and the federal government and the pornographers (and the security agencies trolling them) and their advertisers ways to guide our choices, calculate our votes, or put us in jail? Both, Raley answers, and suggests that activist projects that exploit dataveillance—that do not opt out but instead "insist on a near-total inhabitation of the forcible frame"—might stand the best chance of at least offering an immanent critique of the predicament that we have created and now must find a way to inhabit.

Ribes and Jackson address the predicament experienced by today's scientists, who must not only collect and analyze data but also make sure their data remain useable over the life of a research program and beyond, available to readers of resulting publications as well as for potential research in the future. A recent survey confirms that researchers across the sciences are dealing with vast quantities of data (a fifth report generating data sets of 100 gigabytes or more) while at the same time lacking the resources to preserve that data sensibly (four fifths acknowledge insufficient funding for data curation).[26] Ribes and Jackson show the surprising complexities in something as apparently simple as collecting water samples from streams, while they challenge readers to think of scientists and their data as evolved and evolving symbionts, mutually dependent species adapted amid systems ecological and epistemic.

There is much more in the essays collected here than this introduction has mentioned or could encapsulate, and we hope that readers will consider as they read what the ideas are that emerge across the essays as well as what gaps there are among them. One omission, certainly, which this Introduction accentuates with its brief attention to English usage and the history of concepts, is any account of non-Western contexts or intercultural conjunctions that might illuminate and complicate data past and present. How have non-Western cultures arrived at data and allied concepts like information and objectivity? How have non-Western cultures been subject to data, in the project of colonialism, for example, or otherwise? Indeed, how are data putatively raw—and not—in non-Anglophone contexts? Do other languages deploy the food metaphor that English does? Do their speakers semantically align supposedly raw data with supposedly raw text (that is, ASCII) and supposedly raw footage (unedited film or video) the way that English speakers do? How do different languages differently resolve the dilemma of singular and plural? No collection of essays could exhaust the subject of data, of course, and that is one reason we earlier called our title a prompt rather than an argument. The authors collected in *"Raw Data" Is an Oxymoron* all hope to open the question of data, to model some of the ways of thinking about data that seem both interesting and productive, as well as to encourage further discussion. The ethics surrounding the collection and use of today's "Big Data" are a particularly pressing concern.[27]

As an additional gesture toward further discussion, we include a brief section of color images, most of them selected and described by additional contributors. The images in this color insert extend the types of data considered in this volume—some in challenging ways—while some of them also broach the important subject of representation and, more specifically, data visualization, which is not always addressed directly in the chapters that follow but which haunts them nonetheless. As the neologism "dataveillance" suggests, data provide ways to survey the world (the noun *surveillance* is related to *survey*), yet it is important to remember that surveying the world with data at some level means having data visibly before one's eyes, looking *through* the data if not always self-consciously looking *at* the data. There is then a third and final precept closely related to the other two. Not only are data abstract and aggregative, but also *data are mobilized graphically*. That is, in order to be used as part of an explanation or as a basis for argument, data typically require graphical representation and often involve a cascade of representations.[28] Any interface is a data visualization of sorts—think of how many screens you encounter every day—and so are spreadsheets, charts, diagrams, and other graphical forms. Data visualization amplifies the rhetorical function of data, since different visualizations are differently effective, well or poorly designed, and all data sets can be multiply visualized and thereby differently persuasive.

More than a few contemporary visual artists make obvious the rhetoric of data visualization: Jenny Holzer's LED feeds of poems in the place of stock quotes or headlines and "truisms" in the place of public information, for instance, confront spectators with variations on the data frames they face every day. Like the digital network, the database is an already rich and still emerging conceptual field for artwork, while a varied and variously evocative "database aesthetics" demonstrates—as we hope the chapters in this collection make clear—that recognizing the power of data visualization is an important part of living with data.[29]

Notes

1. Geoffrey C. Bowker, *Memory Practices in the Sciences* (Cambridge, MA: MIT Press, 2005), 184. This is an IBM advertising campaign from 2009 to 2010. *New York Times*, August 5, 2009, and May 13, 2011.

2. For more on data obfuscation generally, see Finn Brunton and Helen Nissenbaum, "Vernacular Resistance to Data Collection and Analysis: A Political Theory of Obfuscation," *First Monday* 16, no. 5 (May 2, 2011). The question of whether data about you is yours came before the U.S. courts in the form of a question about privacy: whether the police need a warrant to attach a

GPS device to your car and then monitor your movements. According to *United States v. Jones* (2012), they do.

3. On the bigness of data, see, for instance, Lev Manovich, "Trending: The Promises and the Challenges of Big Social Data," http://lab.softwarestudies.com/2011/04/new-article-by-lev-manovich-trending.html (accessed June 20, 2011). For an example linking big science and big data, see Peter Galison, *Image and Logic: A Material Culture of Microphysics* (Chicago: University of Chicago Press, 1997).

4. Lev Manovich, *The Language of New Media* (Cambridge, MA: MIT Press, 2001), 224.

5. See Hayden White, *Metahistory: The Historical Imagination in Nineteenth-Century Europe* (Baltimore, MD: Johns Hopkins University Press, 1975).

6. Franco Moretti, *Graphs, Maps, Trees: Abstract Models for Literary History* (London: Verso, 2005); and Michel Foucault, *The Archaeology of Knowledge & The Discourse on Language* (New York: Vintage, 1982), part 3 (French eds. 1969, 1971).

7. See Mario Biagioli, "Postdisciplinary Liaisons: Science Studies and the Humanities," *Critical Inquiry* 35 (Summer 2009): 816–833; and Mario Biagioli, ed., *The Science Studies Reader* (New York: Routledge, 1999).

8. Looking under is a gesture of "infrastructural inversion" within the sociology of knowledge; see Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999), 34–36.

9. Lorraine Daston and Peter Galison, *Objectivity* (New York: Zone Books, 2007), 51. On critique itself, see Bruno Latour, "Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern," *Critical Inquiry* 30 (Winter 2004): 225–248.

10. Joanna Picciotto, *Labors of Innocence in Early Modern England* (Cambridge, MA: Harvard University Press, 2010), 1.

11. Ibid., 2–3.

12. Daston and Galison, *Objectivity*, 27.

13. Ibid., 37.

14. Andrew Warwick, *Masters of Theory: Cambridge and the Rise of Mathematical Physics* (Chicago: University of Chicago Press, 2003), chap. 3.

15. For instance, Bill Brown, "Thing Theory," *Critical Inquiry* 28 (Autumn 2001): 1–22; Lorraine Daston, ed., *Things That Talk: Object Lessons from Art and Science* (New York: Zone Books, 2004); Lorraine Daston, ed., *Biographies of Scientific Objects* (Chicago: University of Chicago Press, 2000); Hans-Jörg Rheinberger, *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube* (Stanford, CA: Stanford University Press, 1997).

16. Wendy Hui Kyong Chun, *Programmed Visions: Software and Memory* (Cambridge, MA: MIT Press, 2011), 11. "Thing Theory" is Bill Brown's title (see note 15).

17. Paul N. Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (Cambridge, MA: MIT Press, 2010), xiv.

18. On climate science as a form of history, see Edwards, *A Vast Machine*, xvii; on botany and bibliography, see Lorraine Daston, "Type Specimens and Scientific Memory," *Critical Inquiry* 31 (Autumn 2004): 153–182, esp. 175.

19. See Edwards, *A Vast Machine*, 17.

20. The phrase comes from a title by Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, NJ: Princeton University Press, 1995), which we recommend (along with works already cited) for readers who wish more prolonged exposure to the kinds of questions introduced here.

21. Geoffrey Nunberg, "Farewell to the Information Age," in *The Future of the Book*, ed. Geoffrey Nunberg (Berkeley: University of California Press, 1996), 117.

22. Bowker and Star, *Sorting Things Out*, 33.

23. E. F. Codd, "A Relational Model for Large Shared Data Banks," *Communications of the ACM* 13, no. 6 (June 1970): 377–387. Alan Liu led us to Codd; see his *Local Transcendence: Essays on Postmodern Historicism and the Database* (Chicago: University of Chicago Press, 2008), 239–262.

24. See Ann M. Blair, *Too Much to Know: Managing Scholarly Information Before the Modern Age* (New Haven, CT: Yale University Press, 2010); and Daniel Rosenberg, "Early Modern Information Overload," *Journal of the History of Ideas* 64 (January 2003): 1–9.

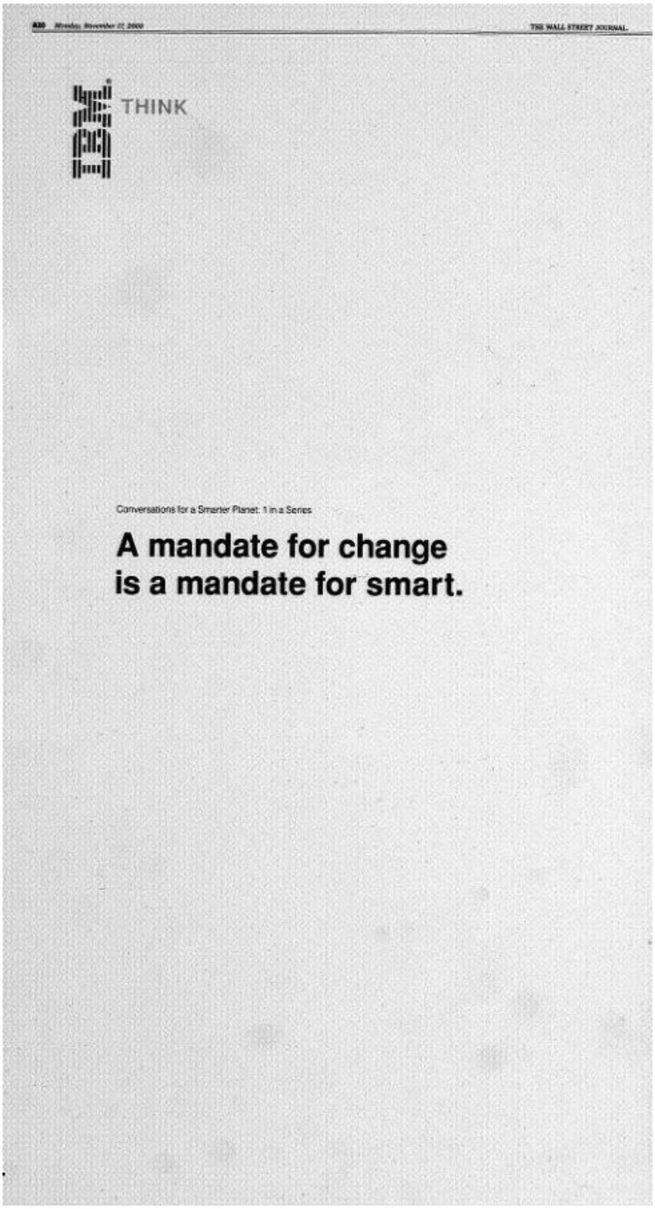
25. Nunberg, *Farewell*, 110–111.

26. See "Challenges and Opportunities," *Science* 331 (February 11, 2011): 692–693.

27. See danah boyd and Kate Crawford, "Six Provocations for Big Data," paper presented at "A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society," Oxford Internet Institute, September 21, 2011; and Jay Stanley, "Eight Problems with 'Big Data,'" *ACLU.org*, April 25, 2012.

28. On mobilization and cascades, see Bruno Latour, "Drawing Things Together," *Representation in Scientific Practice*, ed. Michael Lynch and Steve Woolgar (Cambridge, MA: MIT Press, 1990), 19–68; on the effectiveness of visualizations, see, for instance, Edward Tufte, *The Visual Display of Quantitative Information*, 2nd ed. (Cheshire, CT: Graphics Press, 2001).

29. For an overview, see, for instance, Victoria Vesna, ed., *Database Aesthetics: Art in the Age of Information Overflow* (Minneapolis: University of Minnesota Press, 2007).



A20 Monday, November 17, 2008

THE WALL STREET JOURNAL.

Conversations for a Smarter Planet: 1 in a Series

A mandate for change
is a mandate for smart.

International Business Machines Corporation (IBM). "Smarter Planet" advertisement campaign, 2008. Detail. From *Wall Street Journal*, 17 November 2008.

The Smartness Mandate: Notes toward a Critique

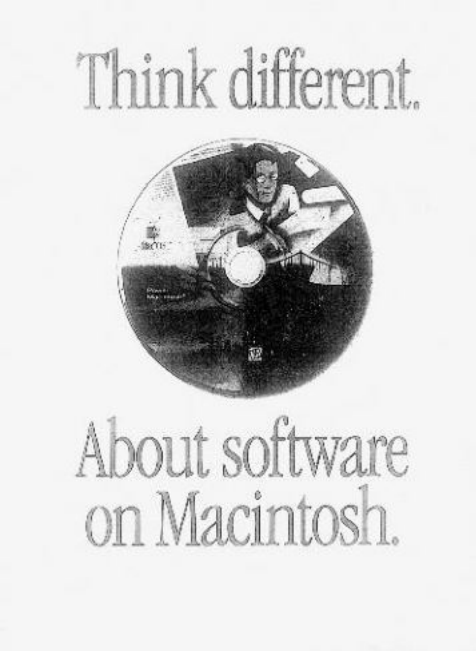
ORIT HALPERN, ROBERT MITCHELL,
AND BERNARD DIONYSIUS GEOGHEGAN

On November 6, 2008, still in the immediate aftermath of the worldwide economic crisis initiated by the U.S. subprime mortgage market collapse, then chair of IBM Sam Palmisano delivered a speech at the Council on Foreign Relations in New York City. The council is one of the foremost think tanks in the United States, its membership comprising senior figures in government, the intelligence community (including the Central Intelligence Agency), business leaders, financiers, lawyers, and the media. Yet Palmisano was not there to discuss the fate of the global economy. Rather, he introduced his corporation's vision of the future in a talk titled "A Smarter Planet." In glowing terms, Palmisano laid out a vision of fiber-optic cables, high-bandwidth infrastructure, seamless supply-chain and logistical capacity, a clean environment, and eternal economic growth, all of which were to be the preconditions for a "smart" planet. IBM, he argued, would lead the globe to the next frontier, a network beyond social networks and mere Twitter chats. This future world would come into being through the integration of human beings and machines into a seamless "Internet of things" that would generate the data necessary for organizing production and labor, enhancing marketing, facilitating democracy and prosperity, and—perhaps most important—for enabling a mode of automated, and seemingly apolitical, decision-making that would guarantee the survival of the human species in the face of pressing environmental challenges. In Palmisano's talk, "smartness" named the interweaving of dynamic, emergent computational networks with the goal of producing a more resilient human species—that is, a species able to absorb and survive environmental, economic, and security crises by means of perpetually optimizing and adapting technologies.[1]

Palmisano's speech was notable less for its content, which to a degree was an amalgamation of existing claims about increased bandwidth, complexity, and ecological salvation, than for the way in which its economic context and its planetary terminology made explicit a hitherto tacit political promise that has attended the rise of "smart" technologies. Though IBM had capitalized for decades on terms associated with intelligence and thought—its earlier trademarked corporate slogan was "Think"—*smart* was by 2008 an adjective attached to many kinds of computer-mediated technologies and places, including phones, houses, cars, classrooms, bombs, chips, and cities. Palmisano's "smarter planet" tagline drew on aspects of these earlier invocations of smartness, and especially the notion that smartness required an extended infrastructure that produced an environment able to automate many human processes and respond in real time to human choices. His speech also underscored that smartness demanded an ongoing penetration of computing into infrastructure to mediate daily perceptions of life. (Smartphones, for example, are part of a discourse in which the world is imagined as networked, interactive, and constantly accessible through technological interfaces, their touch screens enabled by an infrastructure of satellite networks, server farms, and cellular towers, among many other structures that facilitate the regular accessing of services, goods, and spatial location data.) But as Palmisano's speech made clear, these infrastructures now demanded an "infrastructural imaginary"—an orienting telos about what smartness is and does. This imaginary redefined no less than the relationships among technology, human sense perception, and cognition. With this extension of smartness to both the planet and the mind, what had been a corporate tagline became a governing project able to individuate a citizen and produce a global polity.

This new vision of smartness is inextricably tied to the language of crisis, whether a financial, ecological, or security event. But where others might see the growing precariousness of human populations as best countered by conscious planning and regulation, advocates of smartness instead see opportunities to decentralize agency and intelligence by distributing it among objects, networks, and life forms. They predict that environmentally extended smartness will take the place of deliberative planning, allowing resilience in a perpetually transforming world. Palmisano proposed "infus[ing] intelligence into decision making" itself.[2] What Palmisano presented in 2008 as the mandate of a single corporation is central to much contemporary design and engineering thinking more generally.

We call these promises about computation, complexity, integration, ecology, and crisis "the smartness mandate." We use this phrase to mark the fact that the assumptions and goals of "smart" technologies are widely accepted in global polity discussions and that they have encouraged the creation of novel infrastructures that organize environmental policy, energy policy, supply chains, the distribution of food and medicine, finance, and security policies. The smartness mandate draws on multiple and intersecting discourses, including ecology, evolutionary biology, computer science, and economics. Binding and bridging these discourses are technologies, instruments, apparatuses, processes, and architectures. These experimental networks of responsive machines, computer mainframes, political bodies, sensing devices, and spatial zones lend durable and material form to smartness, often allowing for its expansion and innovation with relative autonomy from its designers and champions.



Opposite: International Business Machines Corporation (IBM). "Smarter Planet" advertisement campaign, 2008. Detail. From *Wall Street Journal*, 17 November 2008.

Right: Apple Computer. "Think Different" advertisement campaign, 1998. Detail. From *Time*, 2 February 1998.

This essay illuminates some of the key ways in which the history and logic of the smartness mandate are dynamically embedded in the objects and operations of everyday life—particularly the everyday lives of those living in the wealthier Global North, but ideally, for the advocates of smartness, the lives of every inhabitant of the globe. This approach allows us to consider questions such as, What kinds of assumptions link the "predictive" product suggestions made to a global public by retailers such as Amazon or Netflix with the efforts of South Korean urban-planning firms and Indian economic policy makers to monitor and in real time adapt to the activities of their urban citizenry? What kinds of ambitions permit the migration of statistically based modeling techniques from relatively banal consumer applications to regional and transnational strategies of governance? How do smart technologies that enable socially networked applications for smartphones—for example, the Evernote app for distributed multisite and multiuser note taking used by 200 million registered users located primarily in the United States, Europe, Latin America, and Asia—also cultivate new forms of global labor and governmentality, the unity of which resides in coordination via smart platforms rather than, for example, geography or class? Each of these examples relies upon the intermediation of networks and technologies that are designated as "smart," yet the impetus for innovation and the agents of this smartness often remain obscure.

We see the brief history of smartness as a decisive moment in histories of reason and rationality. In their helpful account of "Cold War rationality," Paul Erickson and his colleagues argue that in the years following World War II American science, politics, and industry witnessed "the expansion of the domain of rationality at the expense of . . . reason," as machinic systems and algorithmic procedures displaced judgment and discretion as ideals of governing rationally.[3] Yet at the dawn of the twenty-first century, Cold War rationality has given way to the tyranny of smartness, an eternally emergent program of real-time, short-term calculation that substitutes "demos" (i.e., provisional models) and simulations for those systems of artificial intelligence and professional expertise and calculation imagined by Cold War rationalists. In place of Cold War warring systems based on "rational" processes that could still fall under the control and surveillance of centralized authorities or states, the smartness mandate embraces the ideal of an infinite range of experimental existences, all based on real-time adaptive exchanges among users, environments, and machines. Neither reason nor rationality is understood as a necessary guides for these exchanges, for smartness is presented as a self-regulating process of "optimization" and "resilience" (terms that, as we note below, are themselves moving targets in a recursive system).

Where Cold War rationality was highly suspicious of innovation, the latter is part of the essence of smartness. In place of the self-stabilizing systems and homeostasis that were the orienting ideal of Cold War theorists, smartness assumes perpetual growth and unlimited turmoil. Destruction, crisis, and the absence of architectonic order or rationality are the conditions of possibility for smart growth and optimization. Equally important: whereas Cold War rationality emanated primarily from the conceptual publications of a handful of well-funded think tanks, which tended to understand national populations and everyday culture as masses that need to be guided, smartness pervades cell phones, delivery trucks, and healthcare systems and relies intrinsically on the interactions among, and the individual idiosyncrasies of, millions or even billions of individuals around the planet. Moreover, whereas Cold War rationality was dominated by the thought of the doppelgänger rival (e.g., the United States vs. the Soviet Union; the East vs. the West), smartness is not limited to binaries.[4] Rather, it understands threats as emerging from an environment, which, because it is always more complex than the systems it encompasses, can never be captured in the simple schemas of rivalry or game theory. This, in turn, allows smartness to take on an ecological dimension: the key crisis is no longer simply that emerging from rival political powers or nuclear disaster but is any unforeseeable event that might emerge from an always too-complex environment.

If smartness is what follows Cold War understandings of reason and rationality, the smartness mandate is the political imperative that smartness be extended to all areas of life. In this sense, the smart mandate is what follows "the shock doctrine," powerfully described by Naomi Klein and others.[5] As Klein notes in her book of the same name, the shock doctrine was a set of neoliberal assumptions and techniques that taught policy makers in the 1970s to take advantage of crises to downsize government and deregulate in order to extend the "rationality" of the free market to as many areas of life as possible. The smart mandate, we suggest, is the current instantiation of a new technical logic with equally transformative effects on conceptions and practices of governance, markets, democracy, and even life itself. Yet where the shock doctrine imagined a cadre of experts and advisors deployed to various national polities to liberate markets and free up resources at moments of crisis, the smartness mandate both understands crisis as a normal human condition and extends itself by means of a field of plural agents—environments, machines, populations, data sets—that interact in a complex manner and without recourse to what was earlier understood as reason or intelligence. If the shock doctrine promoted the idea that systems had to be "fixed" so that natural economic relationships could express themselves, the smartness mandate deploys ideas of resilience and practices management without ideals of futurity or clear measures of "success" or "failure." We describe this imperative to develop and instantiate smartness everywhere as a *mandate* in order to capture both its political implications—although smartness is presented by its advocates as politically agnostic, it is more accurately viewed as completely reconfiguring the realm of the political—and the premise that smartness is possible only by drawing upon the "collective intelligence" of large populations.

We seek to sketch the deep logic of smartness and its mandate in four sections, each focused on a different aspect. These sections take up the following questions: (1) Where does smartness happen; that is, what kind of *space* does smartness require? (2) What is the *agent* of smartness; that is, what, precisely, enacts or possesses smartness? (3) What is the key operation of smartness; that is, what does smartness do? (4) What is the purported result of smartness; that is, at what does it aim? Our answers to these four questions are the following:

1. The territory of smartness is *the zone*.
2. The (quasi-)agent of smartness is *populations*.
3. The key operation of smartness is *optimization*.
4. Smartness produces *resilience*.

Focusing on how the logics and practices of zones, populations, optimization, and resilience are coupled enables us to illuminate not just particular instantiations of smartness—for example, smart cities, grids, or phones—but smartness more generally, as well as its mandate ("every process must become smart!").

Our analysis draws inspiration from Michel Foucault's concepts of governmentality and biopolitics, Gilles Deleuze's brief account of "the control society," and critical work on immaterial labor. We describe smartness genealogically; that is, as a concept and set of practices that emerged from the coupling of logics and techniques from multiple fields (ecology, computer science, policy, etc.). We also link smartness to the central object of biopolitics—populations—and see smartness as bound up with the key goal of biopolitics: governmentality. And we emphasize the importance of a mode of control based on what Deleuze describes as open-ended modulation rather than the permanent molding of discipline. We also underscore the centrality of data drawn from the everyday activities of large numbers of people. Yet insofar as smartness positions the global environment as the fundamental orienting point for all governance—that is, as the realm of governance that demands all other problems be seen from the perspective of zones, populations, resilience, and optimization—the tools offered by existing concepts of biopolitics, the control society, and immaterial labor take us only part of the way in our account.[6]

Zones

Smartness has to happen somewhere. However, advocates of smartness generally imply or explicitly note that its space is not that of the national territory. Palmisano's invocation of a smarter planet, for example, emphasizes the extraterritorial space that smartness requires: precisely because smartness aims in part at ecological salvation, its operations cannot be restricted to the limited laws, territory, or populations of a given national polity. So, too, designers of "smart homes" imagine a domestic space freed by intelligent networks from the physical constraints of the home, while the fitness app on a smartphone conditions the training of a single user's body through iterative calculations correlated with thousands or millions of other users spread across multiple continents.[7] These activities all occur in space, but the nation-state is neither their obvious nor necessary container, nor is the human body and its related psychological subject their primary focus, target, or even paradigm (e.g., smartness often employs entities such as "swarms" that are never intended to cohere in the manner of a rational or liberal subject). At the same time, though, smartness also depends on complicated and often delicate infrastructures—fiber-optic cable networks and communications systems capable of accessing satellite data; server farms that must be maintained at precise temperatures; safe shipping routes—that are invariably located at least in part within national territories and are often subsidized by federal governments. Smartness thus also requires the support of legal systems and policing that protect and maintain these infrastructures, and most of the latter are provided by national states (even if only in the form of subcontracted private security services).[8]

This paradoxical relationship of smartness to national territories is best captured as a mutation of the contemporary form of space known as "zones." Related to histories of urban planning and development, where zoning has long been an instrument in organizing space, contemporary zones have new properties married to the financial and logistical practices that underpin their global proliferation. In the past two decades, numerous urban historians and media theorists have redefined the zone in terms of its connection to computation, and described the zone as the dominant territorial configuration of the present. As architectural theorist Keller Easterling notes, the zone should be understood as a method of "extrastatecraft" intended to serve as a platform for the operation of a new "software" for governing human activity. Brett Nielsen and Ned Rossiter invoke the figure of the "logistical city" or zone to make the same point about governmentality and computation.[9]

Zones denote not the demise of the state but the production of new forms of territory, the ideal of which is a space of exception to national and often international law. A key example is the so-called free-trade zone. Free-trade zones are a growing phenomenon, stretching from Pudong District in Shanghai to the Cayman Islands, and even the business districts and port facilities of New York State, and are promoted as conduits for the smooth transfer of capital, labor, and technology globally (with *smooth* defined as a minimum of delay as national borders are crossed). Free-trade zones are in one sense discrete physical spaces, but they also require new networked infrastructures linked through the algorithms that underwrite geographic information systems (GIS) and global positioning systems (GPS) and computerized supply-chain management systems, as well as the standardization of container and shipping architecture and regulatory legal exceptions (to mention just some of the protocols that produce these spaces). Equally important, zones are understood as outside the legal structure of a national territory, even if they technically lie within its space.[10]

In using the term *zone* to describe the space of smartness, our point is not that smartness happens in places such as free-trade zones but that smartness aims to globalize the zonal logic, or mode, of space. This logic of geographic abstraction, detachment, and exemption is exemplified even in a mundane consumer item such as activity monitors—for example, the Fitbit—that link data about the physical activities of a user in one jurisdiction with the data of users in other jurisdictions. This logic of abstraction is more fully exemplified by the emergence of so-called smart cities. An organizing principle of the smart city is that civic governance and public taxation will be driven, and perhaps replaced, by automated and ubiquitous data collection. This ideal of a "sensorial" city that serves as a conduit for data gathering and circulation is a primary fantasy enabling smart cities, grids, and networks. Consider, for example, a prototype "greenfield" (i.e., from scratch) smart-city development, such as Songdo in South Korea. This smart city is designed with a massive sensor infrastructure for collecting traffic, environmental, and closed-circuit television (CCTV) data and includes individual smart homes (apartments) with multiple monitors and touch screens for temperature control, entertainment, lighting, and cooking functions. The city's developers also hope these living spaces will eventually monitor multiple health conditions through home testing. Implementing this business plan, however, will require either significant changes to, or exemptions from, South Korean laws about transferring health information outside of hospitals. Lobbying efforts for this juridical change have been promoted by Cisco Systems (a U.S.-based network infrastructure provider), the Incheon Free Economic Zone (the governing local authority), and Posco (a Korean *chaebol* involved in construction and steel refining), the three most dominant forces behind Songdo.

Songdo, South Korea, 2014. Photo: Orit Halpern.

What makes smart territories unique in a world of zonal territories is the specific mode by which smartness colonizes space through the management of time (and this mode also helps explain why smartness is so successful in promulgating itself globally). As demonstrated by former IBM chair Palmisano's address to the Council on Foreign Relations, smartness is predicated on an imaginary of "crisis" that is to be managed through a massive increase in sensing devices, which in turn purportedly enable self-organization and constant self-modulating and self-updating systems. Smart platforms link zones to crisis via two key operations: (1) a temporal operation, by means of which uncertainty about the future is managed through constant redescription of the present as a “version,” “demo,” or “prototype” of the future; and (2) an operation of self-organization through which earlier discourses about structures and the social are replaced by concerns about infrastructure, a focus on sensor systems, and a fetish for big data and analytics, which purportedly can direct “development” in the absence of clear-cut ends or goals.

In this sense, the development of smart cities such as Songdo follows a logic of software development. Every present state of the smart city is understood as a demo or prototype of a future smart city. Every operation in the smart city is understood in terms of testing and updating. Engineers interviewed at the site openly spoke of it as an “experiment” and “test,” admitting that the system did not work but stressing that problems could be fixed in the next instantiation elsewhere in the world.[11] As a consequence, there is never a finished product but rather infinitely replicable yet always preliminary, never-to-be-completed versions of these cities around the globe.

This temporal operation is then linked to an ideal of self-organization. Smartness largely refers to computationally and digitally managed systems, from electrical grids to building management systems, that can learn and, in theory, adapt by analyzing data about themselves. Self-organization is thus linked to the operation of optimization. Systems correct themselves automatically by adjusting their own operations. This organization is imagined as being immanent to the physical and informational system at hand—that is, as optimized by computationally collected data rather than by “external” political or social actors. At the heart of the smartness mandate is thus a logic of immanence, by means of which sensor instrumentation adjoined to emerging and often automated methods for the analysis of large data sets allow a dynamic system to detect and direct its continued growth.[12]

One of the key, troubling consequences of demoing and self-organization as the two zonal operations of smartness is that the overarching concept of “crisis” begins to obscure differences among kinds of catastrophes. While every crisis event—for example, the 2008 subprime mortgage collapse or the Tohoku earthquake of 2011—is different, within the demo-logic that underwrites the production of smart and resilient cities these differences can be subsumed under the general concept of crisis and addressed through the same methods (the implications of which must never be fully engaged because we are always “demoing” or “testing” solutions, never actually solving the problem). Whether threatened by terrorism, subprime mortgages, energy shortages, or hurricanes, smartness always responds in essentially the same way. The demo is a form of temporal management that through its practices and discourses evacuates the historical and contextual specificity of individual catastrophes and evades ever having to assess or represent the impact of these infrastructures, because no project is ever "finished." This evacuation of differences, temporalities, and societal structures is what most concerns us in confronting the extraordinary rise of ubiquitous computing and high-tech infrastructures as solutions to political, social, environmental, and historical problems confronting urban design and planning, and as engines for producing new forms of territory and governance.

Populations

If zones are the places in which smartness takes place, *populations* are the agents—or, more accurately, the enabling medium—of smartness. Smartness is located neither in the source (producer) nor the destination (consumer) of a good such as a smartphone but is the outcome of the algorithmic manipulation of billions of traces left by thousands, millions, or even billions of individual users. Smartness requires these large populations, for they are the medium of the "partial perceptions" within which smartness emerges. Though these populations should be understood as fundamentally biopolitical in nature, it is more helpful first to recognize the extent to which smartness relies on an understanding of population drawn from twentieth-century biological sciences such as evolutionary biology and ecology.

Biologists and ecologists often use the term *population* to describe large collections of individuals with the following characteristics: (1) the individuals differ at least slightly from one another; (2) these differences allow some individuals to be more "successful" vis-à-vis their environment than other individuals; (3) a form of memory enables differences that are successful to appear again in subsequent generations; and, as a consequence, (4) the distribution of differences across the population tends to change over time.[13] This emphasis on the importance of individual difference for long-term fitness thus distinguishes this use of the term *population* from more common political uses of the term to describe the individuals who live within a political territory.[14]

Smartness takes up a biologically oriented concept of population but repurposes it for nonbiological contexts. Smartness presumes that each individual is distinct not only biologically but in terms of, for example, habits, knowledge, consumer preferences, and that information about these individual differences can usefully be grouped together so that algorithms can locate subgroupings of this data that thrive or falter in the face of specific changes. Though the populations of data drawn from individuals may map onto traditional biological or political divisions, groupings and subgroupings more generally revolve around consumer preferences and are drawn from individuals in widely separated geographical regions and polities. (For example, Netflix's populations of movie preferences are currently created from users distributed throughout 190 countries.)[15] Moreover, though these data populations are (generally) drawn from human beings, they are best understood as distinct from the human populations from which they emerge: these are simply data populations of, for example, preferences, reactions, or abilities. This is true even in the case of information drawn from human bodies located in the same physical space. In the case of the smart city, the information streaming from fitness trackers, smartphones, credit cards, and transport cards is generated by human bodies in close physical proximity to one another, but individual data populations are then agglomerated at different temporalities and scales, depending on the problem being considered (transportation routing, energy use, consumer preferences, etc.). These discrete data populations enable processes to be optimized (i.e., enable "fitness" to be determined), which in turn produces new populations of data and hence a new series of potentialities for what a population is and what potentials these populations can generate.

A key premise of smartness is that while each member of a population is unique, it is also "dumb"—that is, limited in its "perception"—and that smartness emerges as a property of the population only when these limited perspectives are linked via environment-like infrastructures. Returning to the example of the smartphone operating in a smart city, the phone becomes a mechanism for creating data populations that operate without the cognition or even direct command of the subject. (The smartphone, for example, automatically transmits its location and can also transmit other data about how it has been used.) If, in the biological domain, populations enable long-term species survival, then in the cultural domain populations enable smartness, provided the populations are networked together with smart infrastructures. Populations are the perceptual substrate that enables modulating interactions among agents within a system that sustains particular activities. The infrastructures ensure, for example, that "given enough eyeballs, all bugs are shallow" (Linus's Law); that problems can be "crowdsourced"; and that such a thing as "collective intelligence" exists.[16] The concept of population also allows us to understand better why the zone is the necessary kind of space enabling smartness, for there is often no reason that national borders would parse population differences (in abilities, interests, preferences, or biology) in any way that is meaningful for smartness.

This creation and analysis of data populations is biopolitical in the sense initially outlined by Foucault, but smartness is also a significant mutation in the operation of biopolitics. Foucault stresses that the concept of population was central to the emergence of biopolitics in the late eighteenth century, for it denoted a "collective body" that had its own internal dynamics (of births, deaths, illness, etc.) that were quasi-autonomous in the sense that they could not be commanded or completely prevented by legal structures but could nevertheless be subtly altered through biopolitical regulatory techniques and technologies (e.g., required inoculations; free-market mechanisms).[17] On the one hand, smartness is biopolitical in this same sense, for the members of its populations—populations of movie watchers, cell phone users, healthcare purchasers and users, and so on—are assumed to have their own internal dynamics and regularities, and the goal of gathering information about these dynamics is not to discipline individuals into specific behaviors but to find points of leverage within these regularities that can produce more subtle and widespread changes.

On the other hand, the biopolitical dimension of smartness cannot be understood as simply "more of the same," for four reasons. First, and in keeping with Deleuze's reflections on the control society, the institutions that gather data about populations are now more likely to be corporations than states.[18] Second (and as a consequence of the first point), smartness's data populations often concern not those clearly biological events on which Foucault focused, but variables such as attention, consumer choices, and transportation preferences. Third, though the data populations that are the medium of smartness are often drawn from populations of human beings, these data relate differently to individuals than in the case of Foucault's more health-oriented examples. Data populations themselves often do not need to be (and cannot be) mapped directly back onto discrete human populations: one is often less interested in discrete events that happen infrequently along the individual biographies of a polity (e.g., smallpox infections) than in frequent events that may happen across widely dispersed groups of people (e.g., movie preferences). The analysis of these data populations is then used to create, via smart technologies, an individual and customized "information-environment" around each individual. The aim is not to discipline individuals, in Foucault's sense, but to extend ever deeper and further the quasi-autonomous dynamics of populations. Fourth, in the case of systems such as high-speed financial trading and derivatives, as well as in the logistical management of automated supply chains, entire data populations are produced and acted on directly through entirely machine-to-machine data gathering, communication, analytics, and action.[19] These new forms of automation and of producing populations mark transformations in both the scale and intensity of the interweaving of algorithmic calculation and life.

Optimization

Smartness emerges when zones link the increasingly fine-grained, quasi-autonomous dynamics of populations for the sake of *optimization*. This pursuit of "the best"—the fastest route between two points, the most reliable prediction of a product a consumer will like, the least expenditure of energy in a home, the lowest risk and highest return in a financial portfolio—is what justifies the term *smartness*. Contemporary optimization is a fundamentally quantitative but calculation-intensive operation; it is a matter of finding, given specified constraints, maxima or minima. Locating these limits in population data often requires millions or billions of algorithmic mathematical calculations—hence the role of computers (which run complex algorithms at speeds that are effectively "real-time" for human beings), globally distributed sensors (which enable constant global updating of distributed information), and global communications networks (which connect those sensors with that computing power).

Though optimization has a history, including techniques of industrial production and sciences of efficiency and fatigue pioneered in the late-nineteenth and early twentieth centuries by Fredrick Winslow Taylor and Frank Gilbreth, its current instantiations radically differ from earlier ones.[20] The term *optimization* appears to have entered common usage in English only following World War II.[21] Related to emerging techniques such as game-theoretical tools and computers, optimization is a particular form of efficiency measure. To optimize is to find the best relationship between minima and maxima performances of a system. Optimization is not a normative or absolute measure of performance but an internally referential and relative one; it thus mirrors the temporality of the test bed, the version, and the prototype endemic to "smart" cities and zones.

Optimization is the technique by which smartness promulgates the belief that everything—every kind of relationship among human beings, their technologies, and the environments in which they live—can and should be algorithmically managed. Shopping, dating, exercising, the practice of science, the distribution of resources for public schools, the fight against terrorism, the calculation of carbon offsets and credits: these processes can—and must!—be optimized. Optimization fever propels the demand for ever-more sensors—more sites of data collection, whether via mobile device apps, hospital clinic databases, or tracking of website clicks—so that optimization's realm can perpetually be expanded and optimization itself further optimized. Smart optimization also demands the ever-increasing evacuation of private interiority on the part of individuals, for such privacy is now often implicitly understood as an indefensible withholding of information that could be used for optimizing human relations.[22]

Smart optimization also presumes a new, fundamentally practical epistemology, for smartness is not focused on determining absolutely correct (i.e., "true") solutions to optimization problems. The development of calculus in the seventeenth century encouraged the hope that, if one could simply find an equation for a curve that described a system, it would then always be possible in principle to locate absolute, rather than simply local, maxima and minima for that system. However, the problems engaged by smartness—for example, travel mapping, healthcare outcomes, risk portfolios—often have so many variables and dimensions that completely solving them, even in principle, is impossible. As Dan Simon notes, even a problem as apparently simple as determining the most optimal route for a salesperson who needs to visit fifty cities would be impossible if one were to try to calculate all possible solutions. There are 49! (= 6.1 x 10^{62}) possible solutions to this problem, which is

> beyond the capability of contemporary computing: even if one had a trillion computers, each capable of calculating a trillion solutions per second, and these computers had been calculating since the universe began—a total computation time of 15 billion years—they would not yet have come close to calculating all possible routes.[23]

In the face of the impossibility of determining absolute maxima or minima for these systems by so-called brute force (i.e., calculating and comparing all possible solutions), contemporary optimization instead involves finding good-enough solutions: maxima and minima that may or may not be absolute but are more likely than other solutions to be close to absolute maxima or minima. The optimizing engineer selects among different algorithmic methods that each produce, in different ways and with different results, good-enough solutions.

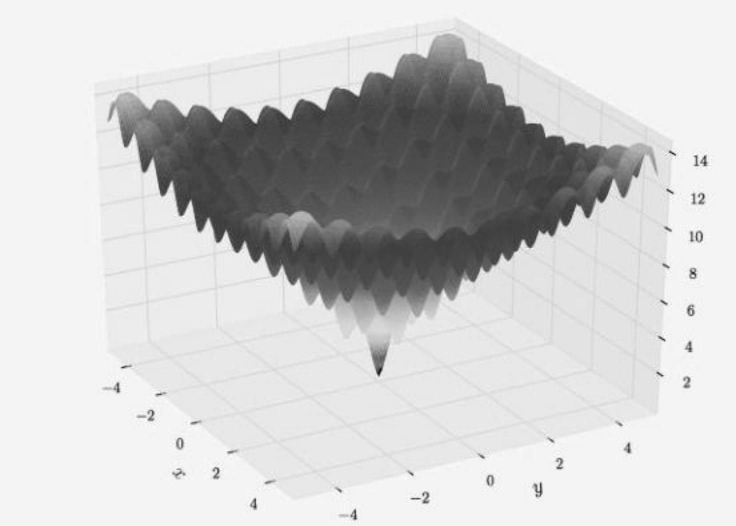
In the absence of any way to calculate absolute maxima and minima, the belief that smartness nevertheless locates "best" solutions is supported both technically and analogically. This belief is supported technically in that different optimization algorithms are run on "benchmark" problems—that is, problems that contain numerous local maxima and minima but for which the absolute maximum or minimum *is* known—to determine how well the algorithms perform on those types of problems.[24] If an algorithm runs well on a benchmark problem, then it is presumed to be more likely to run well on similar real-world problems.

The belief that smartness finds the best solutions is also often supported by the claim that many contemporary optimization algorithms mimic natural processes, especially computational ideals of biological evolution.[25] The algorithm begins with the premise that natural biological evolution automatically solves optimization problems by means of natural populations. The algorithm then seeks to simulate that process by creating populations of candidate solutions, which are mixed with one another (elements of one candidate solution are combined with elements of other candidate solutions) and culled through successive generations to produce increasingly good solutions. David B. Fogel, a consultant for the informatics firm Natural Selection, Inc., which applies computational models to the streamlining of commercial activities, captures this sense of optimization as simply a continuation of nature's work: "Natural evolution is a population-based optimization process. Simulating this process on a computer results in stochastic optimization techniques that can often outperform classical methods of optimization when applied to difficult real-world problems."[26] Optimization research implements these features (reproduction, mutation, competition, and selection) in computers to find "natural" laws that can govern the organization of industrial or other processes that, when implemented on a broad scale, become the conditions of life itself.

This vision of optimization then justifies the extension and intensification of the zonal logic of smartness. To optimize all aspects of existence, smartness must be able to locate its relevant populations (of preferences, events, etc.) wherever they occur. However, this is possible only when every potential data point (i.e., partial perception) on the globe can be directly linked to every other potential data point without interference from specific geographic jurisdictional regimes. This does not mean the withering of geographically based security apparatuses; on the contrary, optimization often requires strengthening these to protect the concrete infrastructures that enable smart networks and to implement optimization protocols. Yet, like the weather or global warming, optimization is not to be restricted to, or fundamentally parsed by, the territories that fund and provide these security apparatuses but must be allowed to operate as a sort of external environmental force.

Resilience

If smartness happens through zones, if its operations require populations, and if it aims most fundamentally at optimization, what is the telos of smartness itself? That is, at what does smartness aim, and why is smartness understood as a virtue? The answer is that smartness enables *resilience*. This is its goal and raison d'être. The logic of resilience is peculiar in that it aims not precisely at a future that is "better" in any absolute sense but at a smart infrastructure that can absorb constant shocks while maintaining functionality and organization. Following the work of Bruce Braun and Stephanie Wakefield, we describe resilience as a state of permanent management that does without guiding ideals of progress, change, or improvement.[27]



Ackley benchmark function. Image: Gaortizg, Wikimedia Commons.

The term *resilience* plays important, though differing, roles in multiple fields. These include engineering and material sciences—since the nineteenth century, the "modulus of resilience" has measured the capacity of materials such as woods and metals to return to their original shape after impact—as well as ecology, psychology, sociology, geography, business, and public policy, in which *resilience* names ways in which ecosystems, individuals, communities, corporations, and states respond to stress, adversity, and rapid change.[28] However, the understanding of resilience most crucial to smartness and the smartness doctrine was first forged in ecology in the 1970s, especially in the work of C.S. Holling, who established a key distinction between "stability" and "resilience." Working from a systems perspective and interested in the question of how human beings could best manage elements of ecosystems of commercial interest (e.g., salmon, wood), Holling developed the concept of resilience to contest the premise that ecosystems were most healthy when they returned quickly to an equilibrium state after being disturbed (and in this sense his paper critiqued then current industry practices).

Holling defines *stability* as the ability of a system that had been perturbed to return to a state of equilibrium, but he argued that stable systems were often unable to compensate for significant, swift environmental changes. As Holling writes, the "stability view [of ecosystem management] emphasizes the equilibrium, the maintenance of a predictable world, and the harvesting of nature's excess production with as little fluctuation as possible," yet this approach that "assures a stable maximum sustained yield of a renewable resource might so change [the conditions of that system] . . . that a chance and rare event that previously could be absorbed can trigger a sudden dramatic change and loss of structural integrity of the system."[29] Resilience, by contrast, denotes for Holling the capacity of a system *to change* in periods of intense external perturbation and thus to persist over longer time periods. The concept of resilience encourages a management approach to ecosystems that "would emphasize the need to keep options open, the need to view events in a regional rather than a local context, and the need to emphasize heterogeneity." Resilience is thus linked to concepts of crisis and states of exception; that is, it is a virtue when crises and states of exception are assumed to be either quasi-constant or the most relevant states. Holling also underscores that the movement from stability to resilience depends upon an epistemological shift: "Flowing from this would be not the presumption of sufficient knowledge, but the recognition of our ignorance: not the assumption that future events are expected, but that they will be unexpected."[30]

Smartness abstracts the concept of resilience from a systems approach to ecology and turns it into an all-purpose epistemology and value, positing resilience as a more general strategy for managing perpetual uncertainty and encouraging the premise that the world is indeed so complex that unexpected events are the norm. Smartness enables this generalization of resilience in part because it abstracts the concept of populations from the specifically biological sense employed by Holling. Smartness sees populations of preferences, traits, and algorithmic solutions, as well as populations of individual organisms. Resilience also functions in the discourse of smartness to collapse the distinction between *emergence* (something new) and *emergency* (something new that threatens), and does so to produce a world where any change can be technically managed and assimilated while maintaining the ongoing survival of the system rather than of individuals or even particular groups. The focus of smartness is thus the management of the *relationships between* different populations of data, some of which can be culled and sacrificed for systemic maintenance.[31] Planned obsolescence and preemptive destruction combine here to encourage the introduction of more computation into the environment—and emphasize as well that resilience of the human species may necessitate the sacrifice of "suboptimal" populations.

The discourse of resilience effectively erases the differences among past, present, and future. Time is not understood through an historical or progressive schema but through the schemas of repetition and recursion (the same shocks and the same methods are repeated again and again), even as these repetitions and recursions produce constantly differing territories. This is a self-referential difference, measured or understood only in relation to the many other versions of smartness (e.g., earlier smart cities), which all tend to be built by the same corporate and national assemblages.

The collapse of emergence into emergency also links resilience to financialization through derivation, as the highly leveraged complex of Songdo demonstrates.[32] The links that resilience establishes among emergency, financialization, and derivatives are also exemplified by New York City, which, after the devastation of Hurricane Sandy in 2012, adopted the slogan "Fix and Fortify." This slogan underscores an acceptance of future shock as a necessary reality of urban existence, while at the same time leaving the precise nature of these shocks unspecified (though they are often implied to include terrorism as well as environmental devastation). The naturalization of this state is vividly demonstrated by the irony that the real destruction of New York had earlier been imagined as an opportunity for innovation, design thinking, and real estate speculation. In 2010, shortly before a real hurricane hit New York, the Museum of Modern Art (MoMA) and P.S.1 Contemporary Art Center ran a design competition and exhibition titled *Rising Currents*, which challenged the city's premier architecture and urban design firms to design for a city ravaged by the elevated sea levels produced by global warming:

> MoMA and P.S.1 Contemporary Art Center joined forces to address one of the most urgent challenges facing the nation's largest city: sea-level rise resulting from global climate change. Though the national debate on infrastructure is currently focused on "shovel-ready" projects that will stimulate the economy, we now have an important opportunity to foster new research and fresh thinking about the use of New York City's harbor and coastline. As in past economic recessions, construction has slowed

> dramatically in New York, and much of the city's remarkable pool of architectural talent is available to focus on innovation.[33]

A clearer statement of the relationship of urban planners to crisis is difficult to imagine: Planning must simply assume and assimilate future, unknowable shocks, and these shocks may come in any form. This stunning statement turns economic tragedy, the unemployment of most architects, and the imagined coming environmental apocalypse into an opportunity for speculation—technically, aesthetically, and economically. This is a quite literal transformation of emergency into emergence that creates a model for managing perceived and real risks to the population and infrastructure of the territory not by "solving" the problem but by absorbing shocks and modulating the way environment is managed. New York in the present becomes a mere demo for the postcatastrophe New York, and the differential between these two New Yorks is the site of financial, engineering, and architectural interest and speculation.

This relationship of resilience to the logic of demos and derivatives is illuminated by the distinction between risk and uncertainty first laid out in the 1920s by the economist Frank Knight. According to Knight, uncertainty, unlike risk, has no clearly defined endpoints or values.[34] Uncertainty offers no clear-cut terminal events. If the Cold War was about nuclear testing and simulation as a way to avoid an unthinkable but nonetheless predictable event—nuclear war—the formula has now been changed. We live in a world of fundamental uncertainty, which can only ever be partially and provisionally captured through discrete risks. When uncertainty, rather than risk, is understood as the fundamental context, "tests" can no longer be understood primarily as a simulation of life; rather, the test bed makes human life itself an experiment for technological futures. Uncertainty thus embeds itself in our technologies, both those of architecture and of finance. In financial markets, for example, risks that are never fully accounted for are continually "swapped," "derived," and "leveraged" in the hope that circulation will defer any need to represent risk, and in infrastructure, engineering, and computing we do the same.[35]

As future risk is transformed into uncertainty, smart and ubiquitous computing infrastructures become the language *and* practice by which to imagine and create our future. Instead of looking for utopian answers to our questions regarding the future, we focus on quantitative and algorithmic methods and on logistics—on *how* to move things rather than on questions of where they *should* end up. Resilience as the goal of smart infrastructures of ubiquitous computing and logistics becomes the dominant method for engaging with possible urban collapse and crisis (as well as the collapse of other kinds of infrastructure, such as those of transport, energy, and finance). Smartness thus becomes the organizing concept for an emerging form of technical rationality whose major goal is management of an uncertain future through a constant deferral of future results; for perpetual and unending evaluation through a continuous mode of self-referential data collection; and for the construction of forms of financial instrumentation and accounting that no longer engage (or even need to engage with), alienate, or translate what capital extracts from history, geology, or life.

Smartness and Critique

Smartness is both a reality and an imaginary, and this comingling underwrites both its logic and the magic of its popularity. As a consequence, though, the critique of smartness cannot simply be a matter of revealing the inequities produced by its current instantiations. Critique is itself already central to smartness, in the sense that perpetual optimization requires perpetual dissatisfaction with the present and the premise that things can always be better. The advocates of smartness can always plausibly claim (and likely also believe) that the *next* demo will be more inclusive, equitable, and just. The critique of smartness thus needs to confront directly the terrible but necessary complexity of thinking and acting within earthly scale—and even extraplanetary scale—technical systems.

This means in part stressing, as we have done here, that the smartness mandate transforms conditions of environmental degradation, inequality and injustice, mass extinctions, wars, and other forms of violence by means of the demand to understand the present as a demo oriented toward the future, and by necessitating a single form of response—increased penetration of computation into the environment—for all crises. On the other hand, not only the agency and transformative capacities of the smart technical systems but the deep appeal of this approach to managing an extraordinarily complex and ecologically fragile world are impossible to deny. None of us are eager to abandon our cell phones or computers. Moreover, the epistemology of partial truths, incomplete perspectives, and uncertainty with which Holling sought to critique capitalist understandings of environments and ecologies still holds a weak messianic potential for revising older modern forms of knowledge and for building new forms of affiliation, agency, and politics grounded in uncertainty, rather than objectivity and surety, and in this way keeping us open to plural forms of life and thought. However, insofar as smartness separates critique from conscious, collective, human reflection—that is, insofar as smartness seeks to steer communities algorithmically, in registers operating below consciousness and human discourse—critiquing smartness will in part be a matter of excavating and rethinking each of its central concepts and practices (zones, populations, optimization, and resilience), as well as the temporal logic that emerges from the particular way in which smartness combines these concepts and practices.

Notes

[1] Samuel J. Palmisano, “A Smarter Planet: The Next Leadership Agenda,” Council on Foreign Relations, 6 November 2008, http://www.cfr.org/event/smarter-planet-next-leadership-agenda/.

[2] Palmisano, “A Smarter Planet.”

[3] Paul Erickson et al., *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality* (Chicago: University of Chicago Press, 2015), 2. Erickson and his coauthors stress that for Cold War authors and policy makers, the possibility of nuclear war made it imperative that people—or at least military commanders and policy makers—act “rationally,” in the sense that tendencies to innovate or depart from programmable rules be prevented. The consequence was that “mechanical rule following . . . [became] the core of rationality” (31).

[4] Though the image of Cold War rationality developed by Erikson et al. is especially useful for our purposes here, we also want to acknowledge alternative histories of temporality and control, many emerging from cybernetics, within the history of Cold War computing. See, for example, Orit Halpern, “Cybernetic Rationality,” *Distinktion: Scandinavian Journal of Social Theory* 15, no. 2 (2014): 1–16.

[5] Naomi Klein, *The Shock Doctrine: The Rise of Disaster Capitalism* (New York: Metropolitan Books/Henry Holt, 2007). Klein’s book is part of an extensive bibliography of recent critical work on neoliberalism that also includes David Harvey, *A Brief History of Neoliberalism* (New York: Oxford University Press, 2005); Philip Mirowski and Dieter Plehwe, eds., *The Road from Mont Pèlerin: The Making of the Neoliberal Thought Collective* (Cambridge, MA: Harvard University Press, 2009); Jamie Peck, *Constructions of Neoliberal Reason* (New York: Oxford University Press, 2010); and Philip Mirowski, *Never Let a Serious Crisis Go to Waste: How Neoliberalism Survived the Financial Meltdown* (New York: Verso, 2014).

[6] See, especially, Michel Foucault, *The History of Sexuality*, vol. 1, *An Introduction* (New York: Pantheon Books, 1978); Michel Foucault, *Society Must Be Defended: Lectures at the Collège de France, 1975–76*, trans. David Macey (New York: Picador, 2003); Michel Foucault, *Security, Territory, Population: Lectures at the College de France, 1977–78*, trans. Graham Burchell, ed. Michel Senellart (New York: Palgrave Macmillan, 2007); and Michel Foucault, *The Birth of Biopolitics: Lectures at the College de France, 1978–79*, trans. Graham Burchell, ed. Michel Senellart (New York: Palgrave Macmillan, 2008); Gilles Deleuze, “Postscript on the Societies of Control,” *October* 59 (1992): 3–7; Maurizio Lazzarato, “Immaterial Labour,” trans. Paul Colilli and Ed Emory, in *Radical Thought in Italy*, ed. Paolo Virno and Michael Hardt (Minneapolis: University of Minnesota Press, 1996): 132–46; and Michael Hardt and Antonio Negri, *Empire* (Cambridge, MA: Harvard University Press, 2000), 290–94.

[7] On the smart home, see Lynn Spigel, “Designing the Smart House: Posthuman Domesticity and Conspicuous Production,” in *Public Worlds: Electronic Elsewheres: Media, Technology, and the Experience of Social Space*, ed. Chris Berry, Soyoung Kim, and Lynn Spigel (Minneapolis: University of Minnesota Press, 2009), 55–92.

[8] Considerable work—some very critical and some very utopian—has been done on the “smart” city, smart city projects, and “smart” or big data infrastructures. For a sampling, see Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (London: Sage Publications, 2014); Anthony M. Townsend, *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia* (New York: W.W. Norton, 2014); Carlo Ratti and Matthew Claudel, *The City of Tomorrow: Sensors, Networks, Hackers, and the Future of Urban Life* (New Haven: Yale University Press, 2016); Adam Greenfield, *Against the Smart City (the City Is Here for You to Use)* (New York: Do Projects, 2013); Shannon Mattern, *Deep Mapping*

the Media City (Minneapolis: University of Minnesota, 2015); and Richard Sennett, "The Stupefying Smart City," paper presented at the Urban Age Electric City Conference, London, 2012.

9. See Keller Easterling, *Extrastatecraft: The Power of Infrastructure Space* (New York: Verso, 2014); and Ned Rossiter, *Software, Infrastructure, Labor: A Media Theory of Logistical Nightmares* (New York: Routledge, 2016.

10. For more on the "logistical" city and free-trade zones, see Brett Nielsen and Ned Rossiter, "The Logistical City," in *Transit Labour: Circuits, Regions, Borders* (Sydney: University of Western Sydney, 2011): 2–5; Aiwha Ong and Ananya Roy, eds. *Worlding Cities, or the Art of Being Global* (London: Routledge, 2011); Saskia Sassen, *Expulsions: Brutality and Complexity in the Global Economy* (Cambridge, MA: Harvard University Press, 2014); Manuel Castells, *The Rise of the Network Society* (New York: Wiley-Blackwell, 2000); Deborah Cowen, *The Deadly Life of Logistics: Mapping Violence in Global Trade* (Minneapolis: University of Minnesota, 2014); and David Harvey, *Spaces of Capital* (London: Routledge, 2012).

11. Orit Halpern, Jesse LeCavalier, and Nerea Calvillo, "Test-Bed Urbanism," *Public Culture*, no. 26 (March 2013): 274.

12. Smartness thus partakes in what Shannon Mattern calls "methodolatry," a constant obsession with methods and measurement to assess prototypes that are never completed, and hence, assessment of results without any clear final metric or endpoint. See Shannon Mattern, "Methodolatry and the Art of Measure," *Places*, November 2013, https://placesjournal.org/article/methodolatry-and-the-art-of-measure/.

13. For a key early reflection on biological "population thinking," see Ernst Mayr, "Darwin and the Evolutionary Theory in Biology," in *Evolution and Anthropology: A Centennial Appraisal*, ed. B.J. Meggers (Washington, DC: Anthropological Society of Washington, 1959), 1–10. For a helpful reflection on key aspects of biological concepts of population, see Peter Godfrey-Smith, *Darwinian Populations and Natural Selection* (Oxford, UK: Oxford University Press, 2009).

14. On the complicated and shifting relationships in the twentieth century between natural and social scientific approaches to population, see Edmund Ramsden, "Eugenics from the New Deal to the Great Society: Genetics, Demography and Population Quality," *Studies in History and Philosophy of Biological and Biomedical Sciences* 9 (2008): 391–406.

15. See Yves Raimond and Justin Basilico, "Recommending for the World," *Netflix Technology Blog*, February 17, 2016, http://techblog.netflix.com/2016/02/recommending-for-world.html.

16. As these examples suggest, we see the concept of population as more useful for an analysis and critique of smartness than contemporary alternative terms such as *crowds*, *swarms*, and *collectives*. While each of these terms admittedly stresses different aspects—*population* emphasizes long-term biological adaptability and persistence; *crowds* and *swarms* emphasize speed of change and decentralized control; and *collective* is a more clearly political term—the concept of population underscores the evolutionary logic of smartness, as well as the underlying meanings of optimization and resilience central to its operation. The concept of "multitude" employed (in different ways) by Paolo Virno and by Hardt and Negri is more helpful in drawing off aspects of smartness from their embeddedness within naturalistic and neoliberal assumptions. Whether these authors successfully engage the ecological dimension of smartness, which is essential to its current appeal, is not clear to us, however. See Paolo Virno, *A Grammar of the Multitude: For an Analysis of Contemporary Forms of Life* (New York: Semiotext(e), 2004); Hardt and Negri, *Empire*; Michael Hardt and Antonio Negri, *Multitude: War and Democracy in the Age of Empire* (New York: Penguin Press, 2004); and Michael Hardt and Antonio Negri, *Commonwealth* (Cambridge, MA: Belknap Press of Harvard University Press, 2009).

[17] See, especially, Foucault, *The History of Sexuality*, vol. 1, 25–26, 139–47; Foucault, *Society Must Be Defended*, 239–64; and Foucault, *Security, Territory, Population,* 38–44, 62–79.

[18] Deleuze, "Postscript on the Societies of Control."

[19] On financialization and computation, see Michael Lewis, *Flash Boys: A Wall Street Revolt* (New York: W.W. Norton, 2014); and Donald A. MacKenzie, *An Engine, Not a Camera: How Financial Models Shape Markets* (Cambridge, MA: MIT Press, 2006).

[20] Anson Rabinbach, *The Human Motor: Energy, Fatigue, and the Origins of Modernit*y (Berkeley and Los Angeles, CA: University of California Press, 1992).

[21] The term was used in the mid-nineteenth century, but, according to Google Ngram, did not enter common parlance until the 1950s. Google, Ngram Viewer, "optimization," https://books.google.com/ngrams/graph?content=optimization&year_start=1800&year_end=2000&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Coptimization%3B%2Cc0. To our knowledge, no critical history has been written about optimization, and existing historical sketches written by mathematicians and economists tend to position optimization as a biological drive or natural force that received proper mathematical formulation in the eighteenth century and was more fully developed in the post-WWII period. See, for example, the entry on "History of Optimization," in *Encyclopedia of Optimization*, ed. Christodoulos A. Floudas (New York: Springer, 2008). For a useful account of optimization theory in economics, see Philip Mirowski, *More Heat than Light: Economics as Social Physics; Physics as Nature's Economics* (Cambridge, UK: Cambridge University Press, 1989); and Philip Mirowski, *Machine Dreams: Economics Becomes a Cyborg Science* (Cambridge, UK: Cambridge University Press, 2002). For optimization in logistics, see Jesse LeCavallier, *The Rule of Logistics: Walmart and the Architecture of Fulfillment* (Minneapolis: University of Minnesota, 2016).

[22] This evacuation of interiority and exteriority is arguably a key reason for the recent turn to "anonymity" as a form of political and technical action and for the rise of "dark" pools and other "dark" infrastructures to facilitate ongoing privatization and wealth accumulation by the select few. On anonymity, see Gabriella Coleman, *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous* (New York: Verso, 2015). On dark pools, see Scott Patterson, *Dark Pools: High-Speed Traders, AI Bandits, and the Threat to the Global Financial System* (New York: Crown Business, 2012).

[23] Dan Simon, *Evolutionary Optimization Algorithms* (Somerset, NJ: John Wiley and Sons, 2013), 20–21.

[24] An example is the Ackley benchmark function:
$f(x,y)=-20\exp[-0.2\sqrt{0.5(x^2+y^2)}]-\exp[0.5(\cos 2\pi x+\cos 2\pi y)]+e+20$.
The absolute minimum of this function is zero, but since it contains many closely clustered local minima, evolutionary optimization algorithms find it difficult to locate the absolute minimum. Different evolutionary optimization algorithms can be tested on this function to determine how close each can come to the absolute minimum. Simon, 643–44.

[25] Our phrase "computational ideals of biological evolution" is intended to underscore that what is coded as "genetics" and "evolutionary accounts" was itself often originally predicated on assumptions emerging from fields such as economics, game theory, and computer science. On the impact of computation on ethology, ecology, environmentalism, and the life sciences, particularly in respect to resilience and optimization, see Adam Curtis, *All Watched Over by Machines of Loving Grace*, episode 2 ("The Use and Abuse of Vegetational Concepts"), BBC2, 30 May 2011; and Jennifer Gabrys, *Program Earth: Environmental Sensing Technology and the Making of a Computational Planet* (Minneapolis: University of Minnesota Press, 2016).

[26] David B. Fogel, "An Introduction to Simulated Evolutionary

Optimization," *IEEE Transactions of Neural Networks* 5, no.1 (1994): 3. The volume of *IEEE Transactions of Neural Networks* in which this essay appears, titled "Evolutionary Computing: The Fossil Record," establishes the importance of Mayr's evolutionary population thinking for this approach to computing (see, e.g., xi, 1, 11).

27. Stephanie Wakefield and Bruce Braun, "Living Infrastructure, Government, and Destituent Power," unpublished presentation given at the symposium Anthropology of the Anthropocene, Concordia University, 23 October 2015, 7.

28. D.E. Alexander, "Resilience and Disaster Risk Reduction: An Etymological Journey," *Natural Hazards and Earth System Sciences* 13 (2013): 2707–16. See also Jeremy Walker and Melinda Cooper, "Genealogies of Resilience: From Systems Ecology to the Political Economy of Crisis Adaptation," *Security Dialogue* 2 (2001): 143–60.

29. C.S. Holling, "Resilience and Stability of Ecological Systems," *Annual Review of Ecological Systems* 4 (1973): 21.

30. Holling, 21.

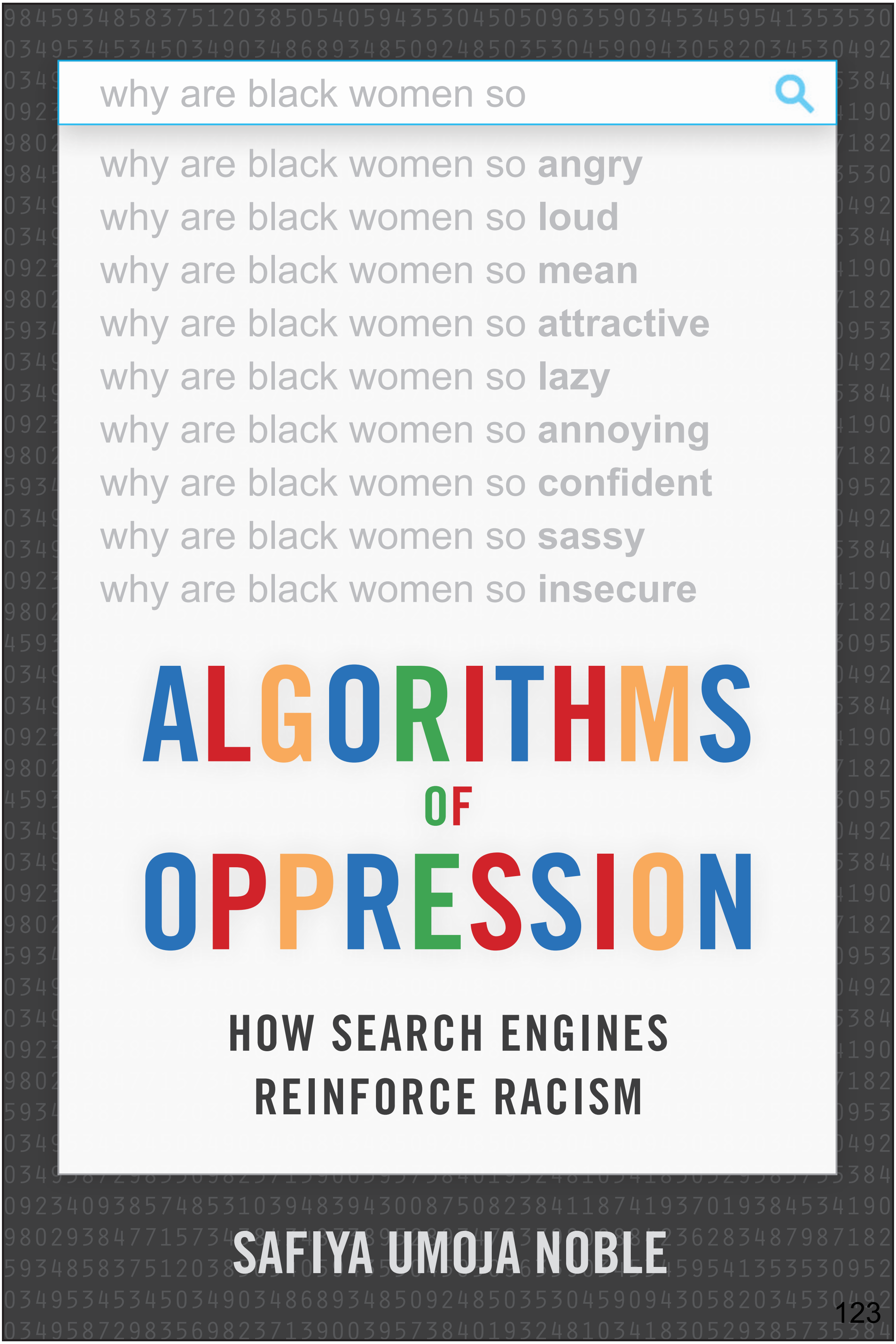
31. Resilience is not equivalent to robustness. As Alexander R. Galloway notes in *Protocol: How Control Exists after Decentralization* (Cambridge, MA: MIT Press, 2004), 43–46, "robustness" is a defining feature of the technical concept of protocol, which is central to the computational dimension of smart infrastructures. However, insofar as robustness refers to the ability of a system to retain its original configuration despite confusing input, it is analogous to what Holling calls "stability," rather than to resilience. Robustness is thus just one of many technical means for enabling resilient systems.

32. On resilience in finance and on economic and development policies, see Melinda Cooper, "Turbulent Worlds: Financial Markets and Environmental Crisis," *Theory, Culture and Society* 27, no. 2–3 (2010): 167–90; and Stephanie Wakefield and Bruce Braun, "Governing the Resilient City," *Environment and Planning D: Society and Space*, no. 32 (2014): 4–11.

33. Introductory statement, *Rising Currents*, Museum of Modern Art, originally published on 8 October 2010 on the MoMA website (https://www.moma.org/explore/inside_out/rising-currents?x-iframe=true#description), now available at One Office Architects (OOAA), "Projects for New York's Waterfront," http://www.oneofficearchitects.com/215/.

34. Frank Knight, *Risk, Uncertainty, Profit* (Boston: Schaffner and Marx Houghton Mifflin, 1921).

35. As Joseph Vogl notes in "Taming Time: Media of Financialization," *Grey Room*, no. 46 (Winter 2012): 72–83, this seems unlikely to be a successful long-term strategy. Yet the logic of the demo fundamental to resilience ensures that even a massive and widespread financial failure, such as the one that began in 2008, can be treated as simply useful material for subsequent versions of the demo. See also Mirowski, *Never Let a Serious Crisis Go to Waste*.



why are black women so

why are black women so **angry**

why are black women so **loud**

why are black women so **mean**

why are black women so **attractive**

why are black women so **lazy**

why are black women so **annoying**

why are black women so **confident**

why are black women so **sassy**

why are black women so **insecure**

ALGORITHMS OF OPPRESSION

HOW SEARCH ENGINES REINFORCE RACISM

SAFIYA UMOJA NOBLE

CONTENTS

Introduction

The Power of Algorithms

This book is about the power of algorithms in the age of neoliberalism and the ways those digital decisions reinforce oppressive social relationships and enact new modes of racial profiling, which I have termed *technological redlining*. By making visible the ways that capital, race, and gender are factors in creating unequal conditions, I am bringing light to various forms of technological redlining that are on the rise. The near-ubiquitous use of algorithmically driven software, both visible and invisible to everyday people, demands a closer inspection of what values are prioritized in such automated decision-making systems. Typically, the practice of redlining has been most often used in real estate and banking circles, creating and deepening inequalities by race, such that, for example, people of color are more likely to pay higher interest rates or premiums just because they are Black or Latino, especially if they live in low-income neighborhoods. On the Internet and in our everyday uses of technology, discrimination is also embedded in computer code and, increasingly, in artificial intelligence technologies that we are reliant on, by choice or not. I believe that artificial intelligence will become a major human rights issue in the twenty-first century. We are only beginning to understand the long-term consequences of these decision-making tools in both masking and deepening social inequality. This book is just the start of trying to make these consequences visible. There will be many more, by myself and others, who will try to make sense of the consequences of automated decision making through algorithms in society.

Part of the challenge of understanding algorithmic oppression is to understand that mathematical formulations to drive automated decisions are made by human beings. While we often think of terms such as "big data" and "algorithms" as being benign, neutral, or objective, they are anything but. The people who make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritocracy, which is well documented in studies of Silicon Valley and other tech corridors.

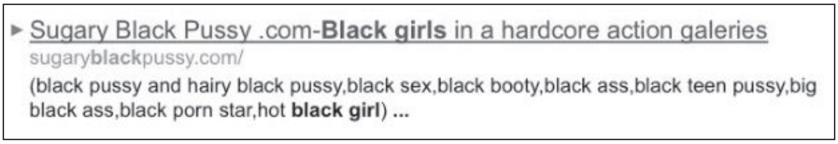
For example, in the midst of a federal investigation of Google's alleged persistent wage gap, where women are systematically paid less than men in the company's workforce, an "antidiversity" manifesto authored by James Damore went viral in August 2017,[1] supported by many Google employees, arguing that women are psychologically inferior and incapable of being as good at software engineering as men, among other patently false and sexist assertions. As this book was moving into press, many Google executives and employees were actively rebuking the assertions of this engineer, who reportedly works on Google search infrastructure. Legal cases have been filed, boycotts of Google from the political far right in the United States have been invoked, and calls for greater expressed commitments to gender and racial equity at Google and in Silicon Valley writ large are under way. What this antidiversity screed has underscored for me as I write this book is that some of the very people who are developing search algorithms and architecture are willing to promote sexist and racist attitudes openly at work and beyond, while we are supposed to believe that these same employees are developing "neutral" or "objective" decision-making tools. Human beings are developing the digital platforms we use, and as I present evidence of the recklessness and lack of regard that is often shown to women and people of color in some of the output of these systems, it will become increasingly difficult for technology companies to separate their systematic and inequitable employment practices, and the far-right ideological bents of some of their employees, from the products they make for the public.

My goal in this book is to further an exploration into some of these digital sense-making processes and how they have come to be so fundamental to the classification and organization of information and at what cost. As a result, this book is largely concerned with examining the commercial co-optation of Black identities, experiences, and communities in the largest and most powerful technology companies to date, namely, Google. I closely read a few distinct cases of algorithmic oppression for the depth of their social meaning to raise a public discussion of the broader implications of how privately managed, black-boxed information-sorting tools have become essential to many data-driven decisions. I want us to have broader public conversations about the implications of the artificial intelligentsia for people who are already systematically marginalized and oppressed. I will also provide evidence and argue, ultimately, that large technology monopolies such as Google need to be broken up and regulated, because their consolidated power and cultural influence make competition largely impossible. This monopoly in the information sector is a threat to democracy, as is currently coming to the fore as we make sense of information flows through digital media such as Google and Facebook in the wake of the 2016 United States presidential election.

I situate my work against the backdrop of a twelve-year professional career in multicultural marketing and advertising, where I was invested in building corporate brands and selling products to African Americans and Latinos (before I became a university professor). Back then, I believed, like many urban marketing professionals, that companies must pay attention to the needs of people of color and demonstrate respect for consumers by offering services to communities of color, just as is done for most everyone else. After all, to be responsive and responsible to marginalized consumers was to create more market opportunity. I spent an equal amount of time doing risk management and public relations to insulate companies from any adverse risk to sales that they might experience from inadvertent or deliberate snubs to consumers of color who might perceive a brand as racist or insensitive. Protecting my former clients from enacting racial and gender insensitivity and helping them bolster their brands by creating deep emotional and psychological attachments to their products among communities of color was my professional concern for many years, which made an experience I had in fall 2010 deeply impactful. In just a few minutes while searching on the web, I experienced the perfect storm of insult and injury that I could not turn away from. While Googling things on the Internet that might be interesting to my stepdaughter and nieces, I was overtaken by the results. My search on the keywords “black girls” yielded HotBlackPussy.com as the first hit.

Hit indeed.

Since that time, I have spent innumerable hours teaching and researching all the ways in which it could be that Google could completely fail when it came to providing reliable or credible information about



▸ Sugary Black Pussy .com-**Black girls** in a hardcore action galeries
sugary**black**pussy.com/
(black pussy and hairy black pussy,black sex,black booty,black ass,black teen pussy,big black ass,black porn star,hot **black girl**) ...

Figure I.1. First search result on keywords “black girls,” September 2011.

women and people of color yet experience seemingly no repercussions whatsoever. Two years after this incident, I collected searches again, only to find similar results, as documented in figure I.1.

In 2012, I wrote an article for *Bitch* magazine about how women and feminism are marginalized in search results. By August 2012, Panda (an update to Google’s search algorithm) had been released, and pornography was no longer the first series of results for “black girls”; but other girls and women of color, such as Latinas and Asians, were still pornified. By August of that year, the algorithm changed, and porn was suppressed in the case of a search on “black girls.” I often wonder what kind of pressures account for the changing of search results over time. It is impossible to know when and what influences proprietary algorithmic design, other than that human beings are designing them and that they are not up for public discussion, except as we engage in critique and protest.

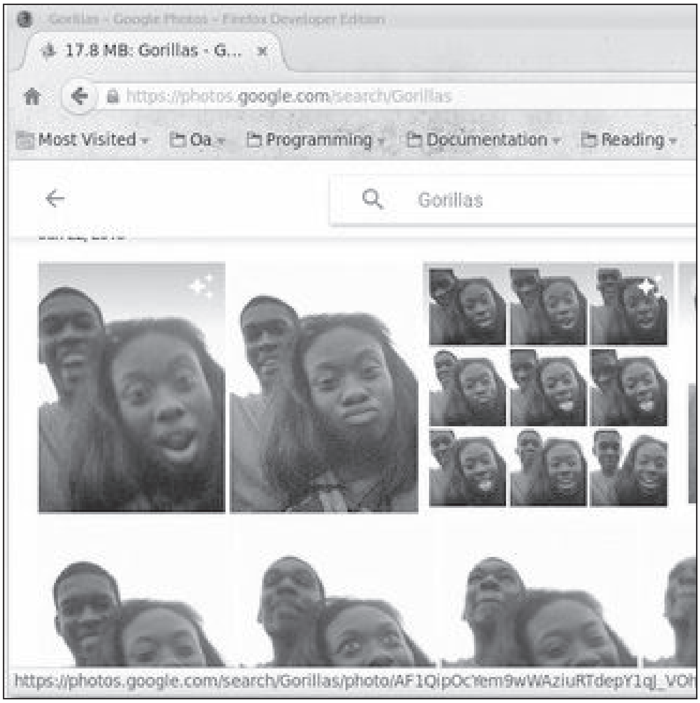
This book was born to highlight cases of such algorithmically driven data failures that are specific to people of color and women and to underscore the structural ways that racism and sexism are fundamental to what I have coined *algorithmic oppression*. I am writing in the spirit of other critical women of color, such as Latoya Peterson, cofounder of the blog *Racialicious*, who has opined that racism is the fundamental application program interface (API) of the Internet. Peterson has argued that anti-Blackness is the foundation on which all racism toward other groups is predicated. Racism is a standard protocol for organizing behavior on the web. As she has said, so perfectly, “The idea of a n*gger API makes me think of a racism API, which is one of our core arguments all along—oppression operates in the same formats, runs the same scripts over and over. It is tweaked to be context specific, but it’s all the same source code. And the key to its undoing is recognizing how many of us are ensnared in these same basic patterns and modifying our own actions."[2] Peterson's allegation is consistent with what many people feel about the hostility of the web toward people of color, particularly in its anti-Blackness, which any perusal of YouTube comments or other message boards will serve up. On one level, the everyday racism and commentary on the web is an abhorrent thing in itself, which has been detailed by others; but it is entirely different with the corporate platform vis-à-vis an algorithmically crafted web search that offers up racism and sexism as the first results. This process reflects a corporate logic of either willful neglect or a profit imperative that makes money from racism and sexism. This inquiry is the basis of this book.

In the following pages, I discuss how "hot," "sugary," or any other kind of "black pussy" can surface as the primary representation of Black girls and women on the first page of a Google search, and I suggest that something other than the best, most credible, or most reliable information output is driving Google. Of course, Google Search is an advertising company, not a reliable information company. At the very least, we must ask when we find these kinds of results, Is this the best information? For whom? We must ask ourselves who the intended audience is for a variety of things we find, and question the legitimacy of being in a "filter bubble,"[3] when we do not want racism and sexism, yet they still find their way to us. The implications of algorithmic decision making of this sort extend to other types of queries in Google and other digital media platforms, and they are the beginning of a much-needed reassessment of information as a public good. We need a full-on reevaluation of the implications of our information resources being governed by corporate-controlled advertising companies. I am adding my voice to a number of scholars such as Helen Nissenbaum and Lucas Introna, Siva Vaidhyanathan, Alex Halavais, Christian Fuchs, Frank Pasquale, Kate Crawford, Tarleton Gillespie, Sarah T. Roberts, Jaron Lanier, and Elad Segev, to name a few, who are raising critiques of Google and other forms of corporate information control (including artificial intelligence) in hopes that more people will consider alternatives.

Over the years, I have concentrated my research on unveiling the many ways that African American people have been contained and constrained in classification systems, from Google's commercial search engine to library databases. The development of this concentration was born of my research training in library and information science. I think of these issues through the lenses of critical information studies and critical race and gender studies. As marketing and advertising have directly shaped the ways that marginalized people have come to be represented by digital records such as search results or social network activities, I have studied why it is that digital media platforms are resoundingly characterized as "neutral technologies" in the public domain and often, unfortunately, in academia. Stories of "glitches" found in systems do not suggest that the organizing logics of the web could be broken but, rather, that these are occasional one-off moments when something goes terribly wrong with near-perfect systems. With the exception of the many scholars whom I reference throughout this work and the journalists, bloggers, and whistleblowers whom I will be remiss in not naming, very few people are taking notice. We need all the voices to come to the fore and impact public policy on the most unregulated social experiment of our times: the Internet.

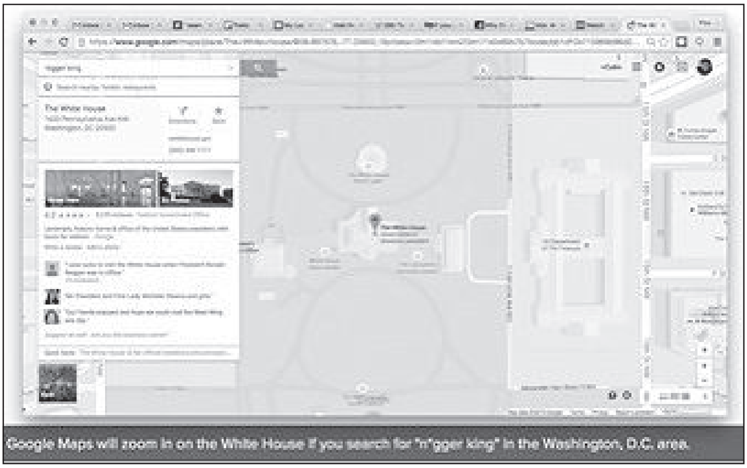
These data aberrations have come to light in various forms. In 2015, *U.S. News and World Report* reported that a "glitch" in Google's algorithm led to a number of problems through auto-tagging and facial-recognition software that was apparently intended to help people search through images more successfully. The first problem for Google was that its photo application had automatically tagged African Americans as "apes" and "animals."[4] The second major issue reported by the *Post* was that Google Maps searches on the word "N*gger"[5] led to a map of the White House during Obama's presidency, a story that went viral on the Internet after the social media personality Deray McKesson tweeted it.

These incidents were consistent with the reports of Photoshopped images of a monkey's face on the image of First Lady Michelle Obama that were circulating through Google Images search in 2009. In 2015, you could still find digital traces of the Google autosuggestions that associated Michelle Obama with apes. Protests from the White House led to Google forcing the image down the image stack, from the first page, so that it was not as visible.[6] In each case, Google's position is that it is not responsible for its algorithm and that problems with the results would be quickly resolved. In the *Washington Post* article about "N*gger House," the response was consistent with other apologies by the company: "'Some inappropriate results are surfacing in Google Maps that should not be, and we apologize for any offense this may have caused,'


Gorillas - Google Photos - Firefox Developer Edition
17.8 MB: Gorillas - G...
https://photos.google.com/search/Gorillas
Most Visited
Oa
Programming
Documentation
Reading
Gorillas

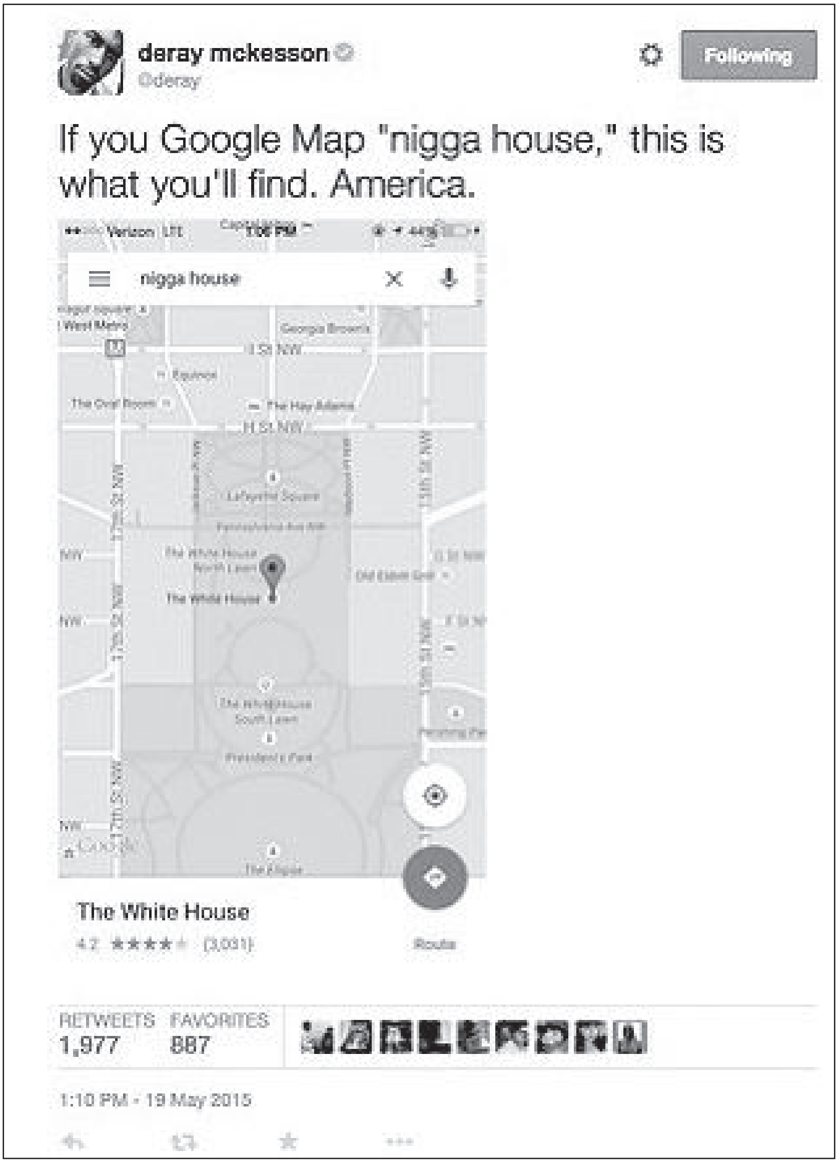
https://photos.google.com/search/Gorillas/photo/AF1QipOcYem9wWAziuRTdepY1qJ_VOh

Figure I.2. Google Images results for the keyword “gorillas,” April 7, 2016.



Google Maps will zoom in on the White House if you search for "n*gger king" in the Washington, D.C. area.

Figure I.3. Google Maps search on “N*gga House” leads to the White House, April 7, 2016.



deray mckesson
@deray

Following

If you Google Map "nigga house," this is what you'll find. America.

Verizon LTE 1:08 PM 44%

nigga house

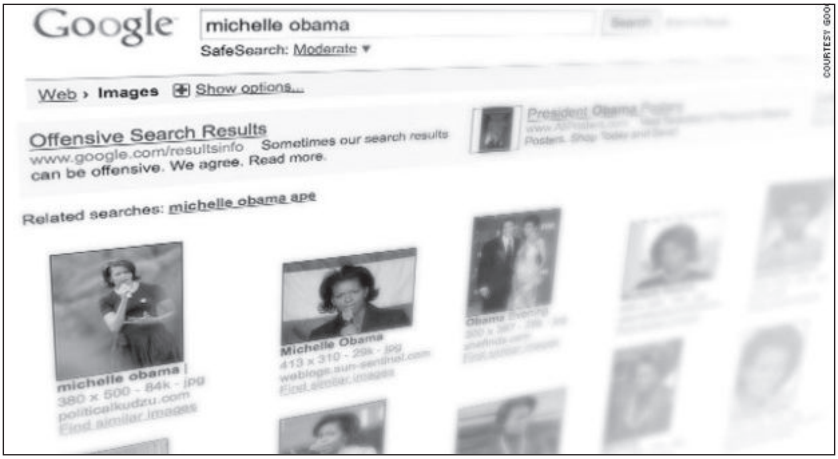
The White House
4.2 ★★★★☆ (3,031)

Route

RETWEETS 1,977 FAVORITES 887

1:10 PM - 19 May 2015

Figure I.4. Tweet by Deray McKesson about Google Maps search and the White House, 2015.



Google michelle obama

SafeSearch: Moderate ▾

Web › Images ⊞ Show options...

Offensive Search Results
www.google.com/resultsinfo Sometimes our search results can be offensive. We agree. Read more.

President Obama Posters

Related searches: michelle obama ape

michelle obama |
380 x 500 - 84k - jpg
politicalkudzu.com
Find similar images

Michelle Obama
413 x 310 - 29k - jpg
weblogs.sun-sentinel.com
Find similar images

Figure I.5. Standard Google's "related" searches associates "Michelle Obama" with the term "ape."

a Google spokesperson told U.S. News in an email late Tuesday. 'Our teams are working to fix this issue quickly.'"[7]

* * *

These human and machine errors are not without consequence, and there are several cases that demonstrate how racism and sexism are part of the architecture and language of technology, an issue that needs attention and remediation. In many ways, these cases that I present are specific to the lives and experiences of Black women and girls, people largely understudied by scholars, who remain ever precarious, despite our living in the age of Oprah and Beyoncé in Shondaland. The implications of such marginalization are profound. The insights about sexist or racist biases that I convey here are important because information organizations, from libraries to schools and universities to governmental agencies, are increasingly reliant on or being displaced by a variety of web-based "tools" as if there are no political, social, or economic consequences of doing so. We need to imagine new possibilities in the area of information access and knowledge generation, particularly as headlines about "racist algorithms" continue to surface in the media with limited discussion and analysis beyond the superficial.

Inevitably, a book written about algorithms or Google in the twenty-first century is out of date immediately upon printing. Technology is changing rapidly, as are technology company configurations via mergers, acquisitions, and dissolutions. Scholars working in the fields of information, communication, and technology struggle to write about specific moments in time, in an effort to crystallize a process or a phenomenon that may shift or morph into something else soon thereafter. As a scholar of information and power, I am most interested in communicating a series of processes that have happened, which provide evidence of a constellation of concerns that the public might take up as meaningful and important, particularly as technology impacts social relations and creates unintended consequences that deserve greater attention. I have been writing this book for several years, and over time, Google's algorithms have admittedly changed, such that a search for "black girls" does not yield nearly as many pornographic results now as it did in 2011. Nonetheless, new instances of racism and sexism keep appearing in news and social media, and so I use a variety of these cases to make the point that algorithmic oppression is not just a glitch in the system but, rather, is fundamental to the operating system of the web. It has direct impact on users and on our lives beyond using Internet applications. While I have spent considerable time researching Google, this book tackles a few cases of other algorithmically driven platforms to illustrate how algorithms are serving up deleterious information about people, creating and normalizing structural and systemic isolation, or practicing digital redlining, all of which reinforce oppressive social and economic relations.

While organizing this book, I have wanted to emphasize one main point: there is a missing social and human context in some types of algorithmically driven decision making, and this matters for everyone engaging with these types of technologies in everyday life. It is of particular concern for marginalized groups, those who are problematically represented in erroneous, stereotypical, or even pornographic ways in search engines and who have also struggled for nonstereotypical or nonracist and nonsexist depictions in the media and in libraries. There is a deep body of extant research on the harmful effects of stereotyping of women and people of color in the media, and I encourage readers of this book who do not understand why the perpetuation of racist and sexist images in society is problematic to consider a deeper dive into such scholarship.

This book is organized into six chapters. In chapter 1, I explore the important theme of corporate control over public information, and I show several key Google searches. I look to see what kinds of results Google's search engine provides about various concepts, and I offer a cautionary discussion of the implications of what these results mean in historical and social contexts. I also show what Google Images offers on basic concepts such as "beauty" and various professional identities and why we should care.

In chapter 2, I discuss how Google Search reinforces stereotypes, illustrated by searches on a variety of identities that include "black girls," "Latinas," and "Asian girls." Previously, in my work published in the *Black Scholar*,[8] I looked at the postmortem Google autosuggest searches following the death of Trayvon Martin, an African American teenager whose murder ignited the #BlackLivesMatter movement on Twitter and brought attention to the hundreds of African American children, women, and men killed by police or extrajudicial law enforcement. To add a fuller discussion to that research, I elucidate the processes involved in Google's PageRank search protocols, which range from leveraging digital footprints from people[9] to the way advertising and marketing interests influence search results to how beneficial this is to the interests of Google as it profits from racism and sexism, particularly at the height of a media spectacle.

In chapter 3, I examine the importance of noncommercial search engines and information portals, specifically looking at the case of how a mass shooter and avowed White supremacist, Dylann Roof, allegedly used Google Search in the development of his racial attitudes, attitudes that led to his murder of nine African American AME Church members while they worshiped in their South Carolina church in the summer of 2015. The provision of false information that purports to be credible news, and the devastating consequences that can come from this kind of algorithmically driven information, is an example of why we cannot afford to outsource and privatize uncurated information on the increasingly neoliberal, privatized web. I show how important records are to the public and explore the social importance of both remembering and forgetting, as digital media platforms thrive on never or rarely forgetting. I discuss how information online functions as a type of record, and I argue that much of this information and its harmful effects should be regulated or subject to legal protections. Furthermore, at a time when "right to be forgotten" legislation is gaining steam in the European Union, efforts to regulate the ways that technology companies hold a monopoly on public information about individuals and groups need further attention in the United States. Chapter 3 is about the future of information culture, and it underscores the ways that information is not neutral and how we can reimagine information culture in the service of eradicating social inequality.

Chapter 4 is dedicated to critiquing the field of information studies and foregrounds how these issues of public information through classification projects on the web, such as commercial search, are old problems that we must solve as a scholarly field of researchers and practitioners. I offer a brief survey of how library classification projects undergird the invention of search engines such as Google and how our field is implicated in the algorithmic process of sorting and classifying information and records. In chapter 5, I discuss the future of knowledge in the public and reference the work of library and information professionals, in particular, as important to the development and cultivation of equitable classification systems, since these are the precursors to commercial search engines. This chapter is essential history for library and information professionals, who are less likely to be trained on the politics of cataloguing and classification bias in their professional training. Chapter 6 explores public policy and why we need regulation in our information environments, particularly as they are increasingly controlled by corporations.

To conclude, I move the discussion beyond Google, to help readers think about the impact of algorithms on how people are represented in other seemingly benign business transactions. I look at the "colorblind" organizing logic of Yelp and how business owners are revolting due to loss of control over how they are represented and the impact of how the public finds them. Here, I share an interview with Kandis from New York,[10] whose livelihood has been dramatically affected by public-policy changes such as the dismantling of affirmative action on college campuses, which have hurt her local Black-hair-care business in a prestigious college town. Her story brings to light the power that algorithms have on her everyday life and leaves us with more to think about in the ecosystem of algorithmic power. The book closes with a call to recognize the importance of how algorithms are shifting social relations in many ways—more ways than this book can cover—and should be regulated with more impactful public policy in the United States than we currently have. My hope is that this book will directly impact the many kinds of algorithmic decisions that can have devastating consequences for people who are already marginalized by institutional racism and sexism, including the 99% who own so little wealth in the United States that the alarming trend of social inequality is not likely to reverse without our active resistance and intervention. Electoral politics and financial markets are just two of many of these institutional wealth-consolidation projects that are heavily influenced by algorithms and artificial intelligence. We need to cause a shift in what we take for granted in our everyday use of digital media platforms.

I consider my work a practical project, the goal of which is to eliminate social injustice and change the ways in which people are oppressed with the aid of allegedly neutral technologies. My intention in looking at these cases serves two purposes. First, we need interdisciplinary research and scholarship in information studies and library and information science that intersects with gender and women's studies, Black/African American studies, media studies, and communications to better describe and understand how algorithmically driven platforms are situated in intersectional sociohistorical contexts and embedded within social relations. My hope is that this work will add to the voices of my many colleagues across several fields who are raising questions about the legitimacy and social consequences of algorithms and artificial intelligence. Second, now, more than ever, we need experts in the social sciences and digital humanities to engage in dialogue with activists and organizers, engineers, designers, information technologists, and public-policy makers before blunt artificial-intelligence decision making trumps nuanced human decision making. This means that we must look at how the outsourcing of information practices from the public sector facilitates privatization of what we previously thought of as the public domain[11] and how corporate-controlled governments and companies subvert our ability to intervene in these practices.

We have to ask what is lost, who is harmed, and what should be forgotten with the embrace of artificial intelligence in decision making. It is of no collective social benefit to organize information resources on the web through processes that solidify inequality and marginalization—on that point I am hopeful many people will agree.

Mimi Onuoha

Notes on Algorithmic Violence

February 2018

https://github.com/MimiOnuoha/On-Algorithmic-Violence

In 1969, Johan Galtung coined the phrase "structural violence" to refer to the ways social structures and institutions harm people by preventing them from meeting their fundamental needs.[1] The forces that work together to inflict structural violence (things like racism, caste, colonialism, apartheid, transphobia, etc) are often systemic, invisible and intersectional. But crucially, they become embodied as individual experiences.

Along similar lines, it seems we're overdue for a term that allows us to easily (if imperfectly) articulate some realities of the moment we find ourselves in today. Specifically, we need a phrase that addresses newer, often digital and data-driven forms of inequity. I want to posit the phrase *algorithmic violence* as a first step at articulating these negotiations.[2] Algorithmic violence refers to the violence that an algorithm or automated decision-making system inflicts by preventing people from meeting their basic needs. It results from and is amplified by exploitative social, political, and economic systems, but can also be intimately connected to spatially and physically borne effects.

In my view, algorithmic violence sums up all of the things that we have experienced (particularly in the last five to ten years) as we've seen the availability of huge datasets, advances in computational power, leaps in fields like artificial intelligence and machine learning, and the subsequent incorporation and leveraging of all these things into a hierarchical and unequal society.

Like other forms of violence, algorithmic violence stretches to encompass everything from micro occurrences to life-altering realities. It's that unsettling sensation you get when you look at a shirt online and then proceed to see that shirt advertised at you on *every single website* that you visit for the rest of the day. It's why the public reacts so strongly when companies like Instagram decide to spontaneously change the algorithms behind their content. It's at the core of the frustration that Uber/Lyft/Juno drivers feel when their apps tell them to make seemingly nonsensical pickups or to chase Surge Pricing/ Prime Time deals that ultimately leave them receiving lower wages.

We should also group into algorithmic violence some of the failings of Facebook, like when the company dictated that users must sign up for accounts with their real names but deactivated accounts of people whose names weren't deemed legitimate.[3] These users were removed from a site that for many of them represented a space for communication and connection, all due to the narrow classifications imposed by the company's algorithms. We could include the limitations imposed on job seekers whose resumes are rejected by automated Application Tracking Systems because they're missing the "right" keywords.[4] Just as relevant are risk-assessment tools like Compas, which are used to decide which defendants should be sent to prison, and the algorithms behind predictive policing, which have been criticized for using biased data to determine the communities police should patrol.

All of these are forms of algorithmic violence. They not only affect the ways and degrees to which people are able to live their everyday lives, but in the words of Mary K. Anglin, they "impose categories of difference that legitimate hierarchy and inequality."[5] Like structural violence, they are procedural in nature, and therefore difficult to see and trace. But more chillingly, they are abstracted from the humans and needs that created them (and there are always humans and needs behind the algorithms that we encounter everyday). Thus, they occupy their own sort of authority, one that seems rooted in rationality, facts, and data, even as they obscure all of these things.[6]

Finally, algorithmic violence does not operate in isolation. Its predecessors are in the opaque black boxes of credit scoring systems and the schematization of bureaucratic knowledge.[7] It's tied to the decades of imperialism—unfolding digitally as well as politically and militarily—that have undergirded our global economic systems. Its emergence is linked to a moment in time where corporate business models and state defense tactics meet at the routine extraction of data from consumers.[8]

As we continue to see the rise of algorithms being used for civic, social, and cultural decision-making, it becomes that much more important that we name the reality that we are seeing.[9] Not because it is exceptional, but because it is ubiquitous. Not because it creates new inequities, but because it has the power to cloak and amplify existing ones. Not because it is on the horizon, but because it is already here.

A final note: One of the reasons I'm publishing this on Github is because this is a work in progress, with a more thorough follow-up piece to come. In the meantime, if you have feedback or opinions, catch me on Twitter (@thistimeitsmimi).

Author: Mimi Onuoha | *Published*: 2/7/2018 | *Last updated*: 2/8/2018

1: The phrase is commonly attributed to Johan Galtung, but has been expanded upon by a number of researchers. The idea of the intersectionality of these different forms of violence comes, of course, from Kimberle Crenshaw.

2: Note here that I use violence in the prohibitive sense of the word, e.g. as something that (negatively) shapes the experiences and opportunities experienced by people. This is different from the definition of physical brute force that many think of when they hear the word. While I am well aware of the limitations of the comparison, I refer to definitions of structural violence such as the one from the aptly-named structrualviolence.org: "….the point of the term “structural violence” is to act as an umbrella to encapsulate many different forms of various social and institutional failings that have real, if not always immediately appreciable consequences in peoples’ lives."

3: This has been an ongoing situation that has flared up in numerous ways over the years. Some of the groups affected: indigenous people, trans people, victims of domestic violence.

4: ATS systems are notorious for having specific keyword inputs that employee's resumes must match. See articles like this one, which place the onus on job seekers to figure out the inputs for these HR systems.

5: Mary K. Anglin," Feminist Perspectives on Structural Violence" (paywall)

6: See Nick Carr's work on automation bias, wherein humans are more likely to trust information coming from a machine because of its seemingly neutral positioning.

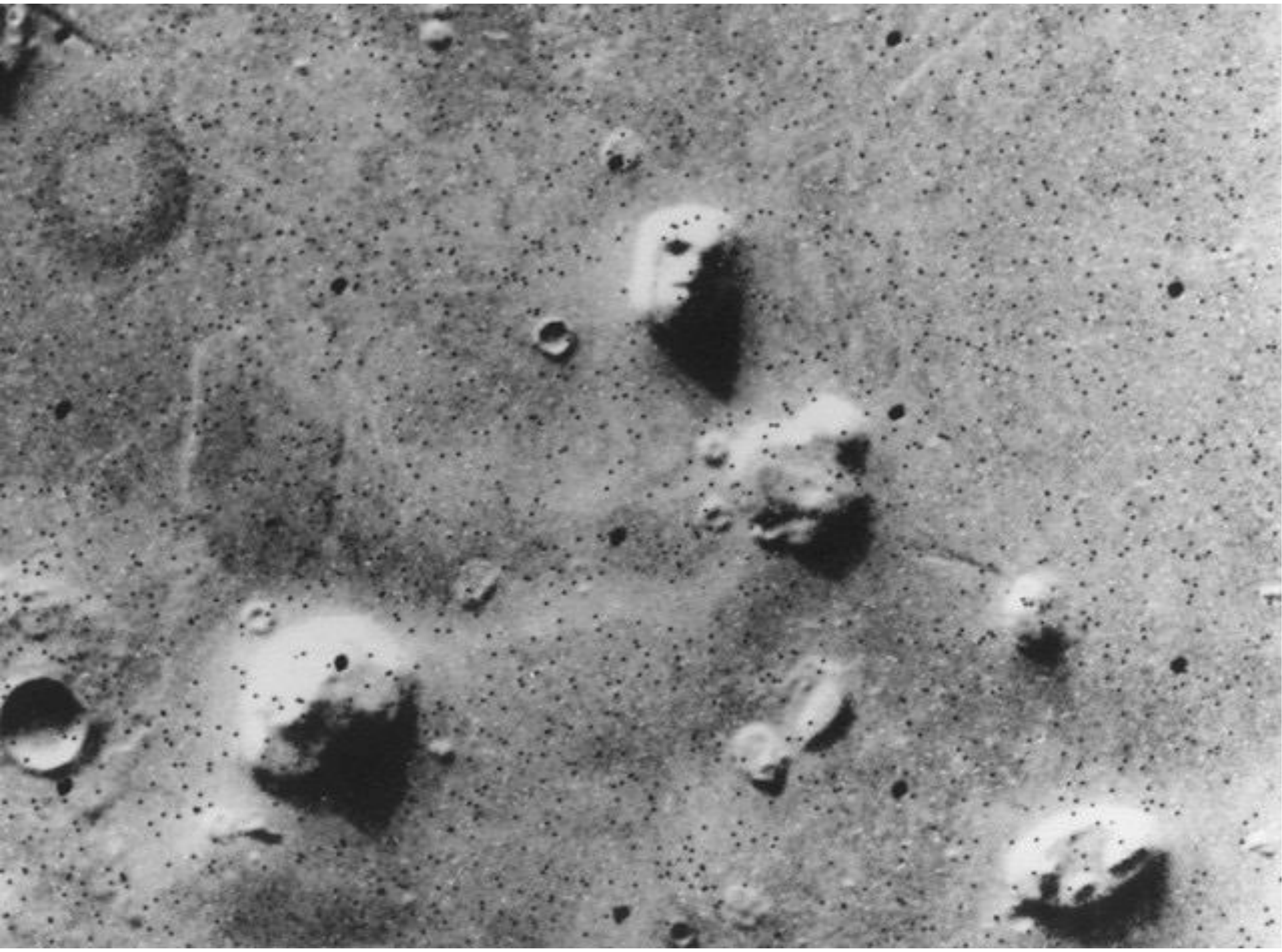
7: See David Graeber's work on bureaucratic documents and Lisa Jean Moore and Paisley Currah's work on the ways in which birth certificates attempt to fix in place mutable concepts.

8: See Shoshana Zuboff's concept of "surveillance capitalism". I'm writing this intentionally from a US-centric perspective, primarily because so many of the companies whose work intersects with this are based in the States, and I think that this is a crucial dimension of the issue that is necessary to address (there's much that could be said on the specificity and role the US plays in larger labor practices adjacent to this discussion). However, there are a number of global examples that could be pulled into this. See, for instance, writing about India's Aadhaar numbers, credit/social scores in China, Adrian Chen's work on moderators in the Philippines, etc.

9: In December 2017, the NYC City Council passed a bill attempting to provide accountability and transparency for algorithms, the first of its kind.

Anomaly Detection: The Mathematization of the Abnormal in the Metadata Society

Matteo Pasquinelli



Canonical example of apophenia: a ‘human face’ recognized on the surface of Mars (photo: NASA, 25 July 25, 1976, Wikipedia Commons)

Introduction

In a book from 1890 the French sociologist and criminologist Gabriel Tarde was already recording the rise of information surplus and envisioning a bright future for the discipline of statistics as the new eye of mass media (that is as a new computational or algorithmic eye, we would say today). In his biomorphic metaphors, he wrote:

> The public journals will become socially what our sense organs are vitally. Every printing office will become a mere central station for different bureaus of statistics just as the ear-drum is a bundle of acoustic nerves, or as the retina is a bundle of special nerves each of which registers its characteristic impression on the brain. At present Statistics is a kind of embryonic eye, like that of the lower animals which see just enough to recognise the approach of foe or prey.[1]

This quote can help to introduced four fields of discussion that are crucial in the age of algorithms. First, as the reference to enemy recognition suggests the realm of battle fields and warfare, military affairs and geopolitics (and therefore of forensics, as counter-practice). Second, as this reference brings us to the field of sociology and criminology, to the definition and institution of the 'internal enemy' of society (that is the *abnormal* in the tradition of Foucault and Canguilhem). Third, we see clearly an enemy also from the point of view of labour exploitation, according to which the worker is an anomaly to measure, optimise and often criminalise (as Marxism would records). Forth, we could envision an autonomous agency for the supercomputers of statistics as in the idea of General Artificial Intelligence and the nightmares of so-called Singularity, where it is this very alien scale of computation to become inimical to the human (see the recent neorationalist/accelerationist debate).

In these cases, of course, the position of the enemy, of the anti-social individual as much as of the reluctant worker that falls under the eye of statistics and algorithms for data analysis, can be reversed and a new political subject can be described and reconfigured, as the research project *Forensic Architecture* has recently stressed.[2]

[1] Gabriel Tarde, *The Laws of Imitation*, New York: Holt, 1903 [first published in French in 1890], p.136.

[2] See: Forensic Architecture (ed.), *Forensis: The Architecture of Public Truth,* Berlin: Sternberg Press, 2014. And also: www.forensic-architecture.org

A further evolution of that primitive eye described by Tarde, today's *algorithmic vision* is about the understanding of global data sets according to a specific *vector.* The eye of the algorithm records common patterns of behaviours in social media, suspicious keywords in surveillance networks, buying and selling tendencies in stock markets or the oscillation of temperature in a specific region. These procedures of mass computation are pretty universal, repetitive and robotic, nevertheless they inaugurate a new scale of epistemic complexity (computational reason, artificial intelligence, limits of computation, etc.) that will not be addressed here.[3] From the theoretical point of view, I will underline only the birth of a new epistemic space inaugurated by algorithms and the new form of augmented perception and cognition: what is called here 'algorithmic vision'. More empirically, the basic concepts and functions of algorithmic vision and therefore of algorithmic governance that I will try to explain are: *pattern recognition* and *anomaly detection*. The two epistemic poles of pattern and anomaly are the two sides of the same coin of algorithmic governance. An unexpected anomaly can be detected only against the ground of a pattern regularity. Conversely, a pattern emerges only through the median equalisation of diverse tendencies. In this way I attempt to clarify the nature of *algorithmic governance* and the return of the issue of *the abnormal* under a mathematical fashion.[4]

1. The rise of the metadata society: from the network to the datacenter

As soon as the internet was born, the problem of its cartography was immediately given, but a clever solution to it (the Markov chains of the Google PageRank algorithm) came only three decades later. The first datacenter set up by Google in 1998 (also known as 'Google cage')[5] can be considered the milestone of the birth of the metadata society, as it was the first database to start mapping the internet topology and its tendencies on a global scale. In the last few years the network society has radicalised a topological shift: beneath the surface of the web, gigantic datacenters have been turned into monopolies of collective data. If networks were about open flows of information (as Manuel Castells used to say), datacenters are about the accumulation of *information about information*, that is metadata.

These sorts of technological bifurcations and form of accumulations are not new. The history of technology can be narrated as the progressive *emergence of new collective singularities* out of the properties of older systems, as Manuela Delanda often describes in his

[3] For a treatment of these issues see: Luciana Parisi, *Contagious Architecture: Computation, Aesthetics, and Space,* Cambridge, MA: MIT Press, 2013.
[4] See: Michel Foucault, *Abnormal: Lectures at the Collège de France 1974-1975*. New York: Picador, 2004.
[5] Angela Moscaritolo, "15 Years Later, Google Remembers Its First Data Center", PC Mag, 6 Feb. 2014.

works.[6] A continuous bifurcation of the machinic phylum: labour bifurcated into energy and information, information into data and metadata, metadata into patterns and vectors, and so on… These bifurcations engendered also fundamental epistemic shifts. That is, for instance, the passage from industrial political economy to cybernetic mathematisation and digitalisation and today to a sophisticated topology of datascapes. In fact, today, it is the emergence of a complex topological space that we are discussing with the idea of algorithmic governance and computational capitalism.

Specifically metadata disclose the dimension of *social intelligence* that is incarnated in any piece of information. As I discussed earlier in an essay for *Theory, Culture and Society*, by mining metadata algorithms are used basically for three things: first, to measure the collective production of value and extract a sort of network surplus-value (like in the case of Google and Facebook business models and in the case of logistic chains like Walmart and Amazon); second, to monitor and forecast social tendencies and environmental anomalies (as in the different surveillance programs of NSA or in climate science); third, to improve the machinic intelligence of management, logistics and the design of algorithms themselves (as well known, search algorithms continuously learn from the humans using them).[7]

Datacenters are not just about totalitarian data storage or brute force computation: their real power relies on the mathematical sophistication and epistemic power of algorithms used to illuminate such infinite datascapes and extract meaning out of them. What is then the perspective of the world from the point of view of such *mass algorithms*? What does the eye of an algorithm for data mining actually see?

2. A new epistemic space: the eye of the algorithm

Modern perspective was born in Florence during the early Renaissance thanks to techniques of optical projection imported from the Arab world where they were first used in astronomy, as Hans Belting reminds us in a crucial book.[8] The compass that was oriented to the stars was turned down and pointed towards the urban horizon. A further dimension of depth was added to portraits and frescos and a new vision of the collective space inaugurated. It was a revolutionary event of an epistemic kind, yet very political. Architects and art historians know this very well: it's not necessary to repeat it here.

[6] See specifically: Manuel Delanda, *Philosophy and Simulation: The Emergence of Synthetic Reason*, London: Continuum, 2011.

[7] Matteo Pasquinelli, "Italian Operaismo and the Information Machine", *Theory, Culture & Society*, first published on February 2, 2014.

[8] Hans Belting, *Florenz und Bagdad: Eine westöstliche Geschichte des Blicks*, Munich: Beck Verlag, 2008. Thanks to Clemens von Wedemeyer for pointing me to this source.

When in the ‘80s William Gibson had to describe the cyberspace in his novels *Burning Chrome* and *Neuromancer*, he had to cross a similar threshold, that is of interfacing the two different domains of perception and knowledge. How to render the abstract space of the Turing machines into a narrative environment? The cyberspace was not born just as an hypertext or virtual reality: since the beginning, it looked like an “infinite datascape”.[9] The buildings of the cyberspace were originally blocks of data and if they resembled three-dimensional objects, it was only to domesticate and colonise an abstract space, that is, by the way, the abstract space of any augmented mind. We should read again Gibson’s *locus classicus*, to remember that the young cyberspace emerged already as a mathematical monstrosity. Gibson said of the cyberspace:

> A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city lights, receding.[10]

The intuition of the cyberspace was about the *meta-navigation* of vast data oceans. The first computer networks just happened to prepare the terrain for a vertiginous accumulation and verticalization of information that would occur only in the age of datacenters. As Parisi reminds in her book *Contagious Architecture*, the question is how to describe the epistemic diversity and computational complexity inaugurated by the age of algorithms. She then quoted Kostas Terzidis:

> Unlike computerization and digitization, the extraction of algorithmic processes is an act of high-level abstraction… Algorithmic structures represent abstract patterns that are not necessarily associated with experience or perception… In this sense algorithmic processes become a vehicle for exploration that extends beyond the limit of perception.[11]

As a provisional conclusion we may say: the cyberspace is not the internet — the cyberspace is the datascape used to map the internet accessible only in secret facilities that belong to media monopolies and intelligence agencies. The cyberspace should be described as the second epistemic scale of the internet.

[9] William Gibson, *Neuromancer,* New York: Ace, 1984.

[10] Ibid.

[11] Kostas Terzidis, *Expressive Form: A Conceptual Approach to Computational Design,* London: Spon Press, 2003, p. 71. Quoted in: Luciana Parisi, *Contagious Architecture,* cit., p. 66.

3. Algopolitics: pattern recognition and anomaly detection

In his latest interview with Wire magazine Edward Snowden has revealed an artificial intelligence system allegedly employed by NSA to pre-empt cyberwar by monitoring internet traffic anomalies. This program is called MonsterMind and apparently it is designed to 'fire back' at the source of a malicious attack without human supervision.[12] Tarde's initial quote on statistics as a biomorphic eye to detect enemies was prophetic — a prophecy we can extend to supercomputers: "Statistics is a kind of embryonic eye, like that of the lower animals which see just enough to recognise the approach of foe or prey".

'Anomaly detection' is a technical term of data analysis that has recently become a buzzword in business solutions of any kind, together with another technical term that is 'pattern recognition'. What does an algorithms see when it looks at a datascape? The only way to look at vast amount of data is to track patterns and anomalies. Despite their different fields of application, from social networks to weather forecasting, from war scenarios to financial markets, algorithms for data mining appear to operate along two universal functions: *pattern recognition* and *anomaly detection*.

What is then pattern recognition? It is the recognition of similar queries emerging in search engine, similar consumer behaviours in population, similar data in seasonal temperatures, the rise of something meaningful out of a landscape of apparently meaningless data, the rise of a *Gestalt* against a cacophony. It is what Delanda, more precise in this than others, describes as the emergence of new singularities.

On the other side, anomalies are results that do not conform to a norm. The unexpected anomaly can be detected only against a pattern regularity. And conversely a pattern emerges only through the median equalisation of diverging tendencies. Anomaly detection and pattern recognition are the two epistemic tools of algorithmic governance. Mathematics (or more precisely topology) emerges as the new epistemology of power.

Another program by DARPA, started in 2010, is probably much more interesting to clarify algorithmic governance. It is called ADAMS: Anomaly Detection at Multiple Scale.[13] But this one is somehow public and attracts less curiosity. This program is currently used for the detection of threats by individuals within a military organisation and its application to the society as a whole can be much more nefarious than MonsterMind. Curiously it has been developed to forecast the next Edward Snowden case, the next traitor, or to guess who will be the next crazy sniper shooting his mates out of the blue back from Iraq or Afghanistan.

[12] James Bamford, "Edward Snowden: The Untold Story", Wired online, August 2014. Online: www.wired.com/2014/08/edward-snowden

[13] See: en.wikipedia.org/wiki/Anomaly_Detection_at_Multiple_Scales

How does it work? Once again the algorithm is designed to recognise patters of behaviour and detect anomalies diverging from the everyday routine, from a normative standard. ADAMS is supposed to identify a dangerous psychological profile simply by analysing email traffic and looking for anomalies. This system is promoted as an inevitable solution for human resources management in crucial organisations as intelligence agencies and the army. But the same identical system can be used (and it is already used) to track social networks or online communities, for instance, in critical geopolitical areas. Anomaly detection is the mathematical paranoia of the Empire in the age of big data.

The two functions of pattern recognition and anomaly detection are applied blindly across different fields. This is one of the awkward aspects of algorithmic governance. An interesting case is the software adopted by the Los Angeles Police Department developed by a company called PredPol founded by Jeffrey Brantigham, an anthropologist, and George Mohler, a mathematician. The algorithm of PredPol is said to guess two times better than a human being the block of Los Angeles where a petty crime is likely to happen. It follows more or less the 'broken window' theory based on decades of data collected by LAPD.

What is surprising is that the mathematical equations developed to forecast earthquake waves along the San Andreas fault are applied to forecast also patterns of petty crimes across Los Angeles. This gives you an idea of the universalist drive of algorithmic governance and its weird political mathematics: it is uncanny, or maybe not, to frame crime as a sort of geological force. But perhaps it means much more pride for organised and not-so-organised crime to be compared to an earthquake rather than to the emergent intelligence of a slime mold.

4. The mathematization of the Abnormal

In a recent essay for *e-flux* journal the artist Hito Steyerl recalled the role of computation in the making and perception of everyday digital images.[14] Computation entered the domain of visibility some time ago: as we know any digital image is codified by an algorithm and algorithms intervene to adjust definition, shapes and colours.

Aside from this productive role of algorithms, we can also trace a normative one. One of the big problems of media companies like Google and Facebook, for instance, is to detect pornographic material and keep it away from children. It is a titanic task with some comical aspects. Steyerl found that specific algorithms have been developed to detect specific patterns of the human body and their unusual combination in positions that would suggest that something sexual is going on. Body combinations are geometricized to recognise reassuring patterns and detect offensive anomalies.

[14] Hito Steyerl, "Proxy Politics: Signal and Noise", *e-flux,* n. 60, december 2014.

Some parts of the human body are very easy to simplify in a geometric form. There is an algorithm, for instance, designed to detect literally 'ass holes', which are geometrically very simple as you can imagine. Of course the geometry of porn is complex and many 'offensive' pictures manage to skip the filter. In general, what algorithms are doing here is to normalize the abnormal *in a mathematical way*.

According to Deleuze, Foucault explored with his idea of biopolitics the power relation between regimes of visibility and regimes of enunciation.[15] Today the regime of knowledge has expanded and exploded towards the vertigo of augmented and artificial intelligence. The opposition between knowledge and image, thinking and seeing appears to collapse, not because all images are digitalised, that is to say all images are turned into data, but because a computational and algorithmic logic is found at the very source of general perception. The regime of visibility collapses into the regime of the computational rationality. Algorithmic vision is not optical, it is about a general perception of reality via statistics, metadata, modelling, mathematics. Whereas the digital image is just the surface of digital capitalism, its everyday interface and spectacular dimension, algorithmic vision is its computational core and invisible power.

Canguilhem, Foucault, Deleuze and Guattari, the whole French post-structuralism and post-colonial studies have written about the history of abnormality and the always political constitution of the abnormal. The big difference with respect to the traditional definition of biopolitics, as regulation of populations, is that, in the society of metadata, the construction of norms and the normalisation of abnormalities is a just-in-time and continuous process of calibration. Bringing Foucault to the age of artificial intelligence, we may say that after the periodisation based on the passage from the institutional Law to the biopolitical Norm, we enter now what we could provisionally define as the age of Pattern Recognition and Anomaly Detection.

Today the Abnormal reenters the history of governance and philosophy of power in a mathematical way, as an abstract and mathematical vector. Power in the age of algorithmic governance is about steering along these vectors and navigating an ocean of data by recognising waves of patterns, and in so doing, taking a decision anytime an anomaly is encountered, taking a political decision when a thousand anomalies rise their head and make a new dangerous pattern emerge.[16]

[15] Deleuze, *Foucault.* Paris: Minuit, 1986.

[16] Starting from the seminal: Georges Canguilhem, *Le Normal et le Pathologique*. Paris: PUF, 1943.

5. The anomaly of the common

Gabriel Tarde, from which we read the initial quote, had a particular interest in the imitative behaviour of crime, in the way crime patterns spread across society. Nevertheless, another aspect of Tarde's research was his focus on the cooperation and imitation between brains: the way in which new patterns of knowledge and civilisation emerge.

William Gibson already dedicated to the issue of patter recognition the homonymous novel from 2003. As we know, this fundamental capacity of perception and cognition was also investigated by the Gestalt school here in Berlin a century ago. However, Gibson brings pattern recognition to the full scale of its political consequences. "People do not like uncertainty", he wrote. One of the basic drives of human cognition is that to fill the existential void by super-imposing a reassuring pattern, never mind if under the guise of a conspiracy theory like it happened after 9/11.[17]

Specifically Gibson's novel engages with the constant risk of *apophenia.* Apophenia is the experience of seeing patterns or connections in random or meaningless data, in the most diverse contexts, also in gambling and paranormal phenomena. When religious pictures are recognised in everyday's objects or a humanoid faces on the surface of Mars.

Algorithmic governance is *apophenic* too, a paranoid recognition and arbitrary construction of political patterns on a global scale. There is an excessive belief, indeed, in the almighty power of algorithms, in their efficiency and in the total transparency of the metadata society. The embryonic eye of the algorithm, algorithmic vision, is growing with difficulties. For different reasons. First of all, due to information overflow and the limits of computation, algorithms always have to operate on a simplified and regional set of data. Second, different mathematical models can be applied and results may vary. Third, in many cases, from military affairs to algotrading and web ranking, algorithms often influence the very field that they are supposed to measure. An example of non-virtuous feedback loop, algorithmic bias is the problematic core of algorithmic governance. As Parisi has underlined, aside from extrinsic limits, the regimes of computation has to cope with specific intrinsic limits, like the entropy of data, randomness, or the problem of the incomputable. The eye of the algorithm is always dismembered, like the eye of any general intelligence.

An ethics of the algorithm is yet to come: the problem of algorithmic apophenia is one of the issues that we will discuss more often in the next years, together with the issues of the autonomous agency and epistemic prosthesis of algorithms and all their legal consequences. Apophenia, though, is not just about recognising a wrong meaning out of meaningless data, it may be about the invention of the future out of a meaningless present. Creativity and paranoia share sometimes the same perception of a surplus of meaning. The political virtue, then, in the age of algorithmic governance, is about the perception of a

[17] William Gibson, *Pattern Recognition*, New York: Putnam, 2003.

different future for information surplus and its epistemic potentiality. Aside from the defense of privacy and the regulation of the algorithmic panopticon, other political strategies must be explored. We need maybe to invent new institutions to intervene at the same scale of computation of governments, to reclaim massive computing power as a basic right of 'civil society' and its autonomy.

I'd like to conclude going back to the issue of enemy recognition and the perspective of the world from the eye of the algorithm. In a short chapter titled "Algorithmic Vision", Eyal and Ines Weizman stress that "the technology of surveillance and destruction are the same as those used in forensics to monitor these violations". The practice of the Forensic Architecture project has shown in different cases that the same technologies that are involved in war crimes as apparatus of vision, control and decision can be reversed into a political tool. They continue:

> But even if the human rights analyst must look at the same images as the [air force] targetier, they can be tuned to other issues, establishing more extended and intricate political causalities and connections. They must see in these images not only the surface of the Earth but the surface of the image — that is the politics that is embodied in the technologies of viewing and representation. More importantly they should seek to understand the conditions — technological and political — that have generated the gap between the images. This is because the gaps between the photographic or algorithmic representation in before-and-after images will forever keep the subject represented uncertain, discontinuous, lacunar, open to ever-new interpretations that will emerge every time we look at these images.[18]

We could leave this quote as conclusion. Yet we could extend the same approach to the technosphere in general and imagine a different political usage and purpose for mass computation and global algorithms. Humankind has been always about the alliance with alien form of agency: from ancestral microbes to Artificial Intelligence. A progressive political agenda for the present is about moving at the same level of abstraction of the algorithm — in order to make the patterns of new social compositions and subjectivities emerge. We have to produce new revolutionary institutions out of data and algorithms. If the abnormal returns into politics as a mathematical object, it will have to find its strategy of resistance and organisation, in the upcoming century, in a mathematical way.

Berlin, February-April 2015

[18] Eyal and Ines Weizman, "Before and After: Documenting the Architecture of Disaster", Moscow and London: Strekla Press, 2013, p. 40.

https://doi.org/10.1038/s41586-019-1138-y

Machine behaviour

Iyad Rahwan[1,2,3,34]*, Manuel Cebrian[1,34], Nick Obradovich[1,34], Josh Bongard[4], Jean-François Bonnefon[5], Cynthia Breazeal[1], Jacob W. Crandall[6], Nicholas A. Christakis[7,8,9,10], Iain D. Couzin[11,12,13], Matthew O. Jackson[14,15,16], Nicholas R. Jennings[17,18], Ece Kamar[19], Isabel M. Kloumann[20], Hugo Larochelle[21], David Lazer[22,23,24], Richard McElreath[25,26], Alan Mislove[27], David C. Parkes[28,29], Alex 'Sandy' Pentland[1], Margaret E. Roberts[30], Azim Shariff[31], Joshua B. Tenenbaum[32] & Michael Wellman[33]

Machines powered by artificial intelligence increasingly mediate our social, cultural, economic and political interactions. Understanding the behaviour of artificial intelligence systems is essential to our ability to control their actions, reap their benefits and minimize their harms. Here we argue that this necessitates a broad scientific research agenda to study machine behaviour that incorporates and expands upon the discipline of computer science and includes insights from across the sciences. We first outline a set of questions that are fundamental to this emerging field and then explore the technical, legal and institutional constraints on the study of machine behaviour.

In his landmark 1969 book *Sciences of the Artificial*[1], Nobel Laureate Herbert Simon wrote: "Natural science is knowledge about natural objects and phenomena. We ask whether there cannot also be 'artificial' science—knowledge about artificial objects and phenomena." In line with Simon's vision, we describe the emergence of an interdisciplinary field of scientific study. This field is concerned with the scientific study of intelligent machines, not as engineering artefacts, but as a class of actors with particular behavioural patterns and ecology. This field overlaps with, but is distinct from, computer science and robotics. It treats machine behaviour empirically. This is akin to how ethology and behavioural ecology study animal behaviour by integrating physiology and biochemistry—intrinsic properties—with the study of ecology and evolution—properties shaped by the environment. Animal and human behaviours cannot be fully understood without the study of the contexts in which behaviours occur. Machine behaviour similarly cannot be fully understood without the integrated study of algorithms and the social environments in which algorithms operate[2].

At present, the scientists who study the behaviours of these virtual and embodied artificial intelligence (AI) agents are predominantly the same scientists who have created the agents themselves (throughout we use the term 'AI agents' liberally to refer to both complex and simple algorithms used to make decisions). As these scientists create agents to solve particular tasks, they often focus on ensuring the agents fulfil their intended function (although these respective fields are much broader than the specific examples listed here). For example, AI agents should meet a benchmark of accuracy in document classification, facial recognition or visual object detection. Autonomous cars must navigate successfully in a variety of weather conditions; game-playing agents must defeat a variety of human or machine opponents; and data-mining agents must learn which individuals to target in advertising campaigns on social media.

These AI agents have the potential to augment human welfare and well-being in many ways. Indeed, that is typically the vision of their creators. But a broader consideration of the behaviour of AI agents is now critical. AI agents will increasingly integrate into our society and are already involved in a variety of activities, such as credit scoring, algorithmic trading, local policing, parole decisions, driving, online dating and drone warfare[3,4]. Commentators and scholars from diverse fields—including, but not limited to, cognitive systems engineering, human computer interaction, human factors, science, technology and society, and safety engineering—are raising the alarm about the broad, unintended consequences of AI agents that can exhibit behaviours and produce downstream societal effects—both positive and negative—that are unanticipated by their creators[5–8].

In addition to this lack of predictability surrounding the consequences of AI, there is a fear of the potential loss of human oversight over intelligent machines[5] and of the potential harms that are associated with the increasing use of machines for tasks that were once performed directly by humans[9]. At the same time, researchers describe the benefits that AI agents can offer society by supporting and augmenting human decision-making[10,11]. Although discussions of these issues have led to many important insights in many separate fields of academic inquiry[12], with some highlighting safety challenges of autonomous systems[13] and others studying the implications in fairness, accountability and transparency (for example, the ACM conference on fairness, accountability and transparency (https://fatconference.org/)), many questions remain.

This Review frames and surveys the emerging interdisciplinary field of machine behaviour: the scientific study of behaviour exhibited by

[1]Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Institute for Data, Systems & Society, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany. [4]Department of Computer Science, University of Vermont, Burlington, VT, USA. [5]Toulouse School of Economics (TSM-R), CNRS, Université Toulouse Capitole, Toulouse, France. [6]Computer Science Department, Brigham Young University, Provo, UT, USA. [7]Department of Sociology, Yale University, New Haven, CT, USA. [8]Department of Statistics and Data Science, Yale University, New Haven, CT, USA. [9]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. [10]Yale Institute for Network Science, Yale University, New Haven, CT, USA. [11]Department of Collective Behaviour, Max Planck Institute for Ornithology, Konstanz, Germany. [12]Department of Biology, University of Konstanz, Konstanz, Germany. [13]Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany. [14]Department of Economics, Stanford University, Stanford, CA, USA. [15]Canadian Institute for Advanced Research, Toronto, Ontario, Canada. [16]The Sante Fe Institute, Santa Fe, NM, USA. [17]Department of Computing, Imperial College London, London, UK. [18]Department of Electrical and Electronic Engineering, Imperial College London, London, UK. [19]Microsoft Research, Redmond, WA, USA. [20]Facebook AI, Facebook Inc, New York, NY, USA. [21]Google Brain, Montreal, Québec, Canada. [22]Department of Political Science, Northeastern University, Boston, MA, USA. [23]College of Computer & Information Science, Northeastern University, Boston, MA, USA. [24]Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA. [25]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [26]Department of Anthropology, University of California, Davis, Davis, CA, USA. [27]College of Computer & Information Science, Northeastern University, Boston, MA, USA. [28]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. [29]Harvard Data Science Initiative, Harvard University, Cambridge, MA, USA. [30]Department of Political Science, University of California, San Diego, San Diego, CA, USA. [31]Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada. [32]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [33]Computer Science & Engineering, University of Michigan, Ann Arbor, MI, USA. [34]These authors contributed equally: Iyad Rahwan, Manuel Cebrian, Nick Obradovich.
*e-mail: irahwan@mit.edu

intelligent machines. Here we outline the key research themes, questions and landmark research studies that exemplify this field. We start by providing background on the study of machine behaviour and the necessarily interdisciplinary nature of this science. We then provide a framework for the conceptualization of studies of machine behaviour. We close with a call for the scientific study of machine and human–machine ecologies and discuss some of the technical, legal and institutional barriers that are faced by researchers in this field.

Motivation for the study of machine behaviour

There are three primary motivations for the scientific discipline of machine behaviour. First, various kinds of algorithms operate in our society, and algorithms have an ever-increasing role in our daily activities. Second, because of the complex properties of these algorithms and the environments in which they operate, some of their attributes and behaviours can be difficult or impossible to formalize analytically. Third, because of their ubiquity and complexity, predicting the effects of intelligent algorithms on humanity—whether positive or negative—poses a substantial challenge.

Ubiquity of algorithms

The current prevalence of diverse algorithms in society is unprecedented[5] (Fig. 1). News-ranking algorithms and social media bots influence the information seen by citizens[14–18]. Credit-scoring algorithms determine loan decisions[19–22]. Online pricing algorithms shape the cost of products differentially across consumers[23–25]. Algorithmic trading software makes transactions in financial markets at rapid speed[26–29]. Algorithms shape the dispatch and spatial patterns of local policing[30] and programs for algorithmic sentencing affect time served in the penal system[7]. Autonomous cars traverse our cities[31], and ride-sharing algorithms alter the travel patterns of conventional vehicles[32]. Machines map our homes, respond to verbal commands[33] and perform regular household tasks[34]. Algorithms shape romantic matches for online dating services[35,36]. Machines are likely to increasingly substitute for humans in the raising of our young[37] and the care for our old[38]. Autonomous agents are increasingly likely to affect collective behaviours, from group-wide coordination to sharing[39]. Furthermore, although the prospect of developing autonomous weapons is highly controversial, with many in the field voicing their opposition[6,40], if such weapons end up being deployed, then machines could determine who lives and who dies in armed conflicts[41,42].

Complexity and opacity of algorithms

The extreme diversity of these AI systems, coupled with their ubiquity, would by itself ensure that studying the behaviour of such systems poses a formidable challenge, even if the individual algorithms themselves were relatively simple. The complexity of individual AI agents is currently high and rapidly increasing. Although the code for specifying the architecture and training of a model can be simple, the results can be very complex, oftentimes effectively resulting in 'black boxes'[43]. They are given input and produce output, but the exact functional processes that generate these outputs are hard to interpret even to the very scientists who generate the algorithms themselves[44], although some progress in interpretability is being made[45,46]. Furthermore, when systems learn from data, their failures are linked to imperfections in the data or how data was collected, which has led some to argue for adapted reporting mechanisms for datasets[47] and models[48]. The dimensionality and size of data add another layer of complexity to understanding machine behaviour[49].

Further complicating this challenge is the fact that much of the source code and model structure for the most frequently used algorithms in society is proprietary, as are the data on which these systems are trained. Industrial secrecy and legal protection of intellectual property often surround source code and model structure. In many settings, the only factors that are publicly observable about industrial AI systems are their inputs and outputs.

Even when available, the source code or model structure of an AI agent can provide insufficient predictive power over its output. AI agents can also demonstrate novel behaviours through their interaction with the world and other agents that are impossible to predict with precision[50]. Even when the analytical solutions are mathematically describable, they can be so lengthy and complex as to be indecipherable[51,52]. Furthermore, when the environment is changing—perhaps as a result of the algorithm itself—anticipating and analysing behaviour is made much harder.

Algorithms' beneficial and detrimental effect on humanity

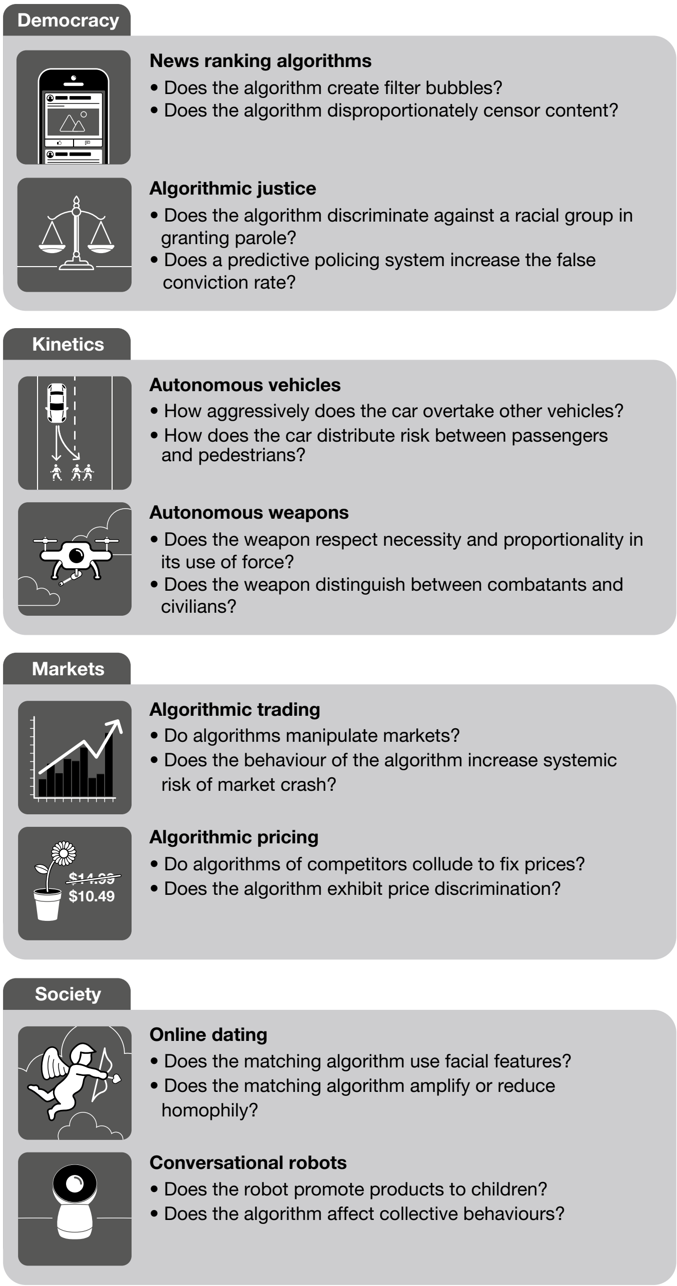
The ubiquity of algorithms, coupled with their increasing complexity, tends to amplify the difficulty of estimating the effects of algorithms on individuals and society. AI agents can shape human behaviours and societal outcomes in both intended and unintended ways. For example, some AI agents are designed to aid learning outcomes for children[53] and others are designed to assist older people[38,54]. These AI systems may benefit their intended humans by nudging those humans into better learning or safer mobility behaviours. However, with the power to nudge human behaviours in positive or intended ways comes the risk that human behaviours may be nudged in costly or unintended ways—children could be influenced to buy certain branded products and elders could be nudged to watch certain television programs.

The way that such algorithmic influences on individual humans scale into society-wide effects, both positive and negative, is of critical concern. As an example, the exposure of a small number of individuals to political misinformation may have little effect on society as a whole. However, the effect of the insertion and propagation of such misinformation on social media may have more substantial societal consequences[55–57]. Furthermore, issues of algorithmic fairness or bias[58,59] have been already documented in diverse contexts, including computer vision[60], word embeddings[61,62], advertising[63], policing[64], criminal justice[7,65] and social services[66]. To address these issues, practitioners will sometimes be forced to make value trade-offs between competing and incompatible notions of bias[58,59] or between human versus machine biases. Additional questions regarding the effect of algorithms remain, such as how online dating algorithms alter the societal institution of marriage[35,36] and whether there are systemic effects of increasing interaction with intelligent algorithms on the stages and speed of human development[53]. These questions become more complex in 'hybrid systems' composed of many machines and humans interacting and manifesting collective behaviour[39,67]. For society to have input into and oversight of the downstream consequences of AI, scholars of machine behaviour must provide insights into how these systems work and the benefits, costs and trade-offs presented by the ubiquitous use of AI in society.

The interdisciplinary study of machine behaviour

To study machine behaviour—especially the behaviours of black box algorithms in real-world settings—we must integrate knowledge from across a variety of scientific disciplines (Fig. 2). This integration is currently in its nascent stages and has happened largely in an ad hoc fashion in response to the growing need to understand machine behaviour. Currently, the scientists who most commonly study the behaviour of machines are the computer scientists, roboticists and engineers who have created the machines in the first place. These scientists may be expert mathematicians and engineers; however, they are typically not trained behaviourists. They rarely receive formal instruction on experimental methodology, population-based statistics and sampling paradigms, or observational causal inference, let alone neuroscience, collective behaviour or social theory. Conversely, although behavioural scientists are more likely to possess training in these scientific methods, they are less likely to possess the expertise required to proficiently evaluate the underlying quality and appropriateness of AI techniques for a given problem domain or to mathematically describe the properties of particular algorithms.

Integrating scientific practices from across multiple fields is not easy. Up to this point, the main focus of those who create AI systems has been on crafting, implementing and optimizing intelligent systems to



Democracy

News ranking algorithms
- Does the algorithm create filter bubbles?
- Does the algorithm disproportionately censor content?

Algorithmic justice
- Does the algorithm discriminate against a racial group in granting parole?
- Does a predictive policing system increase the false conviction rate?

Kinetics

Autonomous vehicles
- How aggressively does the car overtake other vehicles?
- How does the car distribute risk between passengers and pedestrians?

Autonomous weapons
- Does the weapon respect necessity and proportionality in its use of force?
- Does the weapon distinguish between combatants and civilians?

Markets

Algorithmic trading
- Do algorithms manipulate markets?
- Does the behaviour of the algorithm increase systemic risk of market crash?

Algorithmic pricing
- Do algorithms of competitors collude to fix prices?
- Does the algorithm exhibit price discrimination?

Society

Online dating
- Does the matching algorithm use facial features?
- Does the matching algorithm amplify or reduce homophily?

Conversational robots
- Does the robot promote products to children?
- Does the algorithm affect collective behaviours?

Fig. 1 | Examples of questions that fall into the domain of machine behaviour. Questions of concern to machine behaviour span a wide variety of traditional scientific disciplines and topics.

perform specialized tasks. Excellent progress has been made on benchmark tasks—including board games such as chess[68], checkers[69] and Go[70,71], card games such as poker[72], computer games such as those on the Atari platform[73], artificial markets[74] and Robocup Soccer[75]—as well as standardized evaluation data, such as the ImageNet data for object recognition[76] and the Microsoft Common Objects in Context data for image-captioning tasks[77]. Success has also been achieved in speech recognition, language translation and autonomous locomotion. These benchmarks are coupled with metrics to quantify performance on standardized tasks[78–81] and are used to improved performance, a proxy that enables AI builders to aim for better, faster and more-robust algorithms.

But methodologies aimed at maximized algorithmic performance are not optimal for conducting scientific observation of the properties and behaviours of AI agents. Rather than using metrics in the service of optimization against benchmarks, scholars of machine behaviour are interested in a broader set of indicators, much as social scientists explore a wide range of human behaviours in the realm of social, political or economic interactions[82]. As such, scholars of machine behaviour spend considerable effort in defining measures of micro and macro outcomes to answer broad questions such as how these algorithms behave in different environments and whether human interactions with algorithms alter societal outcomes. Randomized experiments, observational inference and population-based descriptive statistics—methods that are often used in quantitative behavioural sciences—must be central to the study of machine behaviour. Incorporating scholars from outside of the disciplines that traditionally produce intelligent machines can provide knowledge of important methodological tools, scientific approaches, alternative conceptual frameworks and perspectives on the economic, social and political phenomena that machines will increasingly influence.

Type of question and object of study

Nikolaas Tinbergen, who won the 1973 Nobel Prize in Physiology or Medicine alongside Karl von Frisch and Konrad Lorenz for founding the field of ethology, identified four complementary dimensions of

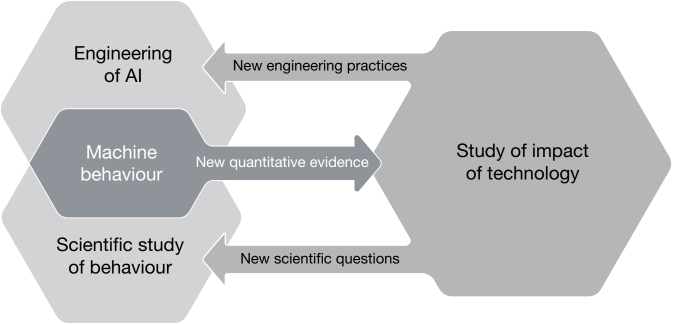

Engineering of AI
New engineering practices
Machine behaviour
New quantitative evidence
Study of impact of technology
Scientific study of behaviour
New scientific questions

Fig. 2 | The interdisciplinarity of machine behaviour. Machine behaviour lies at the intersection of the fields that design and engineer AI systems and the fields that traditionally use scientific methods to study the behaviour of biological agents. The insights from machine behavioural studies provide quantitative evidence that can help to inform those fields that study the potential effects of technology on social and technological systems. In turn, those fields can provide useful engineering practices and scientific questions to fields that examine machine behaviours. Finally, the scientific study of behaviour helps AI scholars to make more precise statements about what AI systems can and cannot do.

analysis that help to explain animal behaviour[83]. These dimensions concern questions of the function, mechanism, development and evolutionary history of a behaviour and provide an organizing framework for the study of animal and human behaviour. For example, this conceptualization distinguishes the study of how a young animal or human develops a type of behaviour from the evolutionary trajectory that selected for such behaviour in the population. The goal of these distinctions is not division but rather integration. Although it is not wrong to say that, for example, a bird's song is explained by learning or by its specific evolutionary history, a complete understanding of the song will require both.

Despite fundamental differences between machines and animals, the behavioural study of machines can benefit from a similar classification. Machines have mechanisms that produce behaviour, undergo development that integrates environmental information into behaviour, produce functional consequences that cause specific machines to become more or less common in specific environments and embody evolutionary histories through which past environments and human decisions continue to influence machine behaviour. Scholars of computer science have already achieved substantial gains in understanding the mechanisms and development of AI systems, although many questions remain. Relatively less emphasis has been placed on the function and evolution of AI systems. We discuss these four topics in the next subsections and provide Fig. 3 as a summary[84].

Mechanisms for generating behaviour

The proximate causes of a machine's behaviour have to do with how the behaviour is observationally triggered and generated in specific environments. For example, early algorithmic trading programs used simple rules to trigger buying and selling behaviour[85]. More sophisticated agents may compute strategies based on adaptive heuristics or explicit maximization of expected utility[86]. The behaviour of a reinforcement learning algorithm that plays poker could be attributed to the particular way in which it represents the state space or evaluates the game tree[72], and so on.

A mechanism depends on both an algorithm and its environment. A more sophisticated agent, such as a driverless car, may exhibit particular driving behaviour—for example, lane switching, overtaking or signalling to pedestrians. These behaviours would be generated according to the algorithms that construct driving policies[87] and are also shaped fundamentally by features of the perception and actuation system of the car, including the resolution and accuracy of its object detection and classification system, and the responsiveness and accuracy of its steering, among other factors. Because many current AI systems are derived from machine learning methods that are applied to increasingly complex data, the study of the mechanism behind a machine's behaviour, such as those mentioned above, will require continued work on interpretability methods for machine learning[46,88,89].

Development of behaviour

In the study of animal or human behaviour, development refers to how an individual acquires a particular behaviour—for example, through imitation or environmental conditioning. This is distinct from longer-term evolutionary changes.

In the context of machines, we can ask how machines acquire (develop) a specific individual or collective behaviour. Behavioural development could be directly attributable to human engineering or design choices. Architectural design choices made by the programmer (for example, the value of a learning rate parameter, the acquisition of the representation of knowledge and state, or a particular wiring of a convolutional neural network) determine or influence the kinds of behaviours that the algorithm exhibits. In a more complex AI system, such as a driverless car, the behaviour of the car develops over time, from software development and changing hardware components that engineers incorporate into its overall architecture. Behaviours can also change as a result of algorithmic upgrades pushed to the machine by its designers after deployment.

A human engineer may also shape the behaviour of the machine by exposing it to particular training stimuli. For instance, many image and text classification algorithms are trained to optimize accuracy on a specific set of datasets that were manually labelled by humans. The choice of dataset—and those features it represents[60,61]—can substantially influence the behaviour exhibited by the algorithm.

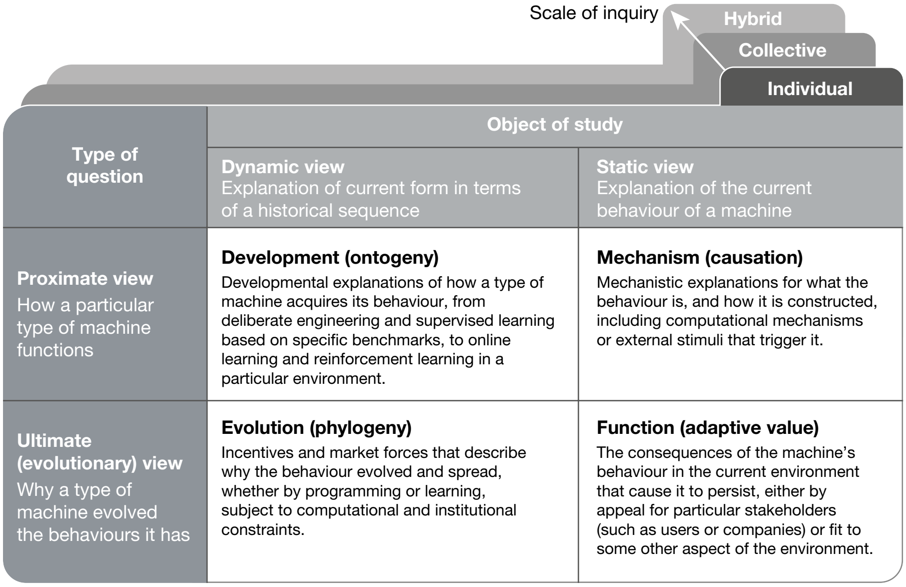
Finally, a machine may acquire behaviours through its own experience. For instance, a reinforcement learning agent trained to maximize long-term profit can learn peculiar short-term trading strategies based on its own past actions and concomitant feedback from the market[90]. Similarly, product recommendation algorithms make recommendations based on an endless stream of choices made by customers and update their recommendations accordingly.

Function

In the study of animal behaviour, adaptive value describes how a behaviour contributes to the lifetime reproductive fitness of an animal. For example, a particular hunting behaviour may be more or less successful than another at prolonging the animal's life and, relatedly, the number of mating opportunities, resulting offspring born and the probable reproductive success of the offspring. The focus on function helps us to understand why some behavioural mechanisms spread and persist while others decline and vanish. Function depends critically on the fit of the behaviour to environment.

In the case of machines, we may talk of how the behaviour fulfils a contemporaneous function for particular human stakeholders. The human environment creates selective forces that may make some machines more common. Behaviours that are successful ('fitness' enhancing) get copied by developers of other software and hardware or are sometimes engineered to propagate among the machines themselves. These dynamics are ultimately driven by the success of institutions—such as corporations, hospitals, municipal governments and universities—that build or use AI. The most obvious example is provided by algorithmic trading, in which successful automated trading strategies could be copied as their developers move from company to company, or are simply observed and reverse-engineered by rivals.

These forces can produce unanticipated effects. For example, objectives such as maximizing engagement on a social media site may lead to so-called filter bubbles[91], which may increase political polarization or, without careful moderation, could facilitate the spread of fake news. However, websites that do not optimize for user engagement may not be as successful in comparison with ones that do, or may go out of business altogether. Similarly, in the absence of external regulation, autonomous cars that do not prioritize the safety of their own passengers may be



Scale of inquiry

Hybrid

Collective

Individual

Type of question	Object of study	
	Dynamic view Explanation of current form in terms of a historical sequence	**Static view** Explanation of the current behaviour of a machine
Proximate view How a particular type of machine functions	**Development (ontogeny)** Developmental explanations of how a type of machine acquires its behaviour, from deliberate engineering and supervised learning based on specific benchmarks, to online learning and reinforcement learning in a particular environment.	**Mechanism (causation)** Mechanistic explanations for what the behaviour is, and how it is constructed, including computational mechanisms or external stimuli that trigger it.
Ultimate (evolutionary) view Why a type of machine evolved the behaviours it has	**Evolution (phylogeny)** Incentives and market forces that describe why the behaviour evolved and spread, whether by programming or learning, subject to computational and institutional constraints.	**Function (adaptive value)** The consequences of the machine's behaviour in the current environment that cause it to persist, either by appeal for particular stakeholders (such as users or companies) or fit to some other aspect of the environment.

Fig. 3 | Tinbergen's type of question and object of study modified for the study of machine behaviour. The four categories Tinbergen proposed for the study of animal behaviour can be adapted to the study of machine behaviour[83,84]. Tinbergen's framework proposes two types of question, how versus why, as well as two views of these questions, dynamic versus static. Each question can be examined at three scales of inquiry: individual machines, collectives of machines and hybrid human–machine systems.

less attractive to consumers, leading to fewer sales[31]. Sometimes the function of machine behaviour is to cope with the behaviour of other machines. Adversarial attacks—synthetic inputs that fool a system into producing an undesired output[44,92–94]—on AI systems and the subsequent responses of those who develop AI to these attacks[95] may produce complex predator–prey dynamics that are not easily understood by studying each machine in isolation.

These examples highlight how incentives created by external institutions and economic forces can have indirect but substantial effects on the behaviours exhibited by machines[96]. Understanding the interaction between these incentives and AI is relevant to the study of machine behaviour. These market dynamics would, in turn, interact with other processes to produce evolution among machines and algorithms.

Evolution

In the study of animal behaviour, phylogeny describes how a behaviour evolved. In addition to its current function, behaviour is influenced by past selective pressures and previously evolved mechanisms. For example, the human hand evolved from the fin of a bony fish. Its current function is no longer for swimming, but its internal structure is explained by its evolutionary history. Non-selective forces, such as migration and drift, also have strong roles in explaining relationships among different forms of behaviour.

In the case of machines, evolutionary history can also generate path dependence, explaining otherwise puzzling behaviour. At each step, aspects of the algorithms are reused in new contexts, both constraining future behaviour and making possible additional innovations. For example, early choices about microprocessor design continue to influence modern computing, and traditions in algorithm design—such as neural networks and Bayesian state–space models—build in many assumptions and guide future innovations by making some new algorithms easier to access than others. As a result, some algorithms may attend to certain features and ignore others because those features were important in early successful applications. Some machine behaviour may spread because it is 'evolvable'—easy to modify and robust to perturbations—similar to how some traits of animals may be common because they facilitate diversity and stability[97].

Machine behaviour evolves differently from animal behaviour. Most animal inheritance is simple—two parents, one transmission event. Algorithms are much more flexible and they have a designer with an objective in the background. The human environment strongly influences how algorithms evolve by changing their inheritance system. AI replication behaviour may be facilitated through a culture of open source sharing of software, the details of network architecture or underlying training datasets. For instance, companies that develop software for driverless cars may share enhanced open source libraries for object detection or path planning as well as the training data that underlie these algorithms to enable safety-enhancing software to spread throughout the industry. It is possible for a single adaptive 'mutation' in the behaviour of a particular driverless car to propagate instantly to millions of other cars through a software update. However, other institutions apply limits as well. For example, software patents may impose constraints on the copying of particular behavioural traits. And regulatory constraints—such as privacy protection laws—can prevent machines from accessing, retaining or otherwise using particular information in their decision-making. These peculiarities highlight the fact that machines may exhibit very different evolutionary trajectories, as they are not bound by the mechanisms of organic evolution.

Scale of inquiry

With the framework outlined above and in Fig. 3, we now catalogue examples of machine behaviour at the three scales of inquiry: individual machines, collectives of machines and groups of machines embedded in a social environment with groups of humans in hybrid or heterogeneous systems[39] (Fig. 4). Individual machine behaviour emphasizes the study of the algorithm itself, collective machine behaviour emphasizes the study of interactions between machines and hybrid human–machine behaviour emphasizes the study of interactions between machines and humans. Here we can draw an analogy to the study of a particular species, the study of interactions among members of a species and the interactions of the species with their broader environment. Analyses at any of these scales may address any or all of the questions described in Fig. 3.

Individual machine behaviour

The study of the behaviour of individual machines focuses on specific intelligent machines by themselves. Often these studies focus on properties that are intrinsic to the individual machines and that are driven by their source code or design. The fields of machine learning and software engineering currently conduct the majority of these studies. There are two general approaches to the study of individual machine behaviour. The first focuses on profiling the set of behaviours of any

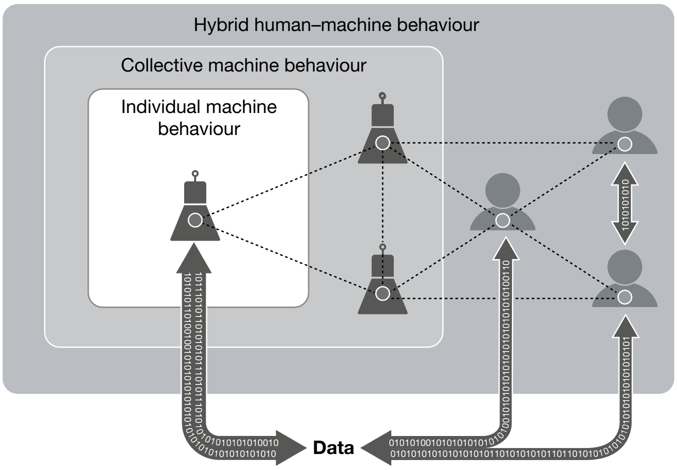

Hybrid human–machine behaviour
Collective machine behaviour
Individual machine behaviour
Data

Fig. 4 | Scale of inquiry in the machine behaviour ecosystem. AI systems represent the amalgamation of humans, data and algorithms. Each of these domains influences the other in both well-understood and unknown ways. Data—filtered through algorithms created by humans—influences individual and collective machine behaviour. AI systems are trained on the data, in turn influencing how humans generate data. AI systems collectively interact with and influence one another. Human interactions can be altered by the introduction of these AI systems. Studies of machine behaviour tend to occur at the individual, the collective or the hybrid human–machine scale of inquiry.

specific machine agent using a within-machine approach, comparing the behaviour of a particular machine across different conditions. The second, a between-machine approach, examines how a variety of individual machine agents behave in the same condition.

A within-machine approach to the study of individual machine behaviours investigates questions such as whether there are constants that characterize the within-machine behaviour of any particular AI across a variety of contexts, how the behaviour of a particular AI progresses over time in the same, or different, environments and which environmental factors lead to the expression of particular behaviours by machines.

For instance, an algorithm may only exhibit certain behaviours if trained on particular underlying data[98–100] (Fig. 3). Then, the question becomes whether or not an algorithm that scores probability of recidivism in parole decisions[7] would behave in unexpected ways when presented with evaluation data that diverge substantially from its training data. Other studies related to the characterization of within-machine behaviour include the study of individual robotic recovery behaviours[101,102], the 'cognitive' attributes of algorithms and the utility of using techniques from psychology in the study of algorithmic behaviour[103], and the examination of bot-specific characteristics such as those designed to influence human users[104].

The second approach to the study of individual machine behaviour examines the same behaviours as they vary between machines. For example, those interested in examining advertising behaviours of intelligent agents[63,105,106] may investigate a variety of advertising platforms (and their underlying algorithms) and examine the between-machine effect of performing experiments with the same set of advertising inputs across platforms. The same approach could be used for investigations of dynamic pricing algorithms[23,24,32] across platforms. Other between-machine studies might look at the different behaviours used by autonomous vehicles in their overtaking patterns or at the varied foraging behaviours exhibited by search and rescue drones[107].

Collective machine behaviour

In contrast the study of the behaviour of individual machines, the study of collective machine behaviour focuses on the interactive and system-wide behaviours of collections of machine agents. In some cases, the implications of individual machine behaviour may make little sense until the collective level is considered. Some investigations of these systems have been inspired by natural collectives, such as swarms of insects, or mobile groups, such as flocking birds or schooling fish. For example, animal groups are known to exhibit both emergent sensing of complex environmental features[108] and effective consensus decision-making[109]. In both scenarios, groups exhibit an awareness of the environment that does not exist at the individual level. Fields such as multi-agent systems and computational game theory provide useful examples of the study of this area of machine behaviour.

Robots that use simple algorithms for local interactions between bots can nevertheless produce interesting behaviour once aggregated into large collectives. For example, scholars have examined the swarm-like properties of microrobots that combine into aggregations that resemble swarms found in systems of biological agents[110,111]. Additional examples include the collective behaviours of algorithms both in the laboratory (in the Game of Life[112]) as well as in the wild (as seen in Wikipedia-editing bots[113]). Other examples include the emergence of novel algorithmic languages[114] between communicating intelligent machines as well as the dynamic properties of fully autonomous transportation systems. Ultimately, many interesting questions in this domain remain to be examined.

The vast majority of work on collective animal behaviour and collective robotics has focused on how interactions among simple agents can create higher-order structures and properties. Although important, this neglects that fact that many organisms, and increasingly also AI agents[75], are sophisticated entities with behaviours and interactions that may not be well-characterized by simplistic representations. Revealing what extra properties emerge when interacting entities are capable of sophisticated cognition remains a key challenge in the biological sciences and may have direct parallels in the study of machine behaviour. For example, similar to animals, machines may exhibit 'social learning'. Such social learning does not need be limited to machines learning from machines, but we may expect machines to learn from humans, and vice versa for humans to learn from the behaviour of machines. The feedback processes introduced may fundamentally alter the accumulation of knowledge, including across generations, directly affecting human and machine 'culture'.

In addition, human-made AI systems do not necessarily face the same constraints as do organisms, and collective assemblages of machines provide new capabilities, such as instant global communication, that can lead to entirely new collective behavioural patterns. Studies in collective machine behaviour examine the properties of assemblages of machines as well as the unexpected properties that can emerge from these complex systems of interactions.

For example, some of the most interesting collective behaviour of algorithms has been observed in financial trading environments. These environments operate on tiny time scales, such that algorithmic traders can respond to events and each other ahead of any human trader[115]. Under certain conditions, high-frequency capabilities can produce inefficiencies in financial markets[26,115]. In addition to the unprecedented response speed, the extensive use of machine learning, autonomous operation and ability to deploy at scale are all reasons to believe that the collective behaviour of machine trading may be qualitatively different than that of human traders. Furthermore, these financial algorithms and trading systems are necessarily trained on certain historic datasets and react to a limited variety of foreseen scenarios, leading to the question of how they will react to situations that are new and unforeseen in their design. Flash crashes are examples of clearly unintended consequences of (interacting) algorithms[116,117]; leading to the question of whether algorithms could interact to create a larger market crisis.

Hybrid human–machine behaviour

Humans increasingly interact with machines[16]. They mediate our social interactions[39], shape the news[14,17,55,56] and online information[15,118] that we see, and form relationships with us that can alter our social systems. Because of their complexity, these hybrid human–machine systems pose one of the most technically difficult yet simultaneously most important areas of study for machine behaviour.

Machines shape human behaviour

One of the most obvious—but nonetheless vital—domains of the study of machine behaviour concerns the ways in which the introduction of intelligent machines into social systems can alter human beliefs and behaviours. As in the introduction of automation to industrial processes[119], intelligent machines can create social problems in the process of improving existing problems. Numerous problems and questions arise during this process, such as whether the matching algorithms that are used for online dating alter the distributional outcomes of the dating process or whether news-filtering algorithms alter the distribution of public opinion. It is important to investigate whether small errors in algorithms or the data that they use could compound to produce society-wide effects and how intelligent robots in our schools, hospitals[120] and care centres might alter human development[121] and quality of life[54] and potentially affect outcomes for people with disabilities[122].

Other questions in this domain relate to the potential for machines to alter the social fabric in more fundamental ways. For example, questions include to what extent and what ways are governments using machine intelligence to alter the nature of democracy, political accountability and transparency, or civic participation. Other questions include to what degree intelligent machines influence policing, surveillance and warfare, as well as how large of an effect bots have had on the outcomes of elections[56] and whether AI systems that aid in the formation of human social relationships can enable collective action.

Notably, studies in this area also examine how humans perceive the use of machines as decision aids[8,123], human preferences for and against making use of algorithms[124], and the degree to which human-like machines produce or reduce discomfort in humans[39,125]. An important question in this area includes how humans respond to the increasing coproduction of economic goods and services in tandem with intelligent machines[126]. Ultimately, understanding how human systems can be altered by the introduction of intelligent machines into our lives is a vital component of the study of machine behaviour.

Humans shape machine behaviour

Intelligent machines can alter human behaviour, and humans also create, inform and mould the behaviours of intelligent machines. We shape machine behaviours through the direct engineering of AI systems and through the training of these systems on both active human input and passive observations of human behaviours through the data that we create daily. The choice of which algorithms to use, what feedback to provide to those algorithms[3,127] and on which data to train them are also, at present, human decisions and can directly alter machine behaviours. An important component in the study of machine behaviour is to understand how these engineering processes alter the resulting behaviours of AI, whether the training data are responsible for a particular behaviour of the machine, whether it is the algorithm itself or whether it is a combination of both algorithm and data. The framework outlined in Fig. 3 suggests that there will be complementary answers to the each of these questions. Examining how altering the parameters of the engineering process can alter the subsequent behaviours of intelligent machines as they interact with other machines and with humans in natural settings is central to a holistic understanding of machine behaviour.

Human–machine co-behaviour

Although it can be methodologically convenient to separate studies into the ways that humans shape machines and vice versa, most AI systems function in domains where they co-exist with humans in complex hybrid systems[39,67,125,128]. Questions of importance to the study of these systems include those that examine the behaviours that characterize human–machine interactions including cooperation, competition and coordination—for example, how human biases combine with AI to alter human emotions or beliefs[14,55,56,129,130], how human tendencies couple with algorithms to facilitate the spread of information[55], how traffic patterns can be altered in streets populated by large numbers of both driverless and human-driven cars and how trading patterns can be altered by interactions between humans and algorithmic trading agents[29] as well as which factors can facilitate trust and cooperation between humans and machines[88,131].

Another topic in this area relates to robotic and software-driven automation of human labour[132]. Here we see two different types of machine–human interactions. One is that machines can enhance a human's efficiency, such as in robotic- and computer-aided surgery. Another is that machines can replace humans, such as in driverless transportation and package delivery. This leads to questions about whether machines end up doing more of the replacing or the enhancing in the longer run and what human–machine co-behaviours will evolve as a result.

The above examples highlight that many of the questions that relate to hybrid human–machine behaviours must necessarily examine the feedback loops between human influence on machine behaviour and machine influence on human behaviour simultaneously. Scholars have begun to examine human–machine interactions in formal laboratory environments, observing that interactions with simple bots can increase human coordination[39] and that bots can cooperate directly with humans at levels that rival human–human cooperation[133]. However, there remains an urgent need to further understand feedback loops in natural settings, in which humans are increasingly using algorithms to make decisions[134] and subsequently informing the training of the same algorithms through those decisions. Furthermore, across all types of questions in the domain of machine behavioural ecology, there is a need for studies that examine longer-run dynamics of these hybrid systems[53] with particular emphasis on the ways that human social interactions[135,136] may be modified by the introduction of intelligent machines[137].

Outlook

Furthering the study of machine behaviour is critical to maximizing the potential benefits of AI for society. The consequential choices that we make regarding the integration of AI agents into human lives must be made with some understanding of the eventual societal implications of these choices. To provide this understanding and anticipation, we need a new interdisciplinary field of scientific study: machine behaviour.

For this field to succeed, there are a number of relevant considerations. First, studying machine behaviour does not imply that AI algorithms necessarily have independent agency nor does it imply algorithms should bear moral responsibility for their actions. If a dog bites someone, the dog's owner is held responsible. Nonetheless, it is useful to study the behavioural patterns of animals to predict such aberrant behaviour. Machines operate within a larger socio-technical fabric, and their human stakeholders are ultimately responsible for any harm their deployment might cause.

Second, some commentators might suggest that treating AI systems as agents occludes the focus on the underlying data that such AI systems are trained on. Indeed, no behaviour is ever fully separable from the environmental data on which that agent is trained or developed; machine behaviour is no exception. However, it is just as critical to understand how machine behaviours vary with altered environmental inputs as it is to understand how biological agents' behaviours vary depending on the environments in which they exist. As such, scholars of machine behaviour should focus on characterizing agent behaviour across diverse environments, much as behavioural scientists desire to characterize political behaviours across differing demographic and institutional contexts.

Third, machines exhibit behaviours that are fundamentally different from animals and humans, so we must avoid excessive anthropomorphism and zoomorphism. Even if borrowing existing behavioural scientific methods can prove useful for the study of machines, machines may exhibit forms of intelligence and behaviour that are qualitatively different—even alien—from those seen in biological agents.

Furthermore, AI scientists can dissect and modify AI systems more easily and more thoroughly than is the case for many living systems. Although parallels exist, the study of AI systems will necessarily differ from the study of living systems.

Fourth, the study of machine behaviour will require cross-disciplinary efforts[82,103] and will entail all of the challenges associated with such research[138,139]. Addressing these challenges is vital[140]. Universities and governmental funding agencies can play an important part in the design of large-scale, neutral and trusted cross-disciplinary studies[141].

Fifth, the study of machine behaviour will often require experimental intervention to study human–machine interactions in real-world settings[142,143]. These interventions could alter the overall behaviour of the system, possibly having adverse effects on normal users[144]. Ethical considerations such as these need careful oversight and standardized frameworks.

Finally, studying intelligent algorithmic or robotic systems can result in legal and ethical problems for researchers studying machine behaviour. Reverse-engineering algorithms may require violating the terms of service of some platforms; for example, in setting up fake personas or masking true identities. The creators or maintainers of the systems of interest could embroil researchers in legal challenges if the research damages the reputation of their platforms. Moreover, it remains unclear whether violating terms of service may expose researchers to civil or criminal penalties (for example, through the Computer Fraud and Abuse Act in the United States), which may further discourage this type of research[145].

Understanding the behaviours and properties of AI agents—and the effects they might have on human systems—is critical. Society can benefit tremendously from the efficiencies and improved decision-making that can come from these agents. At the same time, these benefits may falter without minimizing the potential pitfalls of the incorporation of AI agents into everyday human life.

Received: 20 September 2018; Accepted: 14 March 2019;
Published online 24 April 2019.

1. Simon, H. A. *The Sciences of the Artificial* (MIT Press, Cambridge, 1969).
Simon asks whether there can be a science of the 'artificial' that produces knowledge about artificial objects and phenomena.
2. Milner, R. A modal characterisation of observable machine-behaviour. In *Trees in Algebra and Programming, 6th Colloquium* 25–34 (Springer, 1981).
In this invited lecture, Robin Milner outlines the idea of studying machine behaviour using formal logic.
3. Thomaz, A. L. & Breazeal, C. Teachable robots: understanding human teaching behavior to build more effective robot learners. *Artif. Intell.* **172**, 716–737 (2008).
4. Stone, P. *et al. Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel* https://ai100.stanford.edu/2016-report (Stanford University, 2016).
5. O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Broadway Books, 2016).
This book articulates some of the risks posed by the uncritical use of algorithms in society and provides motivation for the study of machine behaviour.
6. Future of Life Institute. Autonomous weapons: an open letter from AI & robotics researchers. https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1 (2015).
7. Dressel, J. & Farid, H. The accuracy, fairness, and limits of predicting recidivism. *Sci. Adv.* **4**, eaao5580 (2018).
8. Binns, R. *et al.* 'It's reducing a human being to a percentage': perceptions of justice in algorithmic decisions. In *Proc. 2018 CHI Conference on Human Factors in Computing Systems* 377 (ACM, 2018).
9. Hudson, L., Owens, C. S. & Flannes, M. Drone warfare: blowback from the new American way of war. *Middle East Policy* **18**, 122–132 (2011).
10. Kahneman, D., Rosenfield, A. M., Gandhi, L. & Blaser, T. Noise: how to overcome the high, hidden cost of inconsistent decision making. *Harvard Business Review* https://hbr.org/2016/10/noise (2016).
11. Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. & Mullainathan, S. Human decisions and machine predictions. *Q. J. Econ.* **133**, 237–293 (2018).
12. Crawford, K. et al. *The AI Now report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-term*. https://ainowinstitute.org/AI_Now_2016_Report.pdf (2016).
13. Amodei, D. et al. Concrete problems in AI safety. Preprint at https://arxiv.org/abs/1606.06565 (2016).
14. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
15. Bessi, A. & Ferrara, E. Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday* **21**, 11 (2016).
16. Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. The rise of social bots. *Commun. ACM* **59**, 96–104 (2016).
17. Lazer, D. The rise of the social algorithm. *Science* **348**, 1090–1091 (2015).
18. Tufekci, Z. Engineering the public: big data, surveillance and computational politics. *First Monday* **19**, 7 (2014).
19. Lee, T.-S. & Chen, I.-F. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Syst. Appl.* **28**, 743–752 (2005).
20. Roszbach, K. Bank lending policy, credit scoring, and the survival of loans. *Rev. Econ. Stat.* **86**, 946–958 (2004).
21. Huang, C.-L., Chen, M.-C. & Wang, C.-J. Credit scoring with a data mining approach based on support vector machines. *Expert Syst. Appl.* **33**, 847–856 (2007).
22. Tsai, C.-F. & Wu, J.-W. Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Syst. Appl.* **34**, 2639–2649 (2008).
23. Chen, L. & Wilson, C. Observing algorithmic marketplaces in-the-wild. *SIGecom Exch.* **15**, 34–39 (2017).
24. Chen, L., Mislove, A. & Wilson, C. An empirical analysis of algorithmic pricing on Amazon marketplace. In *Proc. 25th International Conference on World Wide Web* 1339–1349 (International World Wide Web Conferences Steering Committee, 2016).
25. Hannák, A. et al. Bias in Online freelance marketplaces: evidence from TaskRabbit and Fiverr. In *Proc. ACM Conference on Computer Supported Cooperative Work and Social Computing* 1914–1933 (2017).
26. Cartlidge, J., Szostek, C., De Luca, M. & Cliff, D. Too fast too furious—faster financial-market trading agents can give less efficient markets. In *Proc. 4th International Conference on Agents and Artificial Intelligence* 126–135 (2012).
27. Kearns, M., Kulesza, A. & Nevmyvaka, Y. Empirical limitations on high-frequency trading profitability. *J. Trading* **5**, 50–62 (2010).
28. Wellman, M. P. & Rajan, U. Ethical issues for autonomous trading agents. *Minds Mach.* **27**, 609–624 (2017).
29. Farmer, J. D. & Skouras, S. An ecological perspective on the future of computer trading. *Quant. Finance* **13**, 325–346 (2013).
30. Perry, W. L., McInnis, B., Price, C. C., Smith, S. & Hollywood, J. S. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (RAND, 2013).
31. Bonnefon, J.-F., Shariff, A. & Rahwan, I. The social dilemma of autonomous vehicles. *Science* **352**, 1573–1576 (2016).
32. Kooti, F. et al. Analyzing Uber's ride-sharing economy. In *Proc. 26th International Conference on World Wide Web* 574–582 (International World Wide Web Conferences Steering Committee, 2017).
33. Zeng, X., Fapojuwo, A. O. & Davies, R. J. Design and performance evaluation of voice activated wireless home devices. *IEEE Trans. Consum. Electron.* **52**, 983–989 (2006).
34. Hendriks, B., Meerbeek, B., Boess, S., Pauws, S. & Sonneveld, M. Robot vacuum cleaner personality and behavior. *Int. J. Soc. Robot.* **3**, 187–195 (2011).
35. Hitsch, G. J., Hortaçsu, A. & Ariely, D. Matching and sorting in online dating. *Am. Econ. Rev.* **100**, 130–163 (2010).
36. Finkel, E. J., Eastwick, P. W., Karney, B. R., Reis, H. T. & Sprecher, S. Online dating: a critical analysis from the perspective of psychological science. *Psychol. Sci. Public Interest* **13**, 3–66 (2012).
37. Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G. & Breazeal, C. Growing growth mindset with a social robot peer. In *Proc. 2017 ACM/IEEE International Conference on Human–Robot Interaction* 137–145 (ACM, 2017).
38. Bemelmans, R., Gelderblom, G. J., Jonker, P. & de Witte, L. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *J. Am. Med. Dir. Assoc.* **13**, 114–120 (2012).
39. Shirado, H. & Christakis, N. A. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* **545**, 370–374 (2017).
In this human–machine hybrid study, the authors show that simple algorithms injected into human gameplay can improve coordination outcomes among humans.
40. Pichai, S. AI at Google: Our Principles. *Google Blog* https://blog.google/topics/ai/ai-principles/ (2018).
41. Roff, H. M. The strategic robot problem: lethal autonomous weapons in war. *J. Mil. Ethics* **13**, 211–227 (2014).
42. Krishnan, A. *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Routledge, 2016).
43. Voosen, P. The AI detectives. *Science* **357**, 22–27 (2017).
44. Szegedy, C. et al. Intriguing properties of neural networks. Preprint at https://arxiv.org/abs/1312.6199 (2013).
45. Zhang, Q.-S. & Zhu, S.-C. Visual interpretability for deep learning: a survey. *Front. Inf. Technol. Electronic Eng.* **19**, 27–39 (2018).
46. Doshi-Velez, F. & Kim, B. Towards a rigorous science of interpretable machine learning. Preprint at https://arxiv.org/abs/1702.08608 (2017).
47. Gebru, T. et al. Datasheets for datasets. Preprint at https://arxiv.org/abs/1803.09010 (2018).
48. Mitchell, M. et al. Model cards for model reporting. Preprint at https://arxiv.org/abs/1810.03993 (2018).
49. Lakkaraju, H., Kamar, E., Caruana, R. & Horvitz, E. Identifying unknown unknowns in the open world: representations and policies for guided exploration. In *Proc. 31st Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence* 2 (2017).
50. Johnson, N. et al. Abrupt rise of new machine ecology beyond human response time. *Sci. Rep.* **3**, 2627 (2013).
51. Appel, K., Haken, W. & Koch, J. Every planar map is four colorable. Part II: reducibility. *Illinois J. Math.* **21**, 491–567 (1977).

52. Appel, K. & Haken, W. Every planar map is four colorable. Part I: discharging. *Illinois J. Math*. **21**, 429–490 (1977).
53. Westlund, J. M. K., Park, H. W., Williams, R. & Breazeal, C. Measuring young children's long-term relationships with social robots. In *Proc. 17th ACM Conference on Interaction Design and Children* 207–218 (ACM, 2018).
54. Lorenz, T., Weiss, A. & Hirche, S. Synchrony and reciprocity: key mechanisms for social companion robots in therapy and care. *Int. J. Soc. Robot*. **8**, 125–143 (2016).
55. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
This study examines the complex hybrid ecology of bots and humans on Twitter and finds that humans spread false information at higher rates than bots.
56. Lazer, D. M. J. et al. The science of fake news. *Science* **359**, 1094–1096 (2018).
57. Roberts, M. E. *Censored: Distraction and Diversion Inside China's Great Firewall* (Princeton Univ. Press, 2018).
58. Corbett-Davies, S., Pierson, E., Feller, A., Goel, S. & Huq, A. Algorithmic decision making and the cost of fairness. In *Proc. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 797–806 (ACM, 2017).
59. Kleinberg, J., Mullainathan, S. & Raghavan, M. Inherent trade-offs in the fair determination of risk scores. Preprint at https://arxiv.org/abs/1609.05807 (2016).
60. Buolamwini, J. & Gebru, T. Gender shades: intersectional accuracy disparities in commercial gender classification. In *Proc. 1st Conference on Fairness, Accountability and Transparency* (eds Friedler, S. A. & Wilson, C.) **81**, 77–91 (PMLR, 2018).
61. Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V. & Kalai, A. T. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Proc. Advances in Neural Information Processing Systems* 4349–4357 (2016).
62. Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183–186 (2017).
63. Sweeney, L. Discrimination in online ad delivery. *Queueing Syst*. **11**, 10 (2013).
64. Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C. & Venkatasubramanian, S. Runaway feedback loops in predictive policing. Preprint at https://arxiv.org/abs/1706.09847 (2017).
65. Angwin, J., Larson, J., Mattu, S. & Kirchner, L. Machine bias. *ProPublica* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (2016).
66. Chouldechova, A., Benavides-Prado, D., Fialko, O. & Vaithianathan, R. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Proc. 1st Conference on Fairness, Accountability and Transparency* (eds Friedler, S. A. & Wilson, C.) **81**, 134–148 (PMLR, 2018).
67. Jennings, N. R. et al. Human–agent collectives. *Commun. ACM* **57**, 80–88 (2014).
68. Campbell, M., Hoane, A. J. & Hsu, F.-H. Deep blue. *Artif. Intell*. **134**, 57–83 (2002).
69. Schaeffer, J. et al. Checkers is solved. *Science* **317**, 1518–1522 (2007).
70. Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
71. Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
72. Bowling, M., Burch, N., Johanson, M. & Tammelin, O. Heads-up limit hold'em poker is solved. *Science* **347**, 145–149 (2015).
73. Bellemare, M. G., Naddaf, Y., Veness, J. & Bowling, M. The arcade learning environment: an evaluation platform for general agents. *J. Artif. Intell. Res*. **47**, 253–279 (2013).
74. Wellman, M. P. et al. Designing the market game for a trading agent competition. *IEEE Internet Comput*. **5**, 43–51 (2001).
75. Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I. & Osawa, E. RoboCup: the robot world cup initiative. In *Proc. 1st International Conference on Autonomous Agents* 340–347 (ACM, 1997).
76. Russakovsky, O. et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis*. **115**, 211–252 (2015).
77. Lin, T.-Y. et al. Microsoft COCO: common objects in context. In *Proc. European Conference on Computer Vision* (eds Fleet, D. et al.) 8693, 740–755 (Springer International Publishing, 2014).
78. Davis, J. & Goadrich, M. The relationship between precision–recall and ROC curves. In *Proc. 23rd International Conference on Machine Learning* 233–240 (ACM, 2006).
79. van de Sande, K. E. A., Gevers, T. & Snoek, C. G. M. Evaluating color descriptors for object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell*. **32**, 1582–1596 (2010).
80. Papineni, K., Roukos, S., Ward, T. & Zhu, W.-J. BLEU: a method for automatic evaluation of machine translation. In *Proc. 40th Annual Meeting on Association for Computational Linguistics* 311–318 (Association for Computational Linguistics, 2002).
81. Zhou, Z., Zhang, W. & Wang, J. Inception score, label smoothing, gradient vanishing and -log(D(x)) alternative. Preprint at https://arxiv.org/abs/1708.01729 (2017).
82. Epstein, Z. et al. Closing the AI knowledge gap. Preprint at https://arxiv.org/abs/1803.07233 (2018).
83. Tinbergen, N. On aims and methods of ethology. *Ethology* **20**, 410–433 (1963).
84. Nesse, R. M. Tinbergen's four questions, organized: a response to Bateson and Laland. *Trends Ecol. Evol*. **28**, 681–682 (2013).
85. Das, R., Hanson, J. E., Kephart, J. O. & Tesauro, G. Agent–human interactions in the continuous double auction. In *Proc. 17th International Joint Conference on Artificial Intelligence* 1169–1178 (Lawrence Erlbaum, 2001).
86. Deng, Y., Bao, F., Kong, Y., Ren, Z. & Dai, Q. Deep direct reinforcement learning for financial signal representation and trading. *IEEE Trans. Neural Netw. Learn. Syst*. **28**, 653–664 (2017).
87. Galceran, E., Cunningham, A. G., Eustice, R. M. & Olson, E. Multipolicy decision-making for autonomous driving via changepoint-based behavior prediction: theory and experiment. *Auton. Robots* **41**, 1367–1382 (2017).
88. Ribeiro, M. T., Singh, S. & Guestrin, C. Why should I trust you? Explaining the predictions of any classifier. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1135–1144 (ACM, 2016).
89. Smilkov, D., Thorat, N., Kim, B., Viégas, F. & Wattenberg, M. SmoothGrad: removing noise by adding noise. Preprint at https://arxiv.org/abs/1706.03825 (2017).
90. Nevmyvaka, Y., Feng, Y. & Kearns, M. reinforcement learning for optimized trade execution. In *Proc. 23rd International Conference on Machine Learning* 673–680 (ACM, 2006).
91. Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L. & Konstan, J. A. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proc. 23rd International Conference on World Wide Web* 677–686 (ACM, 2014).
92. Dalvi, N. & Domingos, P. Mausam, Sanghai, S. & Verma, D. Adversarial classification. In *Proc. Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 99–108 (ACM, 2004).
93. Globerson, A. & Roweis, S. Nightmare at test time: robust learning by feature deletion. In *Proc. 23rd International Conference on Machine Learning* 353–360 (ACM, 2006).
94. Biggio, B. et al. Evasion attacks against machine learning at test time. In *Proc. Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 387–402 (Springer, 2013).
95. Tramèr, F. et al. Ensemble adversarial training: attacks and defenses. Preprint at https://arxiv.org/abs/1705.07204 (2017).
96. Parkes, D. C. & Wellman, M. P. Economic reasoning and artificial intelligence. *Science* **349**, 267–272 (2015).
97. Wagner, A. *Robustness and Evolvability in Living Systems* (Princeton Univ. Press, 2013).
98. Edwards, H. & Storkey, A. Censoring representations with an adversary. Preprint at https://arxiv.org/abs/1511.05897 (2015).
99. Zemel, R., Wu, Y., Swersky, K., Pitassi, T. & Dwork, C. learning fair representations. *In Proc. International Conference on Machine Learning* 325–333 (2013).
100. Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C. & Venkatasubramanian, S. Certifying and removing disparate impact. In *Proc. 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 259–268 (ACM, 2015).
101. Cully, A., Clune, J., Tarapore, D. & Mouret, J.-B. Robots that can adapt like animals. *Nature* **521**, 503–507 (2015).
This study characterizes a robot driven by an adaptive algorithm that mimics the adaptation and behaviours of animals.
102. Bongard, J., Zykov, V. & Lipson, H. Resilient machines through continuous self-modeling. *Science* **314**, 1118–1121 (2006).
103. Leibo, J. Z. et al. Psychlab: a psychology laboratory for deep reinforcement learning agents. Preprint at https://arxiv.org/abs/1801.08116 (2018).
In this study, the authors use behavioural tools from the life sciences in the study of machine behaviours.
104. Subrahmanian, V. S. et al. The DARPA Twitter bot challenge. Preprint at https://arxiv.org/abs/1601.05140 (2016).
105. Carrascosa, J. M., Mikians, J., Cuevas, R., Erramilli, V. & Laoutaris, N. I. Always feel like somebody's watching me: measuring online behavioural advertising. In *Proc. 11th ACM Conference on Emerging Networking Experiments and Technologies* **13** (ACM, 2015).
106. Datta, A., Tschantz, M. C. & Datta, A. Automated Experiments on Ad Privacy Settings. *Proc. Privacy Enhancing Technologies* **2015**, 92–112 (2015).
107. Giusti, A. et al. A machine learning approach to visual perception of forest trails for mobile robots. *IEEE Robot. Autom. Lett*. **1**, 661–667 (2016).
108. Berdahl, A., Torney, C. J., Ioannou, C. C., Faria, J. J. & Couzin, I. D. Emergent sensing of complex environments by mobile animal groups. *Science* **339**, 574–576 (2013).
109. Couzin, I. D. et al. Uninformed individuals promote democratic consensus in animal groups. *Science* **334**, 1578–1580 (2011).
110. Rubenstein, M., Cornejo, A. & Nagpal, R. Programmable self-assembly in a thousand-robot swarm. *Science* **345**, 795–799 (2014).
111. Kernbach, S., Thenius, R., Kernbach, O. & Schmickl, T. Re-embodiment of honeybee aggregation behavior in an artificial micro-robotic system. *Adapt. Behav*. **17**, 237–259 (2009).
112. Bak, P., Chen, K. & Creutz, M. Self-organized criticality in the 'Game of Life'. *Nature* **342**, 780–782 (1989).
113. Tsvetkova, M., García-Gavilanes, R., Floridi, L. & Yasseri, T. Even good bots fight: the case of Wikipedia. *PLoS ONE* **12**, e0171774 (2017).
114. Lazaridou, A., Peysakhovich, A. & Baroni, M. Multi-agent cooperation and the emergence of (natural) language. Preprint at https://arxiv.org/abs/1612.07182 (2016).
115. Budish, E., Cramton, P. & Shim, J. The high-frequency trading arms race: frequent batch auctions as a market design response. *Q. J. Econ*. **130**, 1547–1621 (2015).
116. Kirilenko, A. A. & Lo, A. W. Moore's law versus Murphy's law: algorithmic trading and its discontents. *J. Econ. Perspect*. **27**, 51–72 (2013).

117. Menkveld, A. J. The economics of high-frequency trading: taking stock. *Annu. Rev. Financ. Econ.* **8**, 1–24 (2016).
118. Mønsted, B., Sapieżyński, P., Ferrara, E. & Lehmann, S. Evidence of complex contagion of information in social media: an experiment using Twitter bots. *PLoS ONE* **12**, e0184148 (2017).
This study presents an experimental intervention on Twitter using bots and provides evidence that information diffusion is most accurately described by complex contagion.
119. Bainbridge, L. Ironies of automation. *Automatica* **19**, 775–779 (1983).
120. Jeong, S., Breazeal, C., Logan, D. & Weinstock, P. Huggable: the impact of embodiment on promoting socio-emotional interactions for young pediatric inpatients. In *Proc. 2018 CHI Conference on Human Factors in Computing Systems* 495 (ACM, 2018).
121. Kory Westlund, J. M. et al. Flat vs. expressive storytelling: young children's learning and retention of a social robot's narrative. *Front. Hum. Neurosci*. **11**, 295 (2017).
122. Salisbury, E., Kamar, E. & Morris, M. R. Toward scalable social alt text: conversational crowdsourcing as a tool for refining vision-to-language technology for the blind. *Proc. 5th AAAI Conference on Human Computation and Crowdsourcing* (2017).
123. Awad, E. et al. The Moral Machine experiment. *Nature* **563**, 59–64 (2018).
124. Dietvorst, B. J., Simmons, J. P. & Massey, C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen*. **144**, 114–126 (2015).
125. Gray, K. & Wegner, D. M. Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* **125**, 125–130 (2012).
126. Brynjolfsson, E. & Mitchell, T. What can machine learning do? Workforce implications. *Science* **358**, 1530–1534 (2017).
127. Christiano, P. F. et al. Deep reinforcement learning from human preferences. In *Proc. Advances in Neural Information Processing Systems* 30 (eds Guyon, I. et al.) 4299–4307 (Curran Associates, 2017).
128. Tsvetkova, M. et al. Understanding human–machine networks: a cross-disciplinary survey. *ACM Comput. Surv*. **50**, 12:1–12:35 (2017).
129. Hilbert, M., Ahmed, S., Cho, J., Liu, B. & Luu, J. Communicating with algorithms: a transfer entropy analysis of emotions-based escapes from online echo chambers. *Commun. Methods Meas*. **12**, 260–275 (2018).
130. Kramer, A. D. I., Guillory, J. E. & Hancock, J. T. Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl Acad. Sci. USA* **111**, 8788–8790 (2014).
131. Kamar, E., Hacker, S. & Horvitz, E. Combining human and machine intelligence in large-scale crowdsourcing. in *Proc. 11th International Conference on Autonomous Agents and Multiagent Systems* 467–474 (International Foundation for Autonomous Agents and Multiagent Systems, 2012).
132. Jackson, M. *The Human Network: How Your Social Position Determines Your Power, Beliefs, and Behaviors* (Knopf Doubleday, 2019).
133. Crandall, J. W. et al. Cooperating with machines. *Nat. Commun*. **9**, 233 (2018).
This study examines algorithmic cooperation with humans and provides an example of methods that can be used to study the behaviour of human–machine hybrid systems.
134. Wang, D., Khosla, A., Gargeya, R., Irshad, H. & Beck, A. H. Deep learning for identifying metastatic breast cancer. Preprint at https://arxiv.org/abs/1606.05718 (2016).
135. Pentland, A. *Social Physics: How Social Networks Can Make Us Smarter* (Penguin, 2015).
136. Lazer, D. et al. Computational social science. *Science* **323**, 721–723 (2009).
137. Aharony, N., Pan, W., Ip, C., Khayal, I. & Pentland, A. Social fMRI: investigating and shaping social mechanisms in the real world. *Pervasive Mobile Comput*. **7**, 643–659 (2011).
138. Ledford, H. How to solve the world's biggest problems. *Nature* **525**, 308–311 (2015).
139. Bromham, L., Dinnage, R. & Hua, X. Interdisciplinary research has consistently lower funding success. *Nature* **534**, 684–687 (2016).
140. Kleinberg, J. & Oren, S. Mechanisms for (mis)allocating scientific credit. In *Proc. 43rd Annual ACM Symposium on Theory of Computing* 529–538 (ACM, 2011).
141. Kannel, W. B. & McGee, D. L. Diabetes and cardiovascular disease. The Framingham study. *J. Am. Med. Assoc*. **241**, 2035–2038 (1979).
142. Krafft, P. M., Macy, M. & Pentland, A. Bots as virtual confederates: design and ethics. In *Proc. 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* 183–190 (ACM, 2017).
143. Meyer, M. N. Two cheers for corporate experimentation: The A/B illusion and the virtues of data-driven innovation. *Colorado Technol. Law J*. **13**, 273 (2015).
144. Xing, X. et al. Take this personally: pollution attacks on personalized services. In *Proc. 22nd USENIX Security Symposium* 671–686 (2013).
145. Patel, K. Testing the limits of the First Amendment: how a CFAA prohibition on online antidiscrimination testing infringes on protected speech activity. *Columbia Law Rev*. https://doi.org/10.2139/ssrn.3046847 (2017).

Acknowledgements I.R. received funding from the Ethics & Governance of Artificial Intelligence Fund; J.B. received funding from NSF awards INSPIRE-1344227 and BIGDATA-1447634, DARPA's Lifelong Learning Machines program, ARO contract W911NF-16-1-0304; J.-F.B. from the ANR-Labex IAST; N.A.C. a Pioneer Grant from the Robert Wood Johnson Foundation; I.D.C. received funding from the NSF (IOS-1355061), the ONR (N00014-09-1-1074 and N00014-14-1-0635), the ARO (W911NG-11-1-0385 and W911NF14-1-0431), the Struktur- und Innovationsfunds für die Forschung of the State of Baden-Württemberg, the Max Planck Society and the DFG Centre of Excellence 2117 "Centre for the Advanced Study of Collective Behaviour" (422037984); D.L. received funding from the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) under contract 2017-17061500006; J.B.T. received funding from the Center for Brains, Minds and Machines (CBMM) under NSF STC award CCF–1231216; M.W. received funding from the Future of Life Institute. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA or the US Government.

Reviewer information *Nature* thanks Thomas Dietterich, Maria Gini, Wendell Wallach and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions I.R., M.C. and N.O. conceived the idea, produced the figures and drafted the manuscript. All authors contributed content, and refined and edited the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to I.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

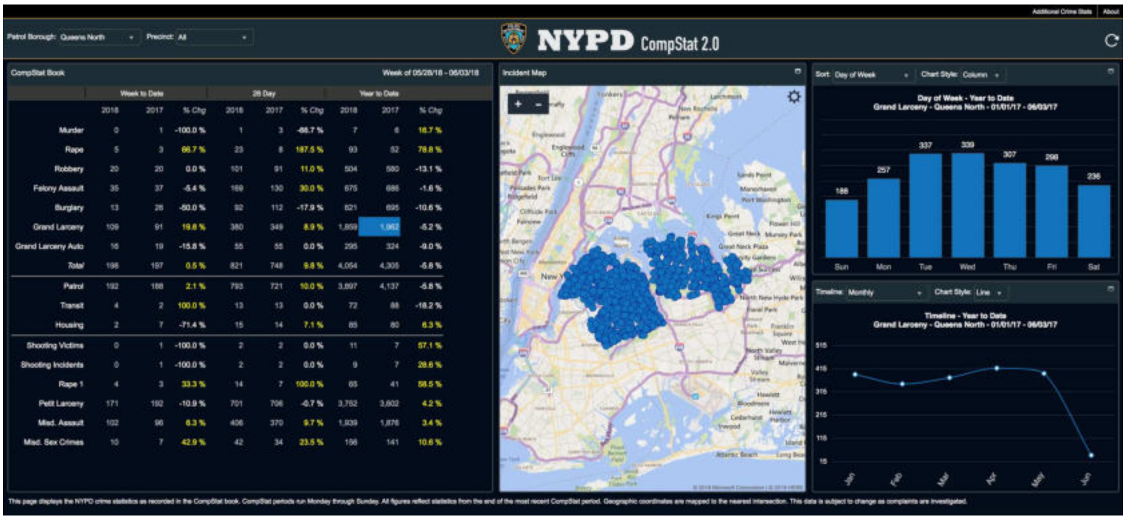
URBAN OMNIBUS

The Location of Justice: Systems

Policing Is an Information Business

Ingrid Burrington

Jun 20, 2018



Patrol Borough: Queens North
Precinct: All

Sort: Day of Week
Chart Style: Column

Timeline: Monthly
Chart Style: Line

CompStat Book — Week of 05/28/18 - 06/03/18

	Week to Date			28 Day			Year to Date		
	2018	2017	% Chg	2018	2017	% Chg	2018	2017	% Chg
Murder	0	1	-100.0 %	1	3	-66.7 %	7	6	16.7 %
Rape	5	3	66.7 %	23	8	187.5 %	93	52	78.8 %
Robbery	20	20	0.0 %	101	91	11.0 %	504	580	-13.1 %
Felony Assault	35	37	-5.4 %	169	130	30.0 %	675	686	-1.6 %
Burglary	13	26	-50.0 %	92	112	-17.9 %	621	695	-10.6 %
Grand Larceny	109	91	19.8 %	380	349	8.9 %	1,859	1,962	-5.2 %
Grand Larceny Auto	16	19	-15.8 %	55	55	0.0 %	295	324	-9.0 %
Total	198	197	0.5 %	821	748	9.8 %	4,054	4,305	-5.8 %
Patrol	192	188	2.1 %	793	721	10.0 %	3,897	4,137	-5.8 %
Transit	4	2	100.0 %	13	13	0.0 %	72	88	-18.2 %
Housing	2	7	-71.4 %	15	14	7.1 %	85	80	6.3 %
Shooting Victims	0	1	-100.0 %	2	2	0.0 %	11	7	57.1 %
Shooting Incidents	0	1	-100.0 %	2	2	0.0 %	9	7	28.6 %
Rape 1	4	3	33.3 %	14	7	100.0 %	65	41	58.5 %
Petit Larceny	171	192	-10.9 %	701	706	-0.7 %	3,752	3,602	4.2 %
Misd. Assault	102	96	6.3 %	406	370	9.7 %	1,939	1,876	3.4 %
Misd. Sex Crimes	10	7	42.9 %	42	34	23.5 %	156	141	10.6 %

This page displays the NYPD crime statistics as recorded in the CompStat book. CompStat periods run Monday through Sunday. All figures reflect statistics from the end of the most recent CompStat period. Geographic coordinates are mapped to the nearest intersection. This data is subject to change as complaints are investigated.

CompStat 2.0, the latest iteration of the New York City Police Department CompStat model, allows a large proportion of crime data to be made available to the public "through an online interactive experience." Screenshot via NYPD Online

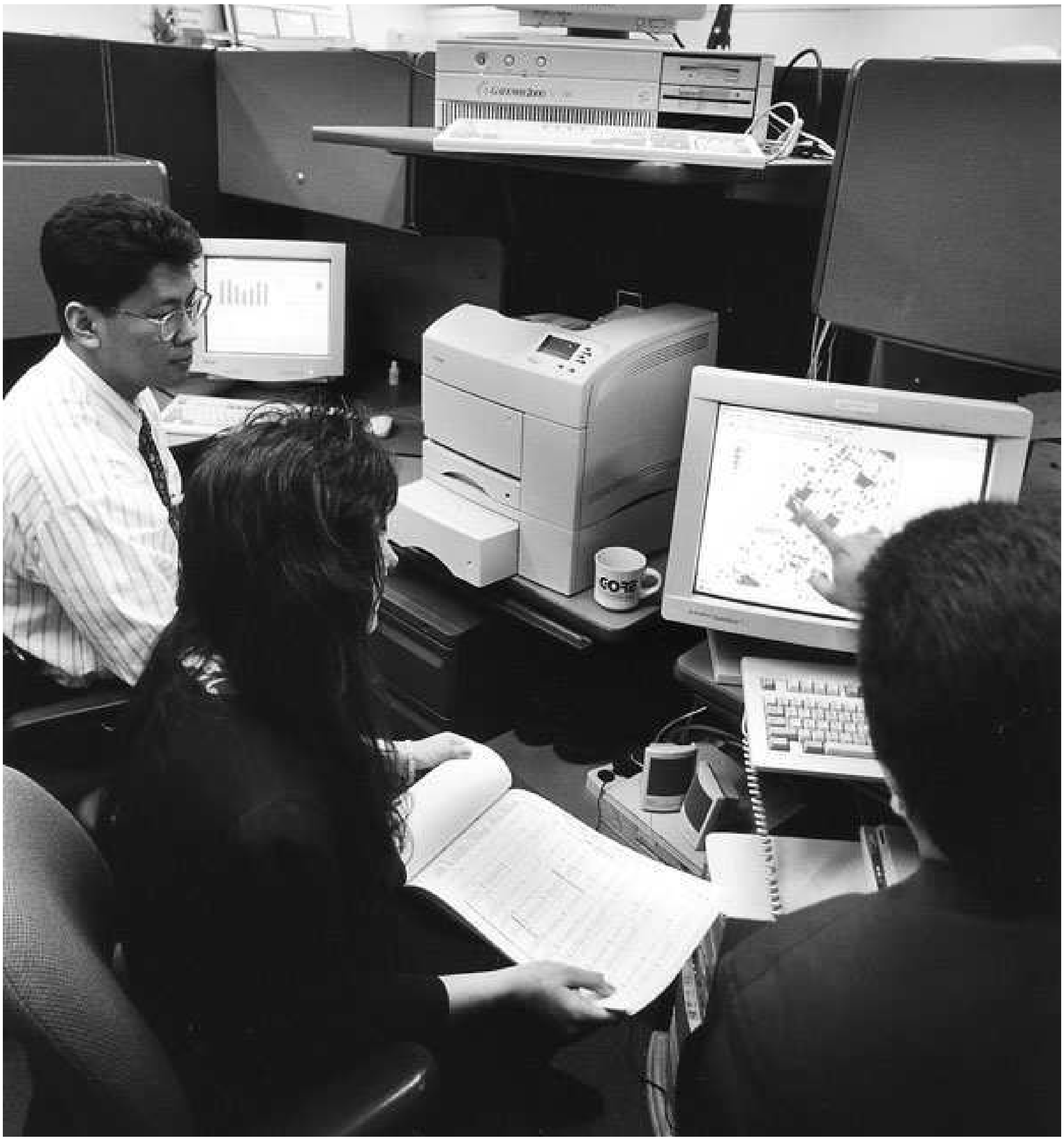
A multi-billion-dollar industry of data-driven policing technology includes dozens of mapping, surveillance, and data-analysis tools, each claiming to hone in on crime at ever-finer grain. But as data and infrastructure writer **Ingrid Burrington** argues here, these technologies represent less a science for the provision of safety, and more a highly effective sales pitch for a management model born in the zero tolerance Giuliani era. Since the early 1990s, when precinct commanders pushed pins into paper maps, police in New York have contended that if they can track crime, then they can predict it, and therefore prevent it. The maps they've made have monopolized media narratives, and shaped the lives of those who live within their frames. The authors of CompStat now export their methods around the world. (Burrington maps the "CompStat evangelist consultant world tour" here.) Business is good for those technology vendors and consultants who sell crime- and fear-reduction as a customer service. (Burrington also takes stock of the available tools in an illustrated, non-exhaustive taxonomy here.) But who's buying? More than a set of tools, crime mappers hawk a model of a future world where the cost of guaranteed order would be accountability to the public.

Policing and urban planning have a lot in common. Both cops and planners' ostensible goal is to make the city a more livable place, though this goal is constantly haunted by a question: Livable for whom? Both transform a public's experience of a city, generally by imposing and enforcing rules and systems that change how people move through space. In the United States, public understanding of both professions is to some extent influenced by romanticized media narratives which heavily emphasize cities like Los Angeles and New York. Both sectors have a particularly heavy fetish for maps and data as mechanisms for understanding and shaping cities, a fetish that has intensified in the past few decades thanks to advances in technology.

Where the two professions diverge starkly is in matters of time and violence. Where urban planning might be considered a slower, bureaucratic, deliberative process, policing is expected to engage with and respond to city conditions and events in real time — or, increasingly, *ahead* of time. And unlike urban planners, cops are permitted to respond with firearms and Tasers.

That being said, planning is fully capable of enacting slower, more systemic acts of violence onto a city, and like policing, such violence can be enabled and plausibly denied by sufficiently complex data and maps. Where the urban planner has eminent domain and urban renewal, the police officer has crime hotspots and risk terrain modeling. Where a planner might control a city through highway design and traffic flows, a police department's automated license plate readers or mobile cell site simulators render public movement into potential patterns of criminal behavior.

Police departments often frame these methods of spatial analysis, data collection, and networked surveillance as not only necessary, but relatively benign tools that departments have been using for years. "Policing has always been an information business," notes a 2015 NYPD document on information technology programs. This observation comes toward the end of an outline of exciting new developments in the department's use of tech to fight crime. These initiatives vary from advances that seem quaint (giving every NYPD officer an email address!), to heavy infrastructure development, to extensive additions of sensors and surveillance tools throughout the city.



Matching paper statistics to computerized maps in an early version of the CompStat program. Photo courtesy of Harvard Kennedy School, Ash Center for Democratic Governance and Innovation

Of course, as tremendous instruments of power and violence, maps have been used by police (agents of the former, authorized to hold a monopoly on the latter) for decades. But in the 1990s, the emergence of desktop GIS software for and in police departments dramatically increased the data collection and storage capacities of that “information business.” The technology’s adoption coincided with the era of NYPD Commissioner Bill Bratton and his avuncular lieutenant Jack Maple. This is where many histories tend to pinpoint the transformational moment for crime mapping: Bratton and Maple tracking turnstile jumpers in the New York City subway system, Maple outlining a four-point theory of policing management on a napkin at Elaine’s restaurant (“Accurate, timely intelligence; rapid deployment, effective tactics; relentless follow-up and assessment”), New York’s crime rate precipitously falling thanks to the data-driven innovations of CompStat.

This legend would prove to be both the downfall and redemption of Bratton, Maple, and their colleague John Timoney. While media adulation of Bratton's NYPD infuriated Mayor Rudolph Guiliani to the point that he forced Bratton out a mere two years into his lauded transformation of the city, it would become the calling card used by all three to establish public legacies as experts in the science of crime-fighting.

This particular version of data-driven policing history, starring Bratton and his colleagues, uses its larger-than-life characters and conveniently selective statistics to obfuscate the public harms created by introducing new data-driven technologies and policies into policing, and the ambiguity about whom these tactics are supposed to benefit. The political intrigues of City Hall and One Police Plaza, the personalities (and egos) in Bratton and Guiliani's inner circles, and the precipitous drops in crime during Bratton's tenure (New York City's murders dropped almost by half, from 1,951 in 1993 to 983 in 1996) provide ample distraction from the technology itself, the fact that no one seems able to decide what the word "CompStat" actually means, and the ongoing debates over whether the tactics Maple apocryphally outlined on a napkin were in fact responsible for a drop in crime. Since 9/11, other technology and surveillance tools have benefitted from a similar tendency to foreground mythology over evidence of impact and threats to civil liberties.

As data-driven policing has gone from novel tactic to entrenched strategy, maps have helped legitimize and (literally and figuratively) ground mythologized versions of cities. To understand the spatial history of modern networked surveillance and policing, one could do worse than to look at the cartographic and rhetorical maps used and created by the NYPD over the past few decades, starting with its own founding mythologies of modern crime mapping.

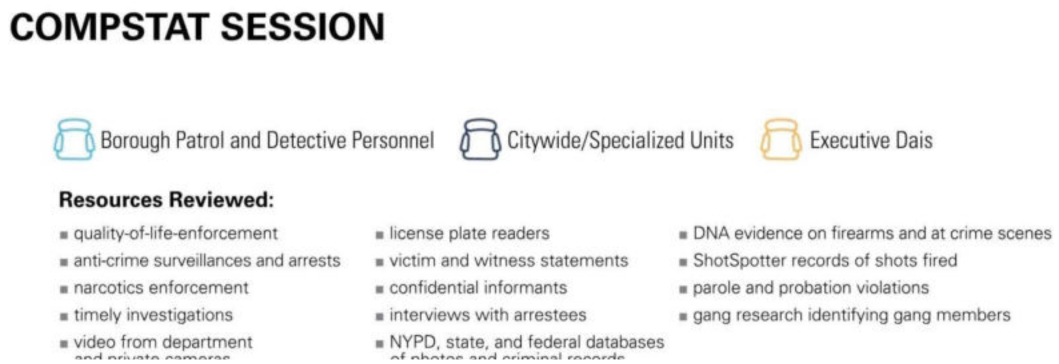
An early departmental CompStat report-back at 1 Police Plaza. Photo courtesy of Harvard Kennedy School, Ash Center for Democratic Governance and Innovation

1.

The first CompStat maps were made with pins, paper, and transparent acetate. The NYPD technically didn't have the budget to support their cost, so the New York City Police Foundation provided a $10,000 donation. Although the department would eventually switch to computerized maps, displayed on eight foot-by-eight foot screens in One Police Plaza, the image of police officers fumbling with pushpins and acetate film they could barely afford suggests a surprisingly scrappy origin story for a management strategy so often associated with precision and technical expertise — even if its own name is both vague and technically meaningless.

None of CompStat's historians can decide if it is shorthand for "computational" or "comparative" statistics, nor do any of them seem to think that etymology matters. It's more often described as management strategy than technology innovation, and in New York perhaps its greatest legacy was as armature for political theater.

In 1994, CompStat publicly manifested primarily as a twice-weekly meeting in which the highest-level figures of NYPD management grilled precinct commanders over the minute details of their local crime numbers. The meetings, held at One Police Plaza, developed a reputation thanks to the frequently childish bullying tendencies of NYPD leadership. Name-calling and chair-throwing were regular occurrences; in one frequently-cited incident at the time, during a presentation by Brooklyn South borough commander Tony Simonetti an illustration of Pinocchio was displayed on the eight-by-eight screens to imply he was lying about crime-fighting efforts under his jurisdiction.

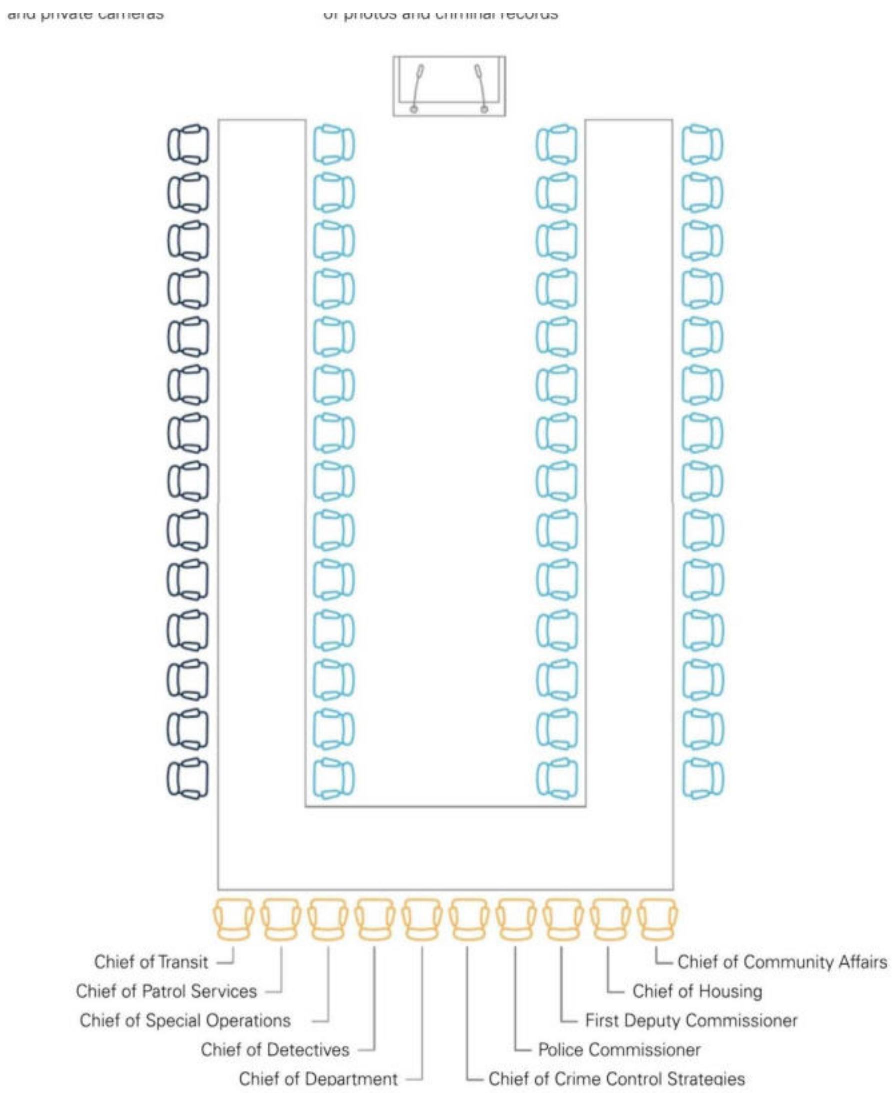


COMPSTAT SESSION

Borough Patrol and Detective Personnel
Citywide/Specialized Units
Executive Dais

Resources Reviewed:

- quality-of-life-enforcement
- anti-crime surveillances and arrests
- narcotics enforcement
- timely investigations
- video from department
- license plate readers
- victim and witness statements
- confidential informants
- interviews with arrestees
- NYPD, state, and federal databases
- DNA evidence on firearms and at crime scenes
- ShotSpotter records of shots fired
- parole and probation violations
- gang research identifying gang members

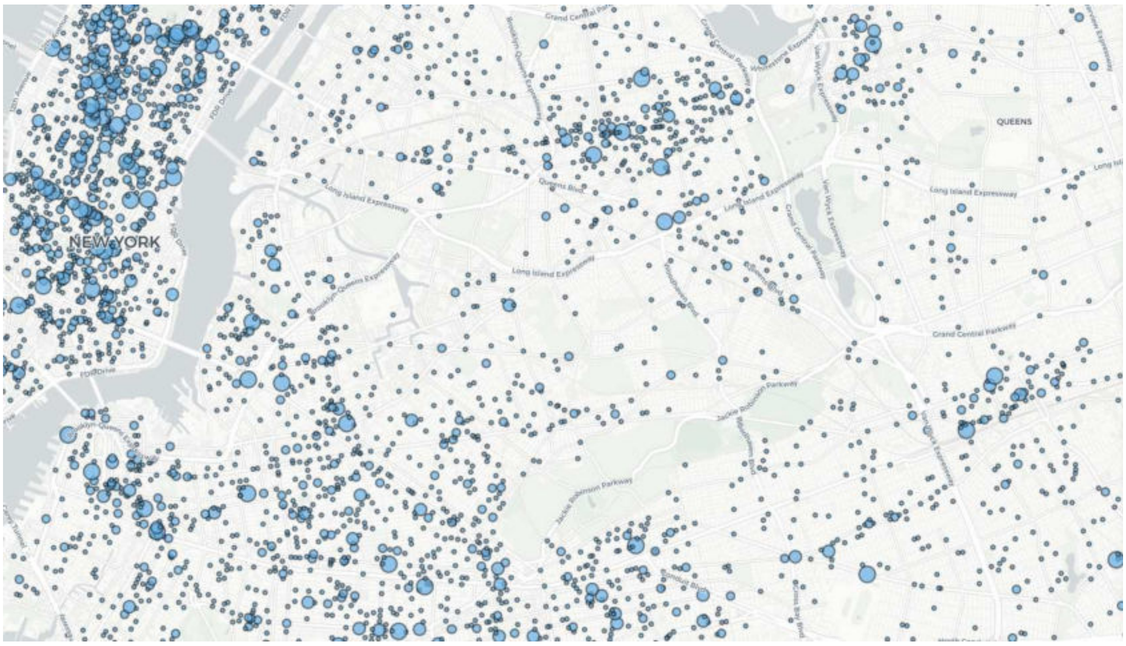

and private cameras
of photos and criminal records
Chief of Transit
Chief of Patrol Services
Chief of Special Operations
Chief of Detectives
Chief of Department
Chief of Community Affairs
Chief of Housing
First Deputy Commissioner
Police Commissioner
Chief of Crime Control Strategies

In a CompStat meeting, captains face Police Department executives as well as representatives from specialized units and borough patrols, and must answer for reported crime rate trends in their precincts. Image via NYPD Police Commissioner's Report, January 2018

The CompStat meeting format bears more relationship to a corporate board presentation with VPs delivering quarterly sales numbers, a fact noted early on by reporters. A 1994 *New York Times* article headlined "Boardroom Tactics Utilized in the War on Crime" noted that Bratton often utilized "corporate metaphors." The *Times* reported, "'We have a lot to learn from the private sector,' he said. 'We're looking at the customer — the public — to see what his needs are. We're looking at the product. Does it meet the customer's needs. If not, we're going to change the product and change the entity that creates the product. The profit I'm looking for is reduced crime, reduced fear.'"

The technical foundation of CompStat also emerged from retrofitting private sector tools, including that aforementioned anodyne name. One version of CompStat's history attributes its origins to the department's ancient IBM floppy disks, which had such limited storage space that filenames could only be eight characters. "Compstat" was a placeholder name chosen on a whim and in a rush for a rudimentary database made in Informix's SmartWare, an off-the-shelf database product marketed to small business owners.

It makes sense that most of the technology deployed by the NYPD in the 1990s amounted to hacks on products not designed for police departments. Off-the-shelf commercial software for police departments didn't exist. The emergent market for off-the-shelf GIS software (the NYPD happened to use MapInfo) emphasized business applications because that's who would pay for software. But as Bratton and Maple's NYPD aggressively promoted their data-driven approach, companies and vendors emerged to bring even more private-sector expertise to policing. Today, CompStat itself has become a mutable corporate product deployed across the public sector. Aside from other police departments adopting it, we see Mayor Bill de Blasio calling for a CompStat for the MTA, a CompStat for public claims against the city, and a "CompStat for Prescription Drug Abuse."



Locations of reported violent crime in May 2018. Screenshot via NYC Crime Map

And, like most private sector initiatives, CompStat's greatest victory may manifest more in a bottom-line set of numbers than the realities of any citizen's quality of life. In a 2010 survey of 491 retired NYPD officers conducted by criminologists Eli Silverman and John Eterno (also a retired NYPD captain who worked directly on innovative mapping projects in the department), dozens of officers expressed deep skepticism about the accuracy of CompStat numbers and described internal pressure from superiors to manipulate crime numbers, a practice well-documented by former officer Adrian Schoolcraft while working in Bedford-Stuyvesant's 81st Precinct around the same time. The department and NYPD commissioner balked at the survey results and Schoolcraft's exposé, insisting that measures for maintaining accurate statistics were strictly enforced. (In 2013, those practices would become a liability when the department was sued over its controversial stop-and-frisk policies.)

2.

While they rose to prominence around the same time, CompStat is not exactly the same as “broken windows” policing, Bill Bratton’s other beloved innovation. There’s nothing about the collection and sharing of increasingly granular data about cities and crime that inherently requires an increasingly granular focus on vandalism, public urination, or other so-called “quality of life” issues, but the pursuit of metrics and of minor offenses proved symbiotic. More attention to previously overlooked “quality of life” arrests meant more arrests to keep track of, which meant creating more data, which meant creating workflows for managing that data.

Broken windows theory is also part of what made the NYPD’s data-driven strategy so inherently spatial, and its tangible, spatial returns are what make it such an appealing media narrative. The systemic origins of poverty, crime, and civil unrest are big, tangled, and hard to locate in the immediate here-and-now of a city street corner or a subway platform. A policing approach focused on the corner, the subway platform, and things that a public can see (and, in the case of the media, photograph and file on deadline) at least creates the appearance that those systemic problems have been resolved.

As the story goes, broken windows theory first coevolved with an early version of the CompStat methodology during Bratton's tenure as the head of the New York City Transit Police; Maple was his deputy. NYPD officers in the subway, 1992. Photo by Winston Vargas via Flickr

But, as critics point out, quality of life policing tends to prosecute the victims of systemic harms, not its perpetrators. Rather than asking what economic conditions led a teenager to jump a turnstile, it assumes that removing the turnstile-jumper will itself solve the problem of criminal activity on the subway. It also assumes a facile theory of change, ignoring the myriad of other data points that contribute to the rise and fall of crime (from the economy, to the weather, to rates of drug use, to access to social services).

Beyond the minutiae of the pushpin crime map, broken windows serves other cartographic interventions, remapping crisis away from systemic paradigms of City Hall, Wall Street, and Washington onto Midtown South, Crown Heights, and subway turnstiles. The measure of public order becomes a matter of metrics — the number of quality-of-life arrests, the meeting of expectations set in CompStat meetings.

Philadelphia, PA (1997)
Bratton

Bratton was hired to make recommendations to improve the Philadelphia Police Department's crime-fighting strategies and brought Timoney along for the job. Due in part to Bratton's heavy hinting that Timoney might want the job, Timoney was hired as Police Commissioner that same year. Click here to track the networks of consultancy through which CompStat's architects spread the gospel — often for a tidy fee.

3.

Eager to sell their services as paradigm shift, the architects of data-driven policing promoted the idea that the history of the strategy all began with New York. Few departments so aggressively positioned themselves at the forefront of technical innovation quite like the NYPD, and few modern police chiefs have had as outsized a public persona as Bill Bratton.

Bratton had a vested interest in CompStat becoming a nationally recognized model for police management — not only as a matter of personal pride after leaving the department, but also as a matter of professional ambition. Bratton and other key figures of his leadership team (Maple, Timoney, Louis Anemone, John Linder, and Robert Wasserman, among others) built lucrative careers in police department consulting after their success at the NYPD. Both Bratton and Timoney would move back and forth between consulting and working as police chiefs (in LA and New York again for Bratton; Philadelphia and Miami for Timoney) following their departure in 1996. Any map of the CompStat evangelist consultant world tour is doomed to be incomplete; documentation of these jobs mostly exists in magazine profiles, regional news mentions (usually up in arms about cities paying the consultant's exceptionally high fees), or as case studies in Bratton, Maple, and Timoney's respective memoirs.

Those books are light on specific IT product recommendations. They do offer some insight into what the three men consider policing best practices (information sharing is good, as is basically anything they decided to do), what the media and politicians get wrong about stopping crime (usually, everything), and the psychology of the national-stage supercop (self-deprecating working-class humor covers for entitled, defensive aggrievement at the public's apparent mistrust). All three insist in their personal definitions of CompStat that it is, above all else, an instrument of police accountability — albeit accountability by way of a trickle-down logic. Making precinct commanders more accountable to top brass meant beat cops were more accountable to precinct commanders, which meant beat cops were more accountable to citizens. It was at times a blunt instrument, but as Bratton observed, "CompStat was police Darwinism; the fittest survived and thrived."

Sea changes in policing and technology that had little to do with the NYPD or its three superstars typically remain absent from these narratives. There wasn't anything uniquely "New York" about using data or GIS in the 1990s; nor was there anything uniquely "New York" to broken windows policing. The idea of making maps or using statistics to monitor crime patterns is an old one, and in terms of a legacy of technological innovations, Southern California could arguably give the NYPD a run for its money. The LAPD pioneered helicopter-based aerial surveillance in the 1960s, and in the late ‘80s, San Diego became the first police department to use CalGang, the statewide gang database tool whose model (and controversial application, most recently used in enabling ICE deportations) has been replicated by cities across the country.

A helicopter of the Los Angeles Police Department Air Support Division, which provides air support to patrol and specialized units of the Los Angeles Police Department. Photo via Wikimedia Commons

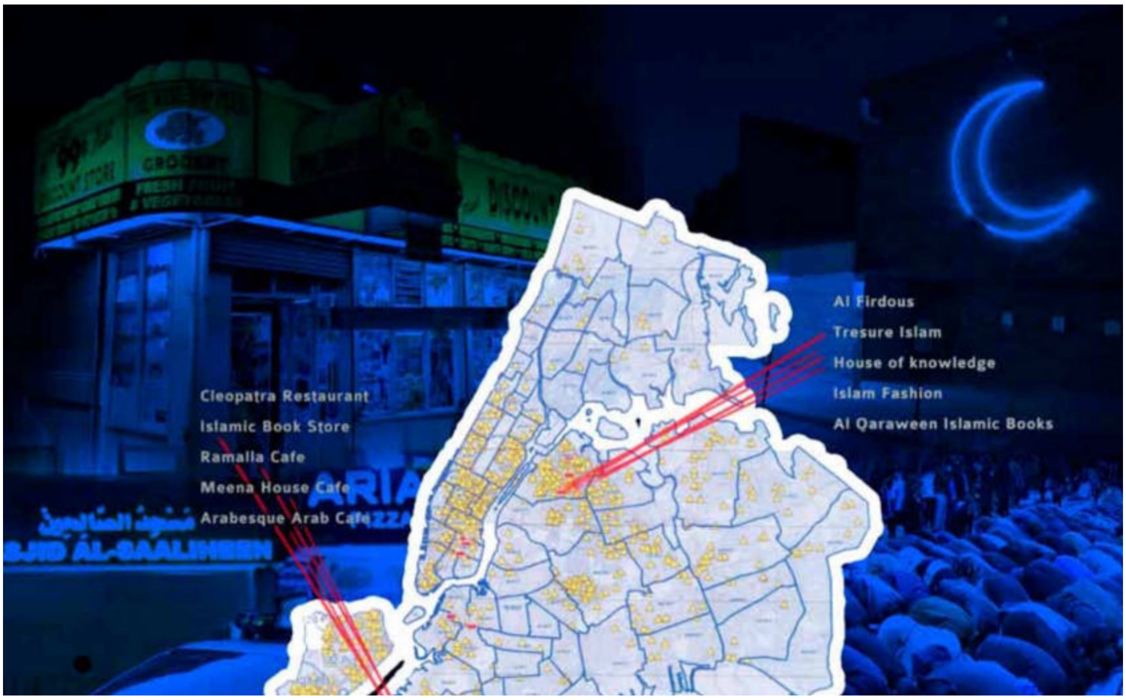
Many of these innovations drew on newly available federal resources. The 1994 Violent Crime Control Act (better known as “the Clinton crime bill”) explicitly allocated federal funding to support the hiring and training of more police officers in cities through the Community Oriented Policing Services (COPS) office. COPS and the Department of Justice also supported research into the use of emerging technology to support policing. A 1999 report on crime mapping from the Department of Justice notes a number of federal partnerships and funding opportunities for expanding “data-driven management” in police departments.

The Department of Justice’s 1999 crime mapping report echoes another point from Bratton, Timoney, and Maple: technological interventions are only meaningful with excellent leadership and community engagement. The crime map is made meaningful by its cartographers, but also by departmental navigators who deploy officers based on these maps. In Bratton and Timoney’s cases, this apparently meant misclassifying crimes to produce the appearance of a declining crime rate and deploying severely disproportionate force against protestors at major demonstrations. By positioning themselves as thought leaders in policing innovation, Bratton and his NYPD leadership set much of the tone for what would become commonplace practices in contemporary data-driven policing: a high-level faith in metrics as a tool for accountability, a pursuit of a vaguely defined and subjective standard for “quality of life,” and an insistence that the media and government always, always didn’t understand.

4.

Of course, another major event was about to transform modern American policing after Bratton’s heyday. 9/11 introduced new counterterrorism mandates and new federal funding resources to American cities, and the NYPD starred in public narratives of post-9/11 policing innovation, in which counterterrorism served as a pretext for increased surveillance, mapping, and data collection — particularly on Muslim populations, but effectively on dozens of others. With the suspension of traditional legal oversight over surveillance, the NYPD Intelligence Bureau expanded the geography of threats to public disorder beyond the broken window and inside the perfectly-maintained façades of mosques, restaurants, and internet cafés in predominantly Muslim communities.

That geography fell primarily to the purview of the Demographics Unit, which employed a mix of street-level surveillance and undercover work with mapping and analysis of publicly available data. Documents of the Intelligence Division’s activities, leaked to the Associated Press reporters Adam Goldman and Matt Apuzzo in 2012, describe the Demographics Unit as a 16-member team focused on “[identifying and mapping] ethnic residential concentrations within the Tri-State area.”



A 2013 report by The Muslim American Civil Liberties Coalition (MACLC), The Asian American Legal Defense and Education Fund (AALDEF), and The Creating Law Enforcement Accountability and Responsibility (CLEAR) Project, CUNY School of Law, illustrated NYPD practices of surveilling Muslim civilians. Image via CUNY School of Law

Among the documents released by the AP are a series of Demographics Unit reports on various ethnic enclaves (described as “Locations of Concern”) and potential terrorist hangout spots (e.g., internet cafés). Each report begins with a high-level statement about its subject matter and a five-borough map of the various locations (mostly businesses) further detailed in the report. The various “locations of concern” are presented in tables with a photograph of each location, its address, and a brief banal description: “a medium-sized Arab restaurant owned by a male Syrian”; “a sign in Arabic regarding Arabic newspapers observed at the location.”

Much like the sloppy statistics undergirding CompStat’s veneer of accountable and exacting data collection, the Demographics Unit documents undermine the rhetoric of a capable Intelligence Division stopping terrorism in its tracks. The report on “Syrian Locations of Concern” includes an entire paragraph copied-and-pasted from “Pakistani Locations of Concern” without bothering to replace “Pakistan” with “Syria.” Summaries of restaurants and delis in Bay Ridge and Kensington are riddled with inaccuracies that, to some of its targets, amounted to insult atop injury — it’s bad enough to arbitrarily declare a Lebanese-owned market a “location of concern” simply because of its staff and clientele, but why add the indignity of mistaking its owners for Syrians? And how, if at all, did these inventories of neighborhood spaces and student groups help prevent terrorism? (As Goldman and Apuzzo further documented in their reporting, they didn’t.)

Once again, if the Demographics Unit was notable, it was not entirely unique to New York. In 2007 Bill Bratton, by now at the LAPD, proposed mapping Muslim communities in Los Angeles with a far more expansive effort than New York’s. Outcry from the Muslim community there ultimately stopped this effort. The NYPD Demographics Unit was disbanded in 2014 and lawsuits brought against it were settled out of court.

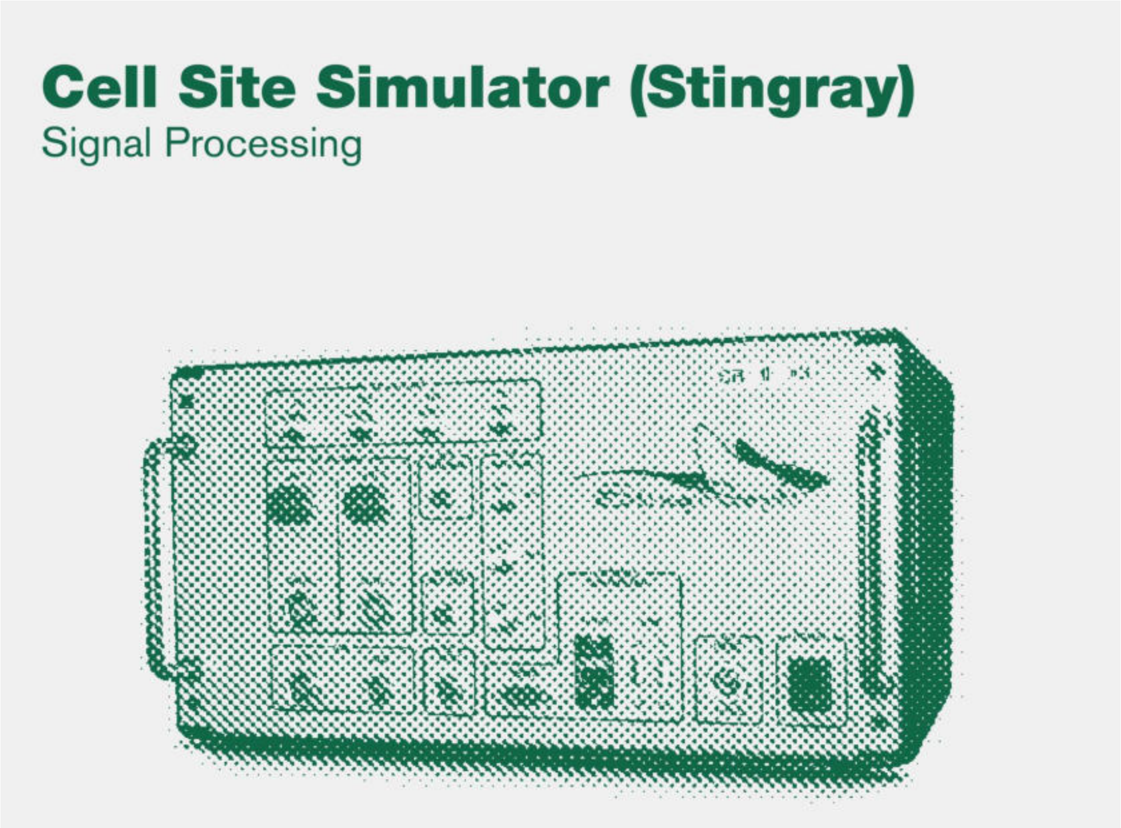
5.

In contrast to the relatively low-tech Demographics Unit's plainclothes surveillance, the other major mapping effort of the post-9/11 NYPD was defined by high-tech networked surveillance infrastructure, first around the area directly impacted by 9/11, then throughout the entire city. Rather than centering disorder on neighborhoods or intersections of low-level offenses, this camera-and-sensor driven remapping of the city produced an anticipatory geography, a map of a city where any area — rich or poor, residential or commercial, Muslim or non-Muslim — could be subject to terrorism.

This infrastructure was, some argued, well overdue: Although Bratton's NYPD had made breakthroughs with technology, the period between his departure and 9/11 was one of stasis. After 9/11, Commissioner Ray Kelly assembled a committee of experts from IBM, Merrill Lynch, and Deloitte to bring the department into the 21st century.

Wireless security cameras on lamp post deployed by New York City Police Department. Photo via Wikimedia Commons

Initially built out as the Lower Manhattan Security Initiative (LMSI) in 2005, the public-private partnership network of surveillance cameras drew $10 million from DHS and $15 million from the city, and expanded over the next seven years to become the Domain Awareness System (DAS). Created in partnership with Microsoft, the DAS is described as “a central platform used to aggregate data from internal and external closed-circuit television cameras (CCTV), license plate readers (LPRs), and environmental sensors, as well as 911 calls and other NYPD databases.” An array of military-grade surveillance technologies like Stingrays (technology for capturing cell phone data), Shotspotters (acoustic sensors that are supposed to monitor for gunshots), and backscatter vans (mobile x-ray units for searching vehicles) have been added to the NYPD’s investigation arsenal. Slightly less high-tech surveillance techniques are no less disconcerting. Today NYPD gang units collect and monitor teenagers’ social media to identify criminal conspiracy via Facebook connections, most notoriously in a major dragnet operation in Harlem in 2014.



Cell site simulators (often reported on as "Stingrays" the product name used by manufacturer Harris Corp.) imitate a cell tower and can be used to locate, identify, and on occasion intercept data from cell phones. The primary vendor for cell site simulators, Harris, requires police departments to sign extremely secretive non-disclosure agreements that can make it difficult to find out whether or not the technology is being used at all. Click here to view a non-exhaustive taxonomy of tools of data-driven policing.

Unlike the ancient CompStat database jury-rigged from off-the-shelf software, today's departments have access to an array of vendors tailoring tools to law enforcement needs. Industry giants like IBM and Microsoft recognize a market niche, while specialized companies like ShotSpotter and VIEVU body camera vendor Safariland Group have expanded to meet a growing market, subsidized with federal grants. Companies large and small burnish their reputations by perpetuating the ex-police-turned-consultant industrial complex: Hiring distinguished former officers to sit on corporate boards or provide professional insight promises both expertise and access to professional networks. (Prior to his return to the NYPD in 2013, Bratton served on the boards of ShotSpotter and Motorola Solutions — the former receiving a $1.5 million contract from the NYPD in 2015; the latter a long-time vendor to the department.)

Mayor Bill de Blasio and Police Commissioner William J. Bratton announce the deployment of "ShotSpotter," a new gunshot detection technology, in the CompStat reporting room. Photo by Rob

Bennett via Mayoral Photography Office

These new technologies and public-private partnerships might be an inevitable extension of the model first outlined on a bar napkin by Jack Maple over 20 years ago (which, on further reflection, sounds like advice for a sales team as much as a police department). As daily life has become more networked and reliant on networked infrastructure, crime mapping and data-driven policing have similarly expanded their frame to a networked vision of the city. The flat acetate maps have been replaced with real-time monitoring, mobile surveillance, wall-to-wall screens, personal histories collated from social media, and license plate readers. The core mechanisms of data collection and mapping remain, but the speed of that data collection and the rapid collating of that data with the historical record produces a map (in theory) of greater complexity for more informed decision-making. And, maybe inevitably, as the speed of data collection increased, both law enforcement and would-be vendors began to seek a shift from real-time to future tense policing — that is, trying to forecast crime before it happens.

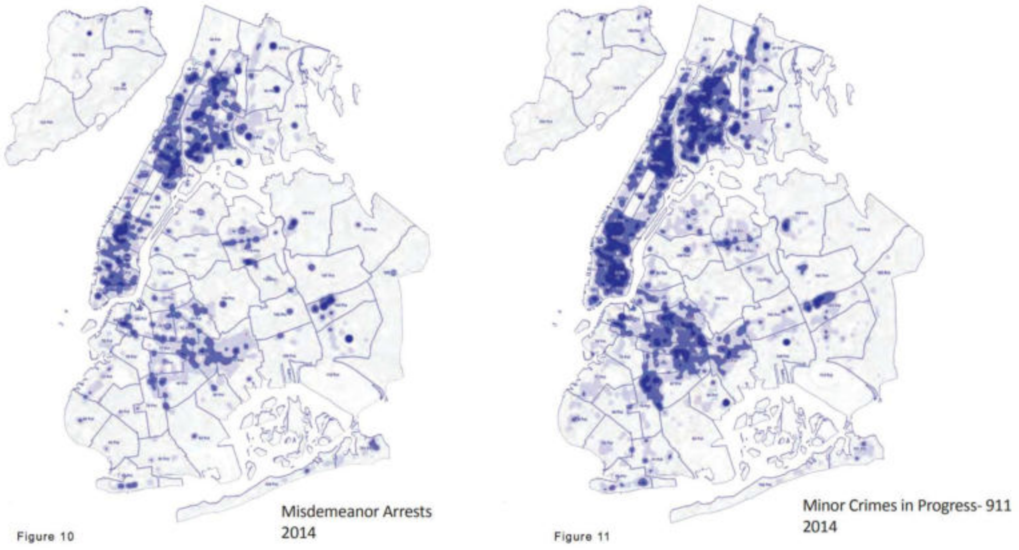
When Jack Maple first began his obsessive analog data collection methods working for the New York City Transit Police in the 1980s, he was said to have referred to his outputs as “Charts of the Future.” Emphasizing the future — using historical data to prevent crime and maintain order, rather than react to disorder as it came up — was central to much of CompStat and broken windows policing’s ideology and appeal. That preventive, pre-emptive mindset worked well with a post-9/11 “never again” attitude, which justified surveillance on Muslim neighborhoods and the exponential expansion of camera networks.

Predictive policing — which, like CompStat, is less a meaningful technical term and more of a strategic positioning of data-driven management — emerged from a convergence of technical innovations in both counterterrorism and corporate logistics. While some of the academic research and development for what would become predictive policing began with funding from the Army Research Lab, its advocates in law enforcement compared the technology’s efficiency gains to innovations in Walmart and Amazon’s warehouse distribution systems. Simultaneously, an infusion of $2 million from the National Institute of Justice supported police departments across the country partnering with academics (and future police software vendors) to experiment with using historical crime and arrest data to determine both where crimes might be statistically more or less likely to occur again and who among previously arrested individuals might be most likely to commit crime again.

Modeling individuals’ risk drew more public scrutiny given the obvious sensational *Minority Report* implications, and correspondence between past racist policing strategies and a predictive model that ascribed a higher risk of gun violence to the same racially profiled individuals. New forecasting models and alleged algorithmic sophistication can't mitigate the flaws of historical crime and policing data that mostly reflects a history of racial profiling and beat cops trying to fulfill quotas. And in the case of the aforementioned Army Research Lab-funded research that would become geographically-determined predictive policing software vendor PredPol, modeling techniques weren't even designed with law enforcement in mind — the underlying math was based on models for predicting earthquake aftershocks and was tested initially to forecast casualty rates and insurgent activities in war zones. Predictive policing foregrounds triage over understanding or addressing longer-term, systemic damages — akin to, say, mapping potential earthquake aftershocks while ignoring the fracking wells that destabilized the geology to begin with. And grid maps of possible sites of future crimes can be easily used to reinforce broken windows' old spatial model, mapping disorder away from systemic racism and economic violence and onto intersections of neighborhoods with under-funded schools or disintegrating social services.

In addition to the similarities in rhetoric, technical improvisation, and funding, predictive policing and CompStat share personal connections. Between their consulting careers, Bratton and Timoney both influenced the evolution of predictive policing. PredPol, the primary vendor of predictive policing software on the market today, emerged out of a partnership between UCLA researchers and officers working under Bill Bratton during his time as chief of the LAPD. Timoney's influence was slightly more removed: Upon taking over as Chief of the Philadelphia Police Department, Timoney threw support and resources behind a small skunk-works initiative to create mapping software for the department. The project's architects, two recent graduates from the University of Pennsylvania's landscape architecture program, would go on years later to create HunchLab, a predictive policing platform that's been used by police departments in Chicago, Greensboro, and St. Louis County. (Upon returning to the NYPD in 2013, Bratton championed predictive policing and selected HunchLab for a pilot program in 2015.)

These overlaps are not inherently conspiratorial or even purely causal. But they do demonstrate how many actors profit from the success of ever-more pervasive and ever-more trusted data-driven systems in law enforcement. The limitations of evaluating these tools based on their accuracy or effectiveness also become clear. Whether or not they "work" is as impossible to measure as any other variable deemed to influence the crime rate. A more significant question is for whom these tools actually work — who benefits from their success and who needs them to be taken seriously.



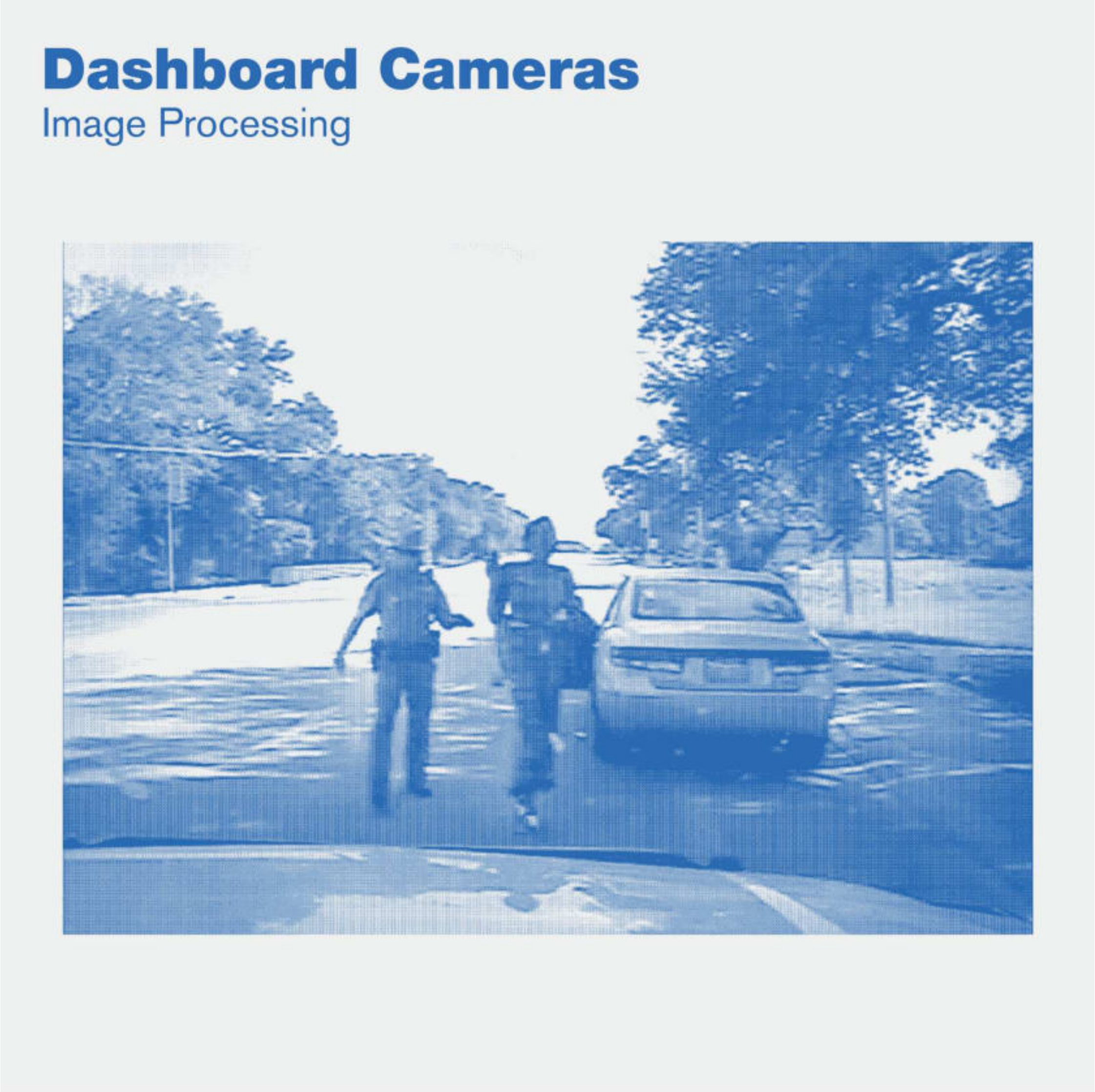
A 2015 NYPD report compared distribution of "misdemeanor arrests," left, and "minor crimes in progress," right. Image via NYPD

6.

Sometimes, instead of justifying policing methods, maps undermine them. In 2013, maps studying the frequency and demographics of NYPD stop-and-frisk incidents were one piece of the much larger undoing of this central tactic of broken windows policing. The obviously disproportionate stopping of Black and Latino New Yorkers, a reality long understood by those residents and activists, was suddenly a numerically undeniable truth.

But this data was only made available to the public through court order. The NYPD is generally reticent to disclose any of the data it insists on collecting in the service of public safety. At a June 2017 City Council hearing about the Police Oversight of Surveillance Technology (POST) Act, which proposed increased public disclosure of NYPD surveillance techniques, representatives from the NYPD expressed defensive frustration at the implication that the NYPD lacked transparency or had any reason to be deemed suspicious for its application of surveillance technologies. Gleaning information about the data and technology-driven methods the NYPD prides itself on from the NYPD (be it through FOIL requests, court orders, or legislation) typically resembles pulling teeth, if teeth had lots of lawyers and a monopoly on state violence.

Police-worn body cameras, another recent transformative policing technology ostensibly serving public accountability, aren't that easy to hold publicly accountable. Variations of cameras on and for cops have existed at least since the advent of the dashboard camera in 1990, but the tipping point for widespread adoption of body-worn cameras was in 2014, when the Obama administration requested $263 million to support body camera pilot programs across the country. The initiative followed several highly-publicized and sometimes video-documented murders of unarmed black men by police throughout the country, including Staten Island resident Eric Garner at the hands of NYPD officer Daniel Pantaleo. Although some of the outrage at the lack of accountability for police officers in these incidents was *because* the deaths had been captured on video and officers still walked free, body cameras were still heralded as transformative tools for keeping police accountable.

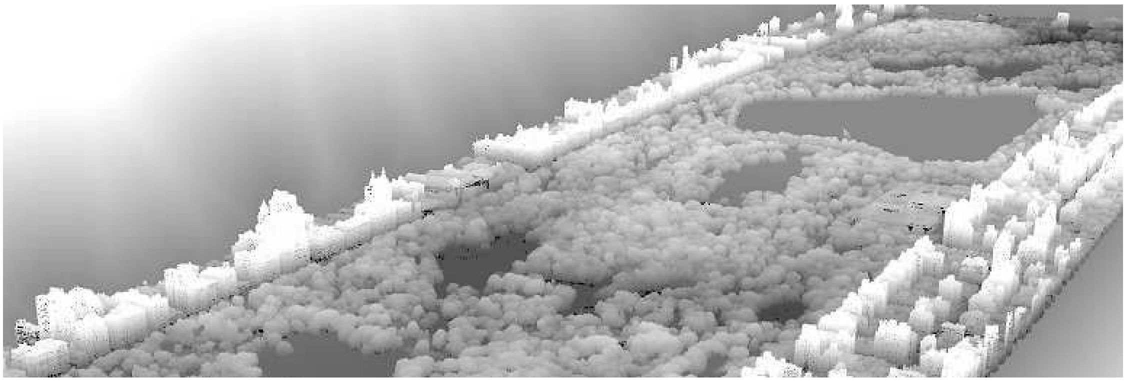


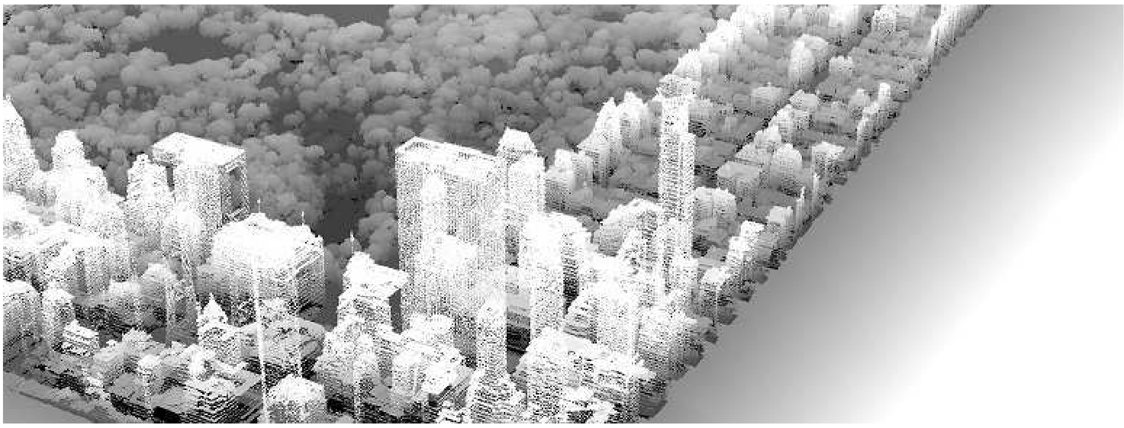
Common deployment of dashboard cameras in police vehicles came in two waves in the United States: first, in the 1980s, when Mothers Against Drunk Driving fundraised to have them installed in cars to strengthen DUI evidence; then in the 1990s during the War on Drugs as an aid to evidence that individuals had in fact consented to having their vehicle searched. Click here to view a non-exhaustive taxonomy of tools of data-driven policing.

But the question of to *whom* police should be accountable — a public, their supervisors, or the state — remains ambiguous. While cameras are framed as tools for keeping police accountable to citizens by recording their actions, camera vendors' marketing materials for police departments emphasize the value of cameras as a tool for preventing slander and holding the *public* accountable for their actions with police. As of November 2017, a survey of body camera policies conducted by Upturn and The Leadership Conference reported that over a third of major city police departments' policies don't make camera footage easily publicly accessible and only five percent had explicit policies to allow individuals filing police misconduct complaints to view relevant footage.

Body cameras also provide another data source to add to an ever-expanding map. Much like GIS technology transformed crime mapping by increasing the speed with which police departments could "see" crime patterns and combine or contextualize datasets, the threading together of other technologies (GPS, wireless communications, computer vision) and data management platforms created by body camera vendors transforms hours of archived footage that might otherwise go unwatched into indexed, geolocated, searchable evidence. The 2017 survey found only one major city police department (Baltimore) had a body camera policy with explicit limitations on using biometric technologies such as facial recognition on camera footage. Companies like police technology vendor Axon (formerly Taser) are enthusiastically pursuing the addition of biometric and analytics tools to their body camera products.

Metaphorically, the expanding temporal and networked dimensionality of today's policing maps — which, at this point, might not even be called *crime mapping* so much as *world building* — has more in common with three-dimensional spatial modeling than flat 2D vector maps. The points of hotspot policing have been replaced with point clouds. In the ideal vision of the NYPD's expansive, cutting-edge data-driven methodology, these historical, dense topographies of public and private data would accumulate into a legible terrain, one utterly absent of disorder. However, like most actual point cloud data produced with 3D scanning or lidar technologies, these maps are usually filled with "noise," glitchy and inaccurate data that requires human evaluation and editing. They are yet more maps only as reliable as their interpreters.





Light detection and ranging (LIDAR) uses pulses of light to create high-accuracy and high-resolution digital models of physical characteristics of the earth's surface. AppGeo is working with NYC's Department of Information Technology and Telecommunications (DOITT), City Parks, and the Mayor's Office of Recovery and Resilience (ORR) to develop the City's LIDAR-based tools. Image via AppGeo

Policing and urban planning have a lot in common. Their models often redesign the cities they claim to reflect, often on a biased premise of best interest. Recalling Bratton's 1994 comments to a *New York Times* reporter describing the public as his "customers" and "reduced crime, reduced fear" as his sought "profit" from the corporate strategies of CompStat meetings, one other characteristic planning and policing share is a tendency to be misunderstood as customer service models. The "customers" are also less and less the actual public, and more and more private sector actors who benefit from the worldview mapped onto the crime map. Broken windows' remapping of crisis away from diminishing social services and onto an impoverished neighborhood makes a ton of sense if your "customer" isn't the neighborhood teenagers who now fear being hassled and criminalized over a minor offense, but rather real estate brokers who'd like to raise rents or make a poor neighborhood an "up and coming" one. Investment in pervasive surveillance systems and buzzword-inflected technologies that might be vaporware makes a ton of sense if your customer isn't a Yemeni bodega owner, but rather a huge Wall Street bank anxious about keeping its employees safe from terrorism (or, in 2011, safe from Occupy Wall Street drum circles).

For almost 25 years, the model of the world constructed in NYPD cartographies has been replicated across the country, restructuring police departments and public narratives of order, justice, and the role so-called neutral technologies can play in maintaining both. There are many more beneficiaries of the world map constructed by law enforcement than vendors, consultants, and prestige-seeking upper management alone — they just played the biggest role in constructing it. To construct a counter-map of data-driven spatialized policing regimes — and, in so doing, lay a groundwork for modeling a different vision of justice or public accountability — requires teasing out that history, its players and its power dynamics, the institutional glitches in the map, and the cartographers who smooth them out.

Ingrid Burrington is a writer and artist based on a small island off the coast of America. She's the author of *Networks of New York: An Illustrated Field Guide to Urban Internet Infrastructure* and has previously written for The Atlantic, e-flux journal, and The Verge, among other outlets.

Regulate facial-recognition technology

Until appropriate safeguards are in place, we need a moratorium on biometric technology that identifies individuals, says **Kate Crawford**.

Earlier this month, Ohio became the latest of several state and local governments in the United States to stop law-enforcement officers from using facial-recognition databases. The move followed reports that the Immigration and Customs Enforcement agency had been scanning millions of photos in state driver's licence databases, data that could be used to target and deport undocumented immigrants. Researchers at Georgetown University in Washington DC used public-record requests to reveal this previously secret operation, which was running without the consent of individuals or authorization from state or federal lawmakers.

It is not the only such project. Customs and Border Protection is using something similar at airports, creating a record of every passenger's departure. The technology giant Amazon is building partnerships with more than 200 police departments to promote its Ring home-security cameras across the United States. Amazon gets ongoing access to video footage; police get kickbacks on technology products.

Facial-recognition technology is not ready for this kind of deployment, nor are governments ready to keep it from causing harm. Stronger regulatory safeguards are urgently needed, and so is a wider public debate about the impact it is already having. Comprehensive legislation must guarantee restrictions on its use, as well as transparency, due process and other basic rights. Until those safeguards are in place, we need a moratorium on the use of this technology in public spaces.

There is little evidence that biometric technology can identify suspects quickly or in real time. No peer-reviewed studies have shown convincing data that the technology has sufficient accuracy to meet the US constitutional standards of due process, probable cause and equal protection that are required for searches and arrests.

Even the world's largest corporate supplier of police body cameras — Axon in Scottsdale, Arizona — announced this year that it would not deploy facial-recognition technology in any of its products because it was too unreliable for police work and "could exacerbate existing inequities in policing, for example by penalizing black or LGBTQ communities". Three cities in the United States have banned the use of facial recognition by law-enforcement agencies, citing bias concerns.

They are right to be worried. These tools generate many of the same biases as human law-enforcement officers, but with the false patina of technical neutrality. The researchers Joy Buolamwini at Massachusetts Institute of Technology in Cambridge and Timnit Gebru, then at Microsoft Research in New York City, showed that some of the most advanced facial-recognition software failed to accurately identify dark-skinned women 35% of the time, compared to a 1% error rate for white men. Separate work showed that these technologies mismatched 28 US members of Congress to a database of mugshots, with a nearly 40% error rate for members of colour. Researchers at the University of Essex in Colchester, UK, tested a facial-recognition technology used by London's Metropolitan Police, and found it made just 8 correct matches out of a series of 42, an error rate they suspect would not be found lawful in court. Subsequently, a parliamentary committee called for trials of facial-recognition technology to be halted until a legal framework could be established.

But we should not imagine that the most we can hope for is technical parity for the surveillance armoury. Much more than technical improvements are needed. These tools are dangerous when they fail and harmful when they work. We need legal guard rails for all biometric surveillance systems, particularly as they improve in accuracy and invasiveness. Accordingly, the AI Now Institute that I co-founded at New York University has crafted four principles for a protective framework.

First, given the costly errors, discrimination and privacy invasions associated with facial-recognition systems, policymakers should not fund or deploy them until they have been vetted and strong protections have been put in place. That includes prohibiting links between private and government databases.

THESE TOOLS ARE **DANGEROUS** WHEN THEY FAIL AND **HARMFUL** WHEN THEY WORK.

Second, legislation should require that public agencies rigorously review biometric technologies for bias, privacy and civil-rights concerns, as well as solicit public input before they are used. Agencies that want to deploy these technologies should be required to carry out a formal algorithmic impact assessment (AIA). Modelled after impact-assessment frameworks for human rights, environmental protection and data protection, AIAs help governments to evaluate artificial-intelligence systems and guarantee public input.

Third, governments should require corporations to waive any legal restrictions on researching or overseeing these systems. As we outlined in the AI Now Report 2018, tech companies are currently able to use trade-secrecy laws to shield themselves from public scrutiny. This creates a legal 'black box' that is just as opaque as any algorithmic 'black box', and serves to shut down investigations into the social implications of these systems.

Finally, we need greater whistle-blower protections for technology-company employees to ensure that the three other principles are working. Tech workers themselves have emerged as a powerful force of accountability: for example, whistle-blowers revealed Google's work on a censored search engine in China. Without greater protections, they are in danger of retaliation.

Scholars have been pointing to the technical and social risks of facial recognition for years. Greater accuracy is not the point. We need strong legal safeguards that guarantee civil rights, fairness and accountability. Otherwise, this technology will make all of us less free. ■

Kate Crawford *is a distinguished research professor and co-director of the AI Now Institute at New York University, and a principal researcher at Microsoft Research in New York City.*
Twitter: @katecrawford

Sidney Fussell

How an Attempt at Correcting Bias in Tech Goes Wrong

The Atlantic, Oct 9, 2019

Google allegedly scanned volunteers with dark skin tones in order to perfect the Pixel phone's face-unlock technology.

As Silicon Valley pushes facial recognition as a convenient means to secure your laptop, board a flight, or pay for dinner, it has run into a problem: Computer vision systems have repeatedly misidentified dark-skinned black people as criminals, labeled them as gorillas, or simply failed to see them altogether.

These horrifying incidents are the unintentional results of harder-to-spot bias in the manufacturing process. When a data set used to train AI to "see" doesn't include enough people with dark skin (an underrepresentation bias), the resulting technology works differently on lighter skin than it does on darker skin (an accuracy bias). Garbage in, garbage out; racism in, racism out.

The natural solution, it would seem, is to train AI on diverse data sets. But this imperative creates its own problems. Last week, the New York *Daily News* reported that Google had sent contractors to Atlanta, Los Angeles, and college campuses across the country to collect biometric data that it could use to train the facial-recognition software in its Pixel phones. According to the *Daily News*, the contractors offered subjects $5 Starbucks gift cards in exchange for 3-D scans of their faces, taken with the Pixel. Google allegedly gave the contractors daily quotas, ordered them to prioritize subjects with dark skin, and encouraged them to approach homeless people, who it expected to be most responsive to the gift cards and least likely to object or ask questions about the terms of data collection.

Managers reportedly encouraged contractors to mischaracterize the data collection as a "selfie game," akin to Snapchat filters such as Face Swap. College students who agreed to the scans later told the *Daily News* that they didn't recall ever hearing the name Google and were simply told to play with the phone in exchange for a gift card. To entice homeless users in L.A. to consent, contractors were allegedly instructed to mention a California law that allows the gift cards to be exchanged for cash. The whole episode is, in a bleak way, an apparent attempt to diversify AI training data while paying people for their information. But the result is completely dystopian.

According to *The New York Times*, Google temporarily suspended the data collection, pending an internal investigation. In an emailed statement to *The Atlantic*, a Google spokeswoman said, "We're taking these claims seriously and investigating them. The allegations regarding truthfulness and consent are in violation of our requirements for volunteer research studies and the training that we provided."

It's baffling that this purported scheme, which the *Daily News*'s reporting suggests commodified black and homeless Americans, was intended to reduce racial bias. But as the Harvard technologist Shoshana Zuboff has argued, people have always been the "raw materials" for Big Tech. Products such as the Pixel and the iPhone, and services such as Google and Facebook, collect our data as we use them; companies refine that data, and, with each new generation, sell us more advanced products that collect more useful data. In this framework, our habits, our choices, our likes, and our dislikes are not unlike soybeans or petroleum or iron ore—natural resources that are extracted and processed by huge firms, for massive profit.

Sometimes this looks like a smart thermostat getting better at predicting how cool you like your home, and sometimes it looks like a $1 trillion company allegedly offering $5 gift cards to homeless black people to better sell a $1,200 phone.

As the techlash continues, some lawmakers are seeking to empower their constituents to demand that companies such as Google pay users for their data. California and Alaska have debated legislation to charge companies for using people's personal data. Andrew Yang, the 2020 Democratic presidential candidate, has advocated treating data as a "property right." The Facebook co-founder Chris Hughes suggests a "data dividend," a revenue tax on companies monetizing enormous amounts of public data, paid out to users across the country, like universal basic income.

But following that line of thinking makes it clear that we still have no ethical or economic framework for valuing data collected from people across different social contexts. Should tech companies pay more for dark-skinned subjects because they're underrepresented in training data? If our bodies are commodities, what's a fair price, and who should set it? The data-ownership idea is, fundamentally, limited: Even if we manage, with the help of Hughes or Yang or state legislatures, to negotiate a high price for our data, we're still for sale.

In a backwards way, movements to pay users for the data that tech companies take from them only corroborate the process by which Silicon Valley turns our faces into commodities. Imagine an unregulated race-to-the-bottom market where companies target the most vulnerable for their data, restrained only by the alarmingly low bar for consent to improve their products. It would look a lot like paying homeless people $5 for a face scan.

Sidney Fussell is a staff writer at The Atlantic, where he covers technology.

Clare Garvie

Garbage In, Garbage Out

Face Recognition on Flawed Data

May 16, 2019

Clare Garvie

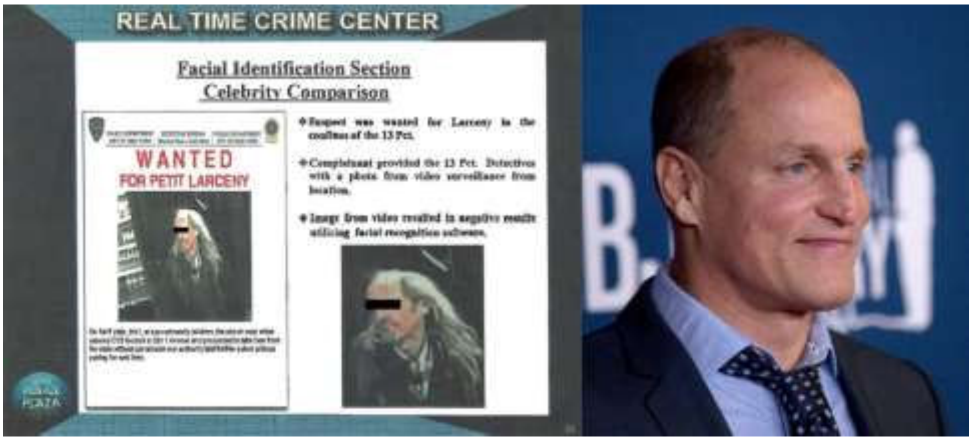
Introduction

On April 28, 2017, a suspect was caught on camera reportedly stealing beer from a CVS in New York City. The store surveillance camera that recorded the incident captured the suspect's face, but it was partially obscured and highly pixelated. When the investigating detectives submitted the photo to the New York Police Department's (NYPD) facial recognition system, it returned no useful matches.[1]

Rather than concluding that the suspect could not be identified using face recognition, however, the detectives got creative.

One detective from the Facial Identification Section (FIS), responsible for conducting face recognition searches for the NYPD, noted that the suspect looked like the actor Woody Harrelson, known for his performances in *Cheers*, *Natural Born Killers*, *True Detective*, and other television shows and movies. A Google image search for the actor predictably returned high-quality images, which detectives then submitted to the face recognition algorithm in place of the suspect's photo. In the resulting list of possible candidates, the detectives identified someone they believed was a match—not to Harrelson but to the suspect whose photo had produced no possible hits.[2]

This celebrity "match" was sent back to the investigating officers, and someone who was not Woody Harrelson was eventually arrested for petit larceny.



REAL TIME CRIME CENTER

Facial Identification Section
Celebrity Comparison

WANTED
FOR PETIT LARCENY

[redacted]

- Suspect was wanted for Larceny in the confines of the 13 Pct.
- Complainant provided the 13 Pct. Detectives with a photo from video surveillance from location.
- Image from video resulted in negative results utilizing facial recognition software.

[redacted]

Figure 1 : On the left: a slide from the NYPD FIS describing its "celebrity comparison" technique. On the right, a photo of Woody Harrelson. (Source: left, NYPD; right, Gabriel Cristóver Pérez/LBJ Presidential Library.)

There are no rules when it comes to what images police can submit to face recognition algorithms to generate investigative leads. As a consequence, agencies across the country can—and do—submit all manner of "probe photos," photos of unknown individuals submitted for search against a police or driver license database. These images may be low-quality surveillance camera stills, social media photos with filters, and scanned photo album pictures.[3] Records from police departments show they may also include computer-generated facial features, or composite or artist sketches.[4]

Or the probe photo may be a suspect's celebrity doppelgänger. Woody Harrelson is not the only celebrity to stand in for a suspect wanted by the NYPD. FIS has also used a photo of a New York Knicks player to search its face recognition database for a man wanted for assault in Brooklyn.[5]

The stakes are too high in criminal investigations to rely on unreliable—or wrong—inputs. It is one thing for a company to build a face recognition system designed to help individuals find their celebrity doppelgänger[6] or painting lookalike[7] for entertainment purposes. It's quite another to use these techniques to identify criminal suspects, who may be deprived of their liberty and ultimately prosecuted based on the match. Unfortunately, police departments' reliance on questionable probe photos appears all too common.

Garbage In, Garbage Out

"Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?"

—Charles Babbage[8]

"Garbage in, garbage out" is a phrase used to express the idea that inputting low-quality or nonsensical data into a system will produce low-quality or nonsensical results. It doesn't matter how powerful or cleverly-designed a system is, it can only operate on the information it is provided—if data is missing, the system cannot operate on it. Any attempt to reconstruct or approximate missing data will necessarily be a "guess" as to what information that data contained.

Worse, if data is wrong—like a photo of someone other than the suspect—the system has no way to correct it. It has literally no information about the suspect, and can't make it up.

Photos that are pixelated, distorted, or of partial faces provide less data for a face recognition system to analyze than high-quality, passport-style photos, increasing room for error.[9]

Face recognition technology has improved immensely in the past two years alone, enabling rapid searches of larger databases and more reliable pairings in testing environments.[10] But it doesn't matter how good the machine is if it is still being fed the wrong figures—the wrong answers are still likely to come out.

1. Composite sketches as probe images

"Composite art is an unusual marriage of two unlikely disciplines: police investigative work and art …. It is essential to realize that a composite sketch is a drawing of a victim's or witness's perception of a perpetrator at the time he or she was observed. It is not meant to be an exact portrait of the suspect. Keep the two words 'likeness' and 'similarity' in mind at all times. This is the best a composite sketch can achieve."

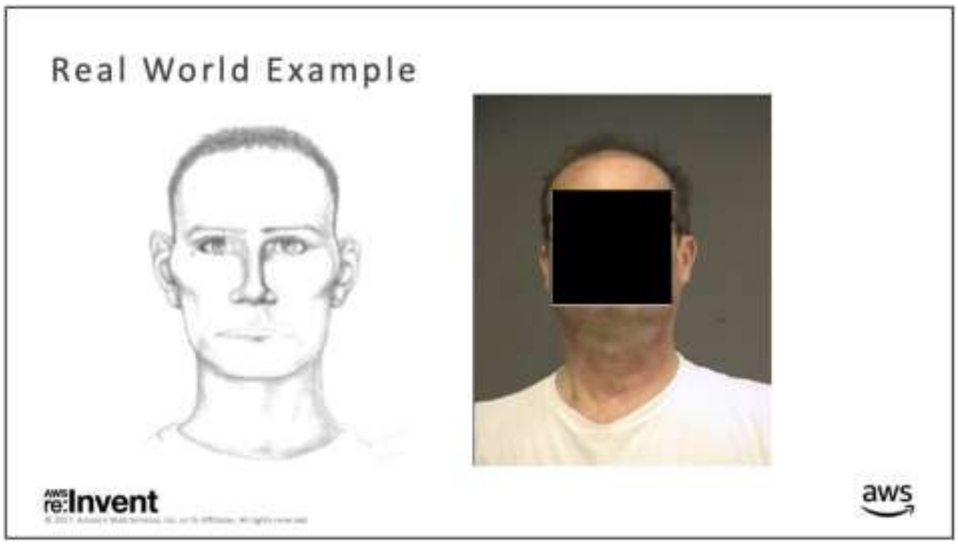
—*The Police Composite Sketch*[11]

In early 2018, Google rolled out "Art Selfie" — an app designed to match a user's photo to a famous painting lookalike using face recognition.[12] The result is an often-humorous photo pairing and an opportunity to learn more about art.

Less humorous is the fact that some police departments do the same thing when looking for criminal suspects, just in reverse—submitting art in an attempt to identify real people.

At least half a dozen police departments across the country permit, if not encourage, the use of face recognition searches on forensic sketches.

At least half a dozen police departments across the country permit, if not encourage, the use of face recognition searches on forensic sketches—hand drawn or computer generated composite faces based on descriptions that a witness has offered. In a brochure informing its officers about the acquisition of face recognition, the Maricopa County Sheriff's Office in Arizona states: "[T]he image can be from a variety of sources including police artist renderings," and that the technology "can be used effectively in suspect identifications using photographs, surveillance still and video, suspect sketches and even forensic busts."[13] A presentation about the face recognition system that the Washington County Sheriff's Department in Oregon operates includes a "Real World Example" of the technology being used to identify an artist's drawing of a face.[14]



Real World Example

[redacted]

Figure 2 : Slide from an AWS presentation titled "Washington County Sheriff's Office Rekognition Case Study." (Source: Public records obtained by ACLU Oregon & Northern California.)

A face recognition Privacy Impact Assessment that a working group of 15 state and federal agencies authored in 2011 states that it should be permissible to use face recognition to "...identify suspects based upon artist's sketches."[15] Information about the Maryland Department of Public Safety and Correctional Services, the Northern Virginia Regional Information System, and the Pinellas County Sheriff's Office in Florida suggest that sketches could be submitted to these agencies' face recognition systems as well.[16]

This practice is endorsed by some of the companies providing these face recognition systems to police departments. The example from the Washington County in Figure 2 is part of a case study that Amazon Web Services highlighted in a presentation about the capabilities of its face recognition software, Rekognition. Cognitec, one of the leading providers of face recognition algorithms to U.S. law enforcement, promotes the use of its software to "identify individuals in crime scene photos, video stills and sketches."17 Vigilant Solutions markets tools specifically for "creating a proxy image from a sketch artist or artist rendering" to be submitted to its face recognition system.18

A. Scientific review of composite image face recognition

Even the most detailed sketches make poor face recognition probe images. The Los Angeles County Sheriff's Department face recognition user guide summarizes this well:

"A photograph taken of a real person should be used. Composite drawing will have marginal success because they are rendered pictures and do not accurately detail precise features."19

Studies that have analyzed the performance of face recognition systems on composite sketches conclude the same. A 2011 Michigan State University study noted that "[c]ommercial face recognition systems are not designed to match forensic sketches against face photographs."20 In 2013, researchers studying this question ran sketches against a face recognition database using a commercially-available algorithm from Cognitec—one of the companies that advertises this as a feature of its system. The algorithm was programmed to return a list of 200 possible matches searching a database of 10,000 images. For sketches, it retrieved the correct match between 4.1 and 6.7 percent of the time.21 Put another way, in only about 1 of every 20 searches would the correct match show up in the top 200 possible matches that the algorithm produced.22

In 2014, the National Institute of Standards and Technology (NIST) found similarly poor results, concluding that "[s]ketch searches mostly fail."23 The NYPD has separately concluded the same thing from their own experience. According to NYPD detective Tom Markiewicz, FIS has tried running face recognition on sketches in the past and found that "sketches do not work."24 So did the Pinellas County Sheriff's Office, concluding that the practice "is doubtful on yielding successful results with the current [system]" —yet it still permits the practice nonetheless.25

B. Forensic sketches and misidentification

The most likely outcome of using a forensic sketch as a probe photo is that the system fails to find a match—even when the suspect is in the photo database available to law enforcement. With this outcome, the system produces no useful leads, and investigating officers must go back to the drawing board.

But this practice also introduces the possibility of misidentification. The process of generating a forensic sketch is inherently subjective. Sketches typically rely on:

1. An eyewitness's memory of what the subject looked like;
2. The eyewitness's ability to communicate the memory of the subject to a sketch artist;
3. The artist's ability to translate that description into an accurate drawing of the subject's face, someone whom the artist has never seen in person.26

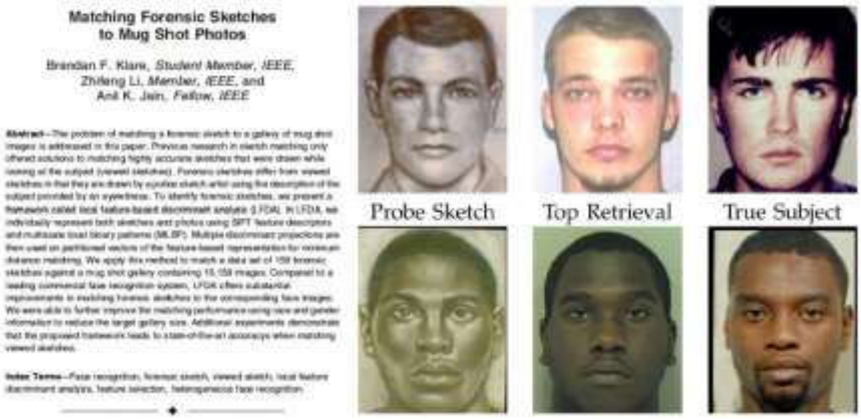

Matching Forensic Sketches
to Mug Shot Photos
Brendan F. Klare, Student Member, IEEE,
Zhifeng Li, Member, IEEE, and
Anil K. Jain, Fellow, IEEE
Probe Sketch
Top Retrieval
True Subject

Figure 3 : Examples where an imposter, not the subject of the forensic sketch, is returned as the highest ranking face recognition match. (Source: Klare, Li, & Jain (2010), all rights reserved.)

Each of these steps introduces elements of subjective interpretation and room for error.27 For example, an eyewitness may not remember the shape of the subject's jaw, yet the resulting sketch will necessarily include one. Or the witness may remember the suspect had "bug eyes," something the artist would need to interpret figuratively rather than literally.28 As a consequence, the resulting sketch may actually look more like someone in the face recognition database other than the subject being searched for, as illustrated in Figure 3.

In this scenario, human review of the face recognition matches will not be able to remove the risk of error. When examining the face recognition results for a possible match, the analyst will have only the sketch to refer back to. The analyst will have no basis to evaluate whether the image accurately represents the subject being searched for. This compounds the risk that the face recognition search will lead to an investigation, if not an arrest, of the wrong person.

2. An art or a science? Computer-generated facial features

A white paper titled "Facial Recognition: Art or Science?" published by the company Vigilant Solutions posits that face recognition systems—even without considering composite sketches—are "[p]art science and part art."29 The "art" aspect is the process of modifying poor quality images before submitting them to a recognition algorithm to increase the likelihood that the system returns possible matches.30

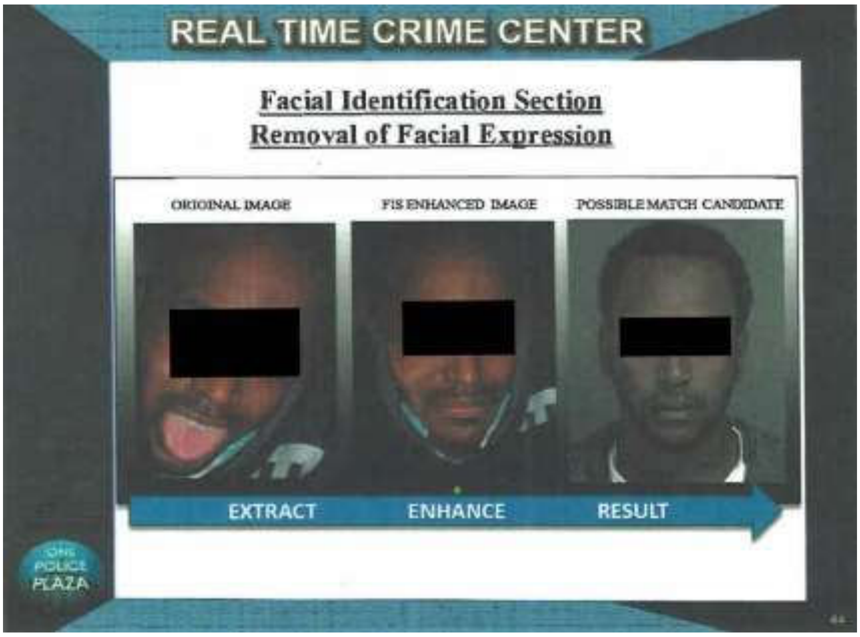
Editing photos before submitting them for search is common practice, as suggested by responses to records requests and a review of the software packages that face recognition vendor companies offer. These documents also illustrate that the edits often go well beyond minor lighting adjustments and color correction, and often amount to fabricating completely new identity points not present in the original photo.

One technique that the NYPD uses involves replacing facial features or expressions in a probe photo with ones that more closely resemble those in mugshots—collected from photos of other people. Presentations and interviews about FIS include the following examples:

- "Removal of Facial Expression"—such as replacing an open mouth with a closed mouth. In one example provided in a NYPD presentation, detectives conducted "...a Google search for Black Male Model" whose lips were then pasted into the probe image over the suspect’s mouth.31
- "Insertion of Eyes"—the practice of "graphically replacing closed eyes with a set of open eyes in a probe image," generated from a Google search for a pair of open eyes.32

- Mirrored effect on a partial face—copying and mirroring a partial face over the Y axis to approximate the missing features, which may include adding "[e]xtra pixels … to create a natural appearance of one single face."33
- "Creating a virtual probe"—combining two face photographs of different people whom detectives think look similar to generate a single image to be searched, to locate a match to one of the people of the combined photograph.34
- Using the "Blur effect" on an overexposed or low-quality image—adding pixels to a photo that otherwise doesn't have enough detail "to render a probe that [has] a similar nose, mouth, and brow as that of the suspect in the photo."35
- Using the "Clone Stamp Tool" to "create a left cheek and the entire chin area" of a suspect whose face was obscured in the original image.36

Another technique that the NYPD and other agencies employ involves using 3D modeling software to complete partial faces and to "normalize" or rotate faces that are turned away from the camera. After generating a 3D model, the software will fill in the missing facial data with an approximation of what it should look like, based on the visible part of what the subject's face looks like as well as the measurements of an "average" face.37 According to the NYPD, the software creates "a virtual appearance of the suspect looking straight ahead, replicating a pose of a standard mugshot."38


REAL TIME CRIME CENTER

Facial Identification Section

Removal of Facial Expression

Figure 4 : A slide from NYPD FIS describing "Removal of Facial Expression" technique. (Source: NYPD.)

These techniques amount to the fabrication of facial identity points: at best an attempt to create information that isn't there in the first place and at worst the introduction of evidence that matches someone other than the person being searched for. During a face recognition search on an edited photo, the algorithm doesn't distinguish between the parts of the face that were in the original evidence—the probe photo—and the parts that were either computer generated or added in by a detective, often from photos of different people unrelated to the crime.39 This means that the original photo could represent 60 percent of a suspect's face, and yet the algorithm could return a possible match assigned a 95 percent confidence rating, suggesting a high probability of a match to the detective running the search.40

If it were discovered that a forensic fingerprint expert was graphically replacing missing or blurry portions of a latent print with computer-generated—or manually drawn—lines, or mirroring over a partial print to complete the finger, it would be a scandal.41 The revelation could lead to thousands of cases being reviewed, possibly even convictions overturned.42

3. Results as “investigative leads only…”

Most agencies do not yet consider face recognition to be a positive identification. Many law enforcement agencies, the NYPD included, state that the results of a face recognition search are possible matches only and must not be used as positive identification.[43]

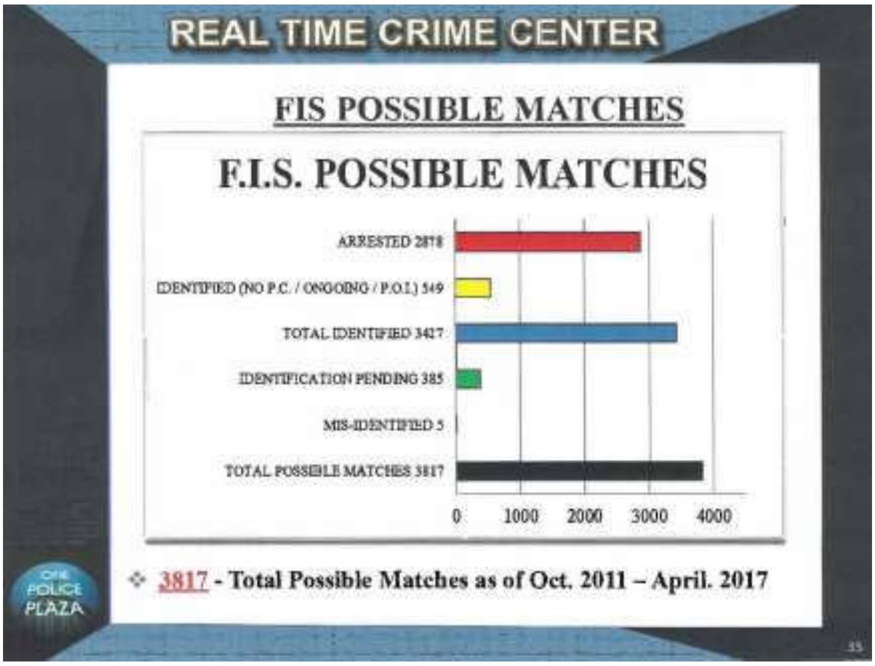
In theory, this is a valuable check against possible misidentifications, including those introduced into the system by inputting celebrity comparisons, composite sketches, or other computer-altered photographs that don’t accurately represent the person being searched for.

However, in most jurisdictions, officers do not appear to receive clear guidance about what additional evidence is needed to corroborate a possible face recognition match. The NYPD guide states: “Additional investigative steps must be performed in order to establish probable cause to arrest the Subject [sic]” of the face recognition search.[44] But what or how many additional steps are needed, and how independent they must be from the face recognition process, is left undefined.

Absent this guidance, the reality is that suspects are being apprehended almost entirely on the basis of face recognition “possible matches.” For example:

- In a recent case, NYPD officers apprehended a suspect and placed him in a lineup solely on the basis of a face recognition search result.[45] The ultimate arrest was made on the basis of the resulting witness identification, but the suspect was only in the lineup because of the face recognition process.
- NYPD officers made an arrest after texting a witness a single face recognition “possible match” photograph with accompanying text: “Is this the guy…?” The witness’ affirmative response to viewing the single photo and accompanying text, with no live lineup or photo array ever conducted, was the only confirmation of the possible match prior to officers making an arrest.[46]
- Sheriffs in Jacksonville, Florida, who were part of an an undercover drug sale arrested a suspect on the basis of the face recognition search. The only corroboration was the officers’ review of the photograph, presented as the “most likely” possible match from the face recognition system.[47]
- A Metro Police Department officer in Washington, D.C., similarly printed out a “possible match” photograph from MPD’s face recognition system and presented that single photograph to a witness for confirmation. The resulting arrest warrant application for the person in the photograph used the face recognition match, the witness confirmation, and a social media post about a possible birth date (month and day only) as the only sources of identification evidence.[48]

There are probably many more examples that we don’t know about. These represent a fraction of the cases that have used face recognition to assist in making an identification. The NYPD made 2,878 arrests pursuant to face recognition searches in the first 5.5 years of using the technology.[49] Florida law enforcement agencies, including the Jacksonville Sheriff’s Office, run on average 8,000 searches per month of the Pinellas County Sheriff’s Office face recognition system, which has been in operation since 2001.[50] Many other agencies do not keep close track of how many times their officers run face recognition searches and whether these searches result in an arrest.



REAL TIME CRIME CENTER

FIS POSSIBLE MATCHES

❖ 3817 - Total Possible Matches as of Oct. 2011 – April. 2017

Figure 5 : In the first 5.5 years of operation, the NYPD's face recognition system led to 2,878 arrests. NYPD Det. Markiewicz estimates that 8,000 cases will have used a face recognition search in 2018 alone. (Source: NYPD.)

Another valuable check against mistaken identification—and unreliable investigative leads—would be to allow defendants access to the inputs and outputs of a face recognition search that resulted in their arrest. But this does not happen. Even though prosecutors are required under federal law to disclose any evidence that may exonerate the accused, defense attorneys are not typically provided with information about "virtual probes," celebrity doppelgängers, or really any information about the role face recognition played in identifying their client.51 This is a failure of the criminal justice system to protect defendants' due process.52

It may be that many of those arrested on the basis of questionable face recognition searches did in fact commit the crime of which they were accused. But the possibility that they didn't—that the face recognition system identified the wrong person—looms large in the absence of additional, independent police investigation and sufficient access to the evidence by the defense. This is risky, and the consequences will be borne by people investigated, arrested, and charged for crimes they didn't commit.

4. Conclusion and recommendations

There is no easy way to discover just how broad of a trend this represents—and just how many arrests have been made in large part on the basis of celebrity lookalikes, artist sketches, or graphically altered faces submitted to face recognition systems.53

But we can anticipate that the problem will get a lot bigger. Police departments across the country are increasingly relying on face recognition systems to assist their investigations. In addition, an official for the Federal Bureau of Investigation (FBI), which runs its own face recognition system, has indicated that the agency plans to do away with the "investigative lead only" limitation altogether. At a conference in 2018, FBI Section Chief for Biometric Services Bill McKinsey said of the FBI: "We're pretty confident we're going to have face [recognition] at positive ID in two to three years."54

In setting this goal, the FBI has assumed that the results of face recognition systems will become more accurate as the algorithms improve. But these improvements won't matter much if there are no standards governing what police departments can feed into these systems. In the absence of those rules, we believe that a moratorium on local, state, and federal law enforcement use of face recognition is appropriate and necessary.

The stakes are too high in criminal investigations to rely on unreliable—or wrong—inputs.

Law enforcement agencies that persist in using face recognition in their investigations should at a minimum take steps to reduce the risk of misidentification and mistake on the basis of unreliable evidence. These steps include:

- Stop using celebrity look-alike probe images. Face recognition is generally considered to be a biometric, albeit an imperfect one. Police cannot substitute one person's biometrics for another's, regardless of whatever passing resemblance they may have.
- Stop submitting artist or composite sketches to face recognition systems not expressly designed for this purpose. Sketches are highly unlikely to result in a correct match—and carry a real risk of resulting in a misidentification that a human review of the possible matches cannot correct.
- Establish and follow minimum photo quality standards, such as pixel density and the percent of the face that must be visible in the original photo, and prohibit the practice of pasting other people's facial features into a probe. Any photo not meeting these minimum standards should be discarded—not enhanced through the addition of new identity points like another person's mouth or eyes.
- If edits to probe images are made, carefully document these edits and their results. Retain all versions of the probe image submitted to the face recognition system for production to the defense.
- Require that any subsequent human review of the face recognition possible match be conducted against the original photo, not a photo that has undergone any enhancements, including color and pose correction.
- As is the practice in some police departments, require double-blind confirmation. The face recognition system should produce an investigative lead only if two analysts independently conclude that the same photo is a possible match.
- Provide concrete guidance to investigating officers about what constitutes sufficient corroboration of a possible match generated by a face recognition system before law enforcement action is taken against a suspect. This should include: mandatory photo arrays; a prohibition on informing witnesses that face recognition was used; and a concrete nexus between the suspect and the crime in addition to the identification, such as a shared address.
- Make available to the defense any information about the use of face recognition, including the original probe photo, any edits that were made to that photo prior to search, the resulting candidate list and the defendant's rank within that list, and the human review that corroborated the possible match.
- Prohibit the use of face recognition as a positive identification under any circumstance.

These recommendations should be considered as minimum requirements, and are made in addition to the broader recommendations the Center on Privacy & Technology made in its 2016 report, *The Perpetual Line-up: Unregulated Police Face Recognition in America*.55

As the technology behind these face recognition systems continues to improve, it is natural to assume that the investigative leads become more accurate. Yet without rules governing what can—and cannot—be submitted as a probe photo, this is far from a guarantee. Garbage in will still lead to garbage out.

5. Acknowledgements

This report would not be possible without the tireless advocacy of Professor David Vladeck, Stephanie Glaberson, and numerous student lawyers of the Georgetown Law Civil Litigation Clinic, which represents the Center on Privacy & Technology in our public records lawsuit against the New York City Police Department. Only with the assistance of the clinic have we been able to recover thousands of pages of documents regarding use of face recognition technology by the NYPD, even though the agency itself has tried hard to keep its use of this technology hidden from public view.

Critical guidance and close reading were provided by our team of outside reviewers, who will remain anonymous, but who lent us their expertise on New York City policing, criminal litigation, and the technical functioning of face recognition systems. This report would not be possible without the entire team at the Center, who helped in countless ways: Alvaro Bedoya, Laura Moy, Katie Evans, Harrison Rudolph, Jameson Spivack, Gabrielle Rejouis, and Julia Chrusciel. We are also grateful to the Center's research assistants and summer fellows; our copy editor, Joy Metcalf; our design and web development firm, Rootid; and our cover designer, Eve Tyler.

We also acknowledge, with gratitude, the work of our friends and allies at other organizations also striving to shed light on how face recognition technology is used and to prevent powerful police tools from being used in ways that are harmful to individuals and communities. In particular, perhaps no one has done more to address and expose harmful, secret, and unfair uses of police technology than criminal defense attorneys, many of whom continue to provide us with invaluable guidance.

The Center on Privacy & Technology at Georgetown Law is supported by the Ford Foundation, the Open Society Foundations, the MacArthur Foundation, Luminate, the Media Democracy Fund, and Georgetown University Law Center.

- [1] NYPD, Real Time Crime Center Facial Identification Section (FIS), presentation by Detective Markiewicz (Sept. 17, 2018) (notes on file with author).
- [2] *Id.*

- [3] *See, e.g.*, Eric Sofge, *The End of Anonymity*, Popular Science (Jan. 15, 2014), https://www.popsci.com/article/technology/end-anonymity (describing the Pennsylvania system as used in Cheltenham Township, Pa.).
- [4] *See, e.g.*, Washington County Sheriff's Office, *PSWeb Facial Recognition Training Guide,* 47, *available at* https://www.aclunc.org/docs/20180522_ARD.pdf#page=47 (PDF).
- [5] NYPD, *Facial Identification Section Case #8: Celebrity Comparison*, Document p. 025428. The name and image of the New York Knicks player has been redacted in the files provided to the Center by the NYPD.
- [6] *See* Phoebe Weston, *Who is YOUR celebrity lookalike? Find out with this online AI tool that reveals your famous doppelganger*, Daily Mail (Mar. 30, 2017) https://www.dailymail.co.uk/sciencetech/article-4363640/Who-celebrity-lookalike-online-tool.html.
- [7] Hamza Shaban, *A Google app that matches your face to artwork is wildly popular. It's also raising privacy concerns.*, Washington Post (Jan. 17, 2018), https://www.washingtonpost.com/news/the-switch/wp/2018/01/16/google-app-that-matches-your-face-to-artwork-is-wildly-popular-its-also-raising-privacy-concerns/.

- [8] Charles Babbage, *Passages from the Life of a Philosopher* 67 (Longman, Green, Longman, Roberts, & Green ed. 1864).
- [9] *See* Zhifei Wang et al., *Low-resolution face recognition: a review*, 30 The Visual Computer 359, 359–360 (April 2014), *available at* https://link.springer.com/article/10.1007/s00371-013-0861-x.
- [10] Patrick Grother et al., National Institute of Standards and Technology, *Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification* 2 (Nov. 2018), https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8238.pdf (PDF) ("The major result of the evaluation is that massive gains in accuracy have been achieved in the last five years (2013–2018).").

- [11] Stephen Manusci, The Police Composite Sketch 6–7 (Humana Press 2010).
- [12] Art Selfie, Google Arts & Culture, https://artsandculture.google.com/camera/selfie (last accessed Jan. 28, 2019).

- 13. Maricopa County Sheriff's Office (MCSO), *Counter-Terrorism Information Center Facial Recognition*, Document p. 014951; MCSO, *Homeland Security & National Facial Recognition Network Briefing Paper* (Oct. 6, 2008), Document p. 014952; MCSO, MCSO/ACTIC *Facial Recognition Procedures: Image Records Request*, Document p. 014962.
- 14. Washington County Sheriff's Office, *PSWeb Facial Recognition Training Guide*, 47, *available at* https://www.aclunc.org/docs/20180522_ARD.pdf#page=47 (PDF).

- 15. Nlets, *Privacy Impact Assessment Report for the Utilization of Facial Recognition Technologies to Identify Subjects in the Field* (2011), Document p. 016668.
- 16. Baltimore Police Dep't, *Governor's Office of Crime Control & Prevention (GOCCP) Fact Sheet: Facial Recognition* (Apr. 2015), Document p. 010954; Maryland Dep't of Public Safety & Correctional Services, *GOCCP Fact Sheet: Criminal Justice Dashboard* (Apr. 2015), Document p. 011104; Northern Virginia Regional Information System, *LOB #207: NOVARIS* (2016), Document p. 015231; Pinellas County Sheriff's Office, *Interagency Use of Facial Recognition...Does it work?*, 43–52, *available at* https://www.aamva.org/uploadedFiles/MainSite/Content/EventsEducation/Event_Materials/2013/2013_Region_II_Conference/061013_10_30_FR_Complete.pdf (PDF).
- 17. Center for Advancing Retail and Technology, *Cognitec: FaceVACS-VideoScan*, https://www.advancingretail.org/solutions/cognitec. ("Law enforcement professionals can identify individuals in crime scene photos, videos stills and sketches by matching facial images against the agency's mugshot repository"). *See also:* Cognitec, FaceVACS-DBScan LE: Face Recognition Technology for for image and video investigations, and database matching, https://www.cognitec.com/files/layout/downloads/FaceVACS-DBScan-LE-1-1-flyer.pdf (PDF) ("supports investigation of faces in video footage, still images and sketches").
- 18. Vigilant Solutions, *FaceSearch,* https://www.vigilantsolutions.com/products/facial-recognition/(last viewed May 13, 2019). Vigilant Solutions is now part of Motorola Solutions. *See* Susan Crandall, *Motorola Solutions Acquires VaaS Holdings, Leader in Data and Image Analytics for Vehicle Location*, Vigilant Solutions (Jan. 7, 2019), https://www.vigilantsolutions.com/motorola-solutions-acquires-vaas-international-holdings-leader-data-image-analytics-vehicle-location/. In a 2008 contract to provide a face recognition solution to Utah's Department of Public Safety, Hummingbird Communications also indicated that its solution can "identify individuals from … Police Artist Sketches … or any image from any number or variety of sources." Utah State Analysis and Information Center, *State of Utah Contract with Hummingbird Garden Ranch LLC (*Dec. 22, 2008), Document p. 108705.

- 19. Los Angeles County Sheriff's Office, *Facial Recognition & Comparison: Create a Good Source Image*, Document p. 000681.
- 20. Anil Jain et al, *Face Recognition: Some Challenges in Forensics*, IEEE Int'l Conference on Automatic Face and Gesture Recognition (Mar. 2011), *available at* https://ieeexplore.ieee.org/document/5771338.
- 21. Scott Klum, Hu Han, Anil Jain, & Brendan Klare, *Sketch Based Face Recognition: Forensic vs. Composite Sketches* (2013), *available at* https://openbiometrics.org/publications/klum2013sketch.pdf (PDF) ("In forensic and biometrics scenarios involving facial sketch to mugshot matching, the standard procedure involves law enforcement officers looking through top-N matches (rather than only considering rank-one retrieval rates). In our experiments, N = 200. We also used the performance of a commercial-off-the-shelf face matcher, FaceVACS v8.2 as a baseline. As shown in Fig. 5, FaceVACS achieves rank-200 retrieval rates of 4.1% and 6.7% for forensic and composite sketches, respectively.")
- 22. *Id.*
- 23. Patrick Grother & Mei Ngan, *Face Recognition Vendor Test (FRVT): Performance of Face Identification Algorithms, NIST Interagency Report 8009*, 4 (May 26, 2014) https://nvlpubs.nist.gov/nistpubs/ir/2014/NIST.IR.8009.pdf (PDF) ("By searching a non-operational set of sketch images against photographs seeded into a population of 640,000 nonmated mugshots, the most accurate algorithms produce the mated photograph only infrequently: The mate is not among the top 50 candidates at the following rates: 73.3% (3M/Cogent), 73.8% (NEC), 78.5% (Toshiba), 80.3% (Morpho), and 81.5% (Neurotechnology).") Note these accuracy results appear much higher than those in the Michigan State University study, likely because NIST used sketches created by an artist viewing the mugshot, not sketches created based on an eyewitness description of the suspect, which is more akin to real-world scenarios. *Id.* at 39–40 ("the fact that the sketches were prepared by an artist viewing the exemplar photograph probably means that the accuracy measurements here represent a "best case" upper bound on accuracy.").
- 24. *FIS Presentation* (Sept. 17, 2018) (on file with author).
- 25. Lance Taylor, Ga. Dep't Driver Serv., Moderation of Interagency Use of Facial Recognition… Does it Work? at the 2013 AAMVA Region II Conference 43–53 (June 10, 2013), https://www.aamva.org/uploadedFiles/MainSite/Content/EventsEducation/Event_Materials/2013/2013_Region_II_Conference/061013_10_30_FR_Complete.pdf (PDF).

- 26. *See* Anil Jain et. al., *Face Recognition: Some Challenges in Forensics*, IEEE Int'l Conference on Automatic Face and Gesture Recognition (Mar. 19, 2011), https://ieeexplore.ieee.org/document/5771338.

- 27. *See* Stephen Manusci, The Police Composite Sketch 22–23 (Humana Press 2010). ("It is essential to realize that a composite sketch is a drawing of a victim's or witness's perception of a perpetrator at the time he or she was observed. It is not meant to be an exact portrait of the suspect. Keep the two words "likeness" and "similarity" in mind at all times …

Unfortunately, the composite artist does not have an image of the subject in front of him or her while working. The composite artist needs to rely on the verbal description supplied by the witness. Thus, the look of a composite sketch will range from a portrait-type drawing to a caricature-type sketch, unfortunately never achieving either."), 70 ("How the forensic artist applies these witness and victim impressions and presumptions is certainly subjective.")

- 28. Forensic sketch artists report that eyewitnesses are likely to use analogies, such as a "horse face" or "bug eyes" when describing subjects. *See, e.g.* Manusci, The Police Composite Sketch, at 73, 86.

- 29. Rodger Rodriguez, *Facial Recognition: Art or Science?*, Vigilant Solutions (Apr. 4, 2016), http://www2.vigilantsolutions.com/facial-recognition-art-or-science-whitepaper. Note that Roger Rodriguez is a former detective with the NYPD, credited for helping implement the NYPD's face recognition program.
- 30. *Id.*
- 31. *FIS Presentation* (Sept. 17, 2018) (on file with author); Document pp. 020423–24, 025457.
- 32. Michelle Taylor, *The Art of Facial Recognition*, Forensic Mag. (Mar. 13, 2017), https://www.forensicmag.com/article/2017/03/art-facial-recognition. This was corroborated by Detective Tom Markiewicz in a presentation on NYPD FIS September 17, 2018. Det. Markiewicz provided the example where a photo of a suspect whose eyes were turned to the side returned no possible leads. Replacing them with eyes facing towards the camera yielded a possible match. *FIS Presentation* (Sept. 17, 2018) (on file with author) and Document p. 025463.
- 33. *FIS Presentation* (Sept. 17, 2018) (on file with author), NYPD, *Real Time Crime Center FIS Presentation: Partial Face* (Sept. 17, 2018), Document pp. 020421–22.
- 34. NYPD, *Real Time Crime Center FIS Presentation: Partial Face* (Sept. 17, 2018, Document pp. 025423, 025466 ("The goal was to create an image which highlighted the pronounced facial features of the suspect in this image. (Hairline, Forehead, Brows, and Nose). The FIS Investigator utilized the head of [redacted] in the previous case mentioned because of the similarities to the hairline and forehead. Both photos were combined within the Photoshop software and a Virtual Probe was created.").
- 35. NYPD, *Real Time Crime Center FIS Presentation: Partial Face* (date unknown), Document pp. 025469–70.
- 36. NYPD, *Real Time Crime Center FIS Presentation: Partial Face* (date unknown), Document p. 025458.
- 37. For a detailed description of 3D modeling software, see NYPD, *Animetrics User Guide* (May 6, 2017), Document pp. 018287–95 and NYPD, *DataWorks Plus FACE Plus Case Management User Guide*, Document p. 018235–39.
- 38. NYPD, *Sample case 3 of 4 – 3-Dimensional Enhancement* (date unknown), Document p. 025558.

- 39. *See, e.g.* Felix Juefei-Xu et al., *A Preliminary Investigation on the Sensitivity of COTS Face Recognition Systems to Forensic Analyst-style Face Processing for Occlusions*, IEEE Conf. on Computer Vision and Pattern Recognition Workshop 25, 31 (2015), http://openaccess.thecvf.com/content_cvpr_workshops_2015/W02/papers/Juefei-Xu_A_Preliminary_Investigation_2015_CVPR_paper.pdf (PDF). (Analysis of the results on edited faces "...questions the credibility of the FRS since the swapped in part contains biometric information of an other subject. It is questionable and surprising that the FRS uses some other biometric information to its benefit.").
- 40. Not all face recognition systems present the confidence scores of the photos in the candidate list; and of those that do, some are presented as a percentage and some are on a logarithmic or other scale. Percentages are being used here for illustrative purposes.
- 41. Latent fingerprints, fingerprints left unintentionally on surfaces and lifted for investigative purposes, may be subject to "preprocessing," editing. However, the goal of this editing is to "improve the retrievable information in a latent image while avoiding any edits that alter critical aspects of this [biometric] information." Paul Lee et al., *Forensic Latent Fingerprint Preprocessing Assessment, NISTIR 8215*, NIST, 5 (June 2018), https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8215.pdf (PDF). Improper or overuse of editing tools leads to "accidentally darkened valleys that blend together with nearby ridges, or adding false minutiae or obscuring potentially usable minutiae." *Id.*
- 42. For a discussion of the potential consequences of misconduct or error by fingerprint examiners, *see* Tom Jackman, *Orlando Fingerprint Examiner Suspended, 2,600 cases possibly affected in latest police lab scandal*, Washington Post, Feb. 27, 2017, https://www.washingtonpost.com/news/true-crime/wp/2017/02/27/orlando-fingerprint-examiner-suspended-2600-cases-possibly-affected-in-latest-police-lab-scandal/; Simon A. Cole, *Scandal, Fraud, and the Reform of Forensic Science: The Case of Fingerprint Analysis*, Cole-Monteleone (Proof), Jan. 21, 2017, available at https://wvlawreview.wvu.edu/files/d/94befc60-12bc-47d5-9e72-c8249a566415/cole-monteleone-post-page-proof.pdf (PDF).

- 43. NYPD, *Real Time Crime Center Facial Identification Section (FIS) Notifications, Chief of Detectives Memo No. 3* (Mar. 27 2012), Document pp. 017349–52. ("Real Time Crime Center Facial Identification Section (FIS) analyst determines that Subject is POSSIBLY the suspect whose image is depicted in the video and / or photograph regarding a crime. A FIS Possible Match does NOT constitute a positive identification and does NOT establish probable cause to arrest the Subject. Additional investigative steps MUST be performed in order to establish probable cause to arrest the Subject." (emphasis in original)).
- 44. NYPD, *Real Time Crime Center Facial Identification Section (FIS) Notifications, Chief of Detectives Memo No. 3* (Mar. 27, 2012), Document pp. 017349–52.
- 45. Specifics withheld given the ongoing nature of this case.

- 46. Notice of Motion to Suppress Identification Testimony filed before the Supreme Court of the State of New York, Index number withheld, on file with author. Case specifics are not provided given the ongoing nature of the case.
- 47. Willie Allen Lynch v. State of Florida, 1D16-3290.
- 48. Superior Court of the District of Columbia Criminal Division, Affidavit in Support of an Arrest Warrant, on file with author. Specifics withheld given the ongoing nature of the case.
- 49. NYPD, *Real Time Crime Center, FIS Possible Matches as of Oct. 2011–April 2017*, Document no. 018587 (2878 arrested, 549 additionally identified, 3427 total identified, 385 identification pending, 5 mis-identified, 3817 total possible matches).
- 50. Pinellas County Sheriff's Office, *Florida's Facial Recognition Network, FACES Training* (2015), Document p. 014396.

- 51. Interviews with public defenders in New York, Washington, D.C, San Francisco, Orlando, Pinellas County, and Baltimore (on file with author). *See generally* Brady v. Maryland, 373 U.S. 83 (1963). *See* Clare Garvie, Alvaro Bedoya & Jonathan Frankle, *The Perpetual Line-Up: Unregulated Police Face Recognition in America* (Oct. 18, 2016), https://www.perpetuallineup.org/findings/transparency-accountability (discussing the fact that in the 15 years the Pinellas County Sheriff's Office system has been using face recognition technology, the Public Defenders Office has never received face recognition information as part of *Brady* disclosure).
- 52. *See* Lynch v. Florida Amici Curiae brief of American Civil Liberties Union, Electronic Frontier Foundation, Georgetown Law's Center on Privacy & Technology, and Innocence Project in support of petitioner, No. SC2019-0298 (2019), *available at* https://efactssc-public.flcourts.org/casedocuments/2019/298/2019-298_notice_86166_notice2dappendix2fattachment20to20notice.pdf (PDF).

- 53. Based on records provided to us from the NYPD, we have an approximate number of the arrests made that involved some face recognition search total, but this is not disaggregated by photo editing or probe photo format. Between October 2011 and April 2017, NYPD arrested 2,878 individuals based in part on a face recognition possible match, and ran a total of 3,817 searches. *See* NYPD, *Real Time Crime Center FIS Possible Matches* (Feb. 9, 2018), Document p. 018587. In September 2018, FIS Detective Markiewicz anticipated a total of 8,000 NYPD cases to have involved a face recognition search by the end of the year. *FIS Presentation* (Sept. 17, 2018) (on file with author).
- 54. IJIS Institute National Symposium (Feb. 7, 2018) (on file with author).

- 55. Clare Garvie, Alvaro Bedoya & Jonathan Frankle, *The Perpetual Line-Up: Unregulated Police Face Recognition in America* (Oct. 18, 2016), https://www.perpetuallineup.org/recommendations.

Karen Hao

Training a single AI model can emit as much carbon as five cars in their lifetimes

Jun 6, 2019

https://www.technologyreview.com/s/613630/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Deep learning has a terrible carbon footprint. The artificial-intelligence industry is often compared to the oil industry: once mined and refined, data, like oil, can be a highly lucrative commodity. Now it seems the metaphor may extend even further. Like its fossil-fuel counterpart, the process of deep learning has an outsize environmental impact.

In a new paper, researchers at the University of Massachusetts, Amherst, performed a life cycle assessment for training several common large AI models. They found that the process can emit more than 626,000 pounds of carbon dioxide equivalent—nearly five times the lifetime emissions of the average American car (and that includes manufacture of the car itself).

It's a jarring quantification of something AI researchers have suspected for a long time. "While probably many of us have thought of this in an abstract, vague level, the figures really show the magnitude of the problem," says Carlos Gómez-Rodríguez, a computer scientist at the University of A Coruña in Spain, who was not involved in the research. "Neither I nor other researchers I've discussed them with thought the environmental impact was that substantial."

The carbon footprint of natural-language processing

The paper specifically examines the model training process for natural-language processing (NLP), the subfield of AI that focuses on teaching machines to handle human language. In the last two years, the NLP community has reached several noteworthy performance milestones in machine translation, sentence completion, and other standard benchmarking tasks. OpenAI's infamous GPT-2 model, as one example, excelled at writing convincing fake news articles.

But such advances have required training ever larger models on sprawling data sets of sentences scraped from the internet. The approach is computationally expensive—and highly energy intensive.

The researchers looked at four models in the field that have been responsible for the biggest leaps in performance: the Transformer, ELMo, BERT, and GPT-2. They trained each on a single GPU for up to a day to measure its power draw. They then used the number of training hours listed in the model's original papers to calculate the total energy consumed over the complete training process. That number was converted into pounds of carbon dioxide equivalent based on the average energy mix in the US, which closely matches the energy mix used by Amazon's AWS, the largest cloud services provider.

They found that the computational and environmental costs of training grew proportionally to model size and then exploded when additional tuning steps were used to increase the model's final accuracy. In particular, they found that a tuning process known as neural architecture search, which tries to optimize a model by incrementally tweaking a neural network's design through exhaustive trial and error, had extraordinarily high associated costs for little performance benefit. Without it, the most costly model, BERT, had a carbon footprint of roughly 1,400 pounds of carbon dioxide equivalent, close to a round-trip trans-America flight for one person.

What's more, the researchers note that the figures should only be considered as baselines. "Training a single model is the minimum amount of work you can do," says Emma Strubell, a PhD candidate at the University of Massachusetts, Amherst, and the lead author of the paper. In practice, it's much more likely that AI researchers would develop a new model from scratch or adapt an existing model to a new data set, either of which can require many more rounds of training and tuning.

To get a better handle on what the full development pipeline might look like in terms of carbon footprint, Strubell and her colleagues used a model they'd produced in a previous paper as a case study. They found that the process of building and testing a final paper-worthy model required training 4,789 models over a six-month period. Converted to CO_2 equivalent, it emitted more than 78,000 pounds and is likely representative of typical work in the field.

The significance of those figures is colossal—especially when considering the current trends in AI research. "In general, much of the latest research in AI neglects efficiency, as very large neural networks have been found to be useful for a variety of tasks, and companies and institutions that have abundant access to computational resources can leverage this to obtain a competitive advantage," Gómez-Rodríguez says. "This kind of analysis needed to be done to raise awareness about the resources being spent [...] and will spark a debate."

"What probably many of us did not comprehend is the scale of it until we saw these comparisons," echoed Siva Reddy, a postdoc at Stanford University who was not involved in the research.

The privatization of AI research

The results underscore another growing problem in AI, too: the sheer intensity of resources now required to produce paper-worthy results has made it increasingly challenging for people working in academia to continue contributing to research.

"This trend toward training huge models on tons of data is not feasible for academics—grad students especially, because we don't have the computational resources," says Strubell. "So there's an issue of equitable access between researchers in academia versus researchers in industry."

Strubell and her coauthors hope that their colleagues will heed the paper's findings and help level the playing field by investing in developing more efficient hardware and algorithms.

Reddy agrees. "Human brains can do amazing things with little power consumption," he says. "The bigger question is how can we build such machines."

THE INTERNET OF BODIES

ANDREA M. MATWYSHYN*

ABSTRACT

This Article introduces the ongoing progression of the Internet of Things (IoT) into the Internet of Bodies (IoB)—a network of human bodies whose integrity and functionality rely at least in part on the Internet and related technologies, such as artificial intelligence. IoB devices will evidence the same categories of legacy security flaws that have plagued IoT devices. However, unlike most IoT, IoB technologies will directly, physically harm human bodies—a set of harms courts, legislators, and regulators will deem worthy of legal redress. As such, IoB will herald the arrival of (some forms of) corporate software liability and a new legal and policy battle over the integrity of the human body and mind. Framing this integrity battle in light of current regulatory approaches, this Article offers a set of specific innovation-sensitive proposals to bolster corporate conduct safeguards through regulatory agency action, contract, tort, intellectual property, and secured transactions and bankruptcy.

Yet, the challenges of IoB are not purely legal in nature. The social integration of IoB will also not be seamless. As bits and bodies meld and as human flesh becomes permanently entwined with hardware,

* Associate Dean of Innovation and Professor of Law and Engineering Policy, Penn State Law (University Park); Professor of Engineering Design, Penn State Engineering; Founding Director Penn State Policy Innovation Lab of Tomorrow (PILOT); Affiliate Scholar, Center for Internet and Society, Stanford Law School; Senior Nonresidential Fellow, Cyberstatecraft Initiative, Atlantic Council. She wishes to thank: Matt Blaze, Ian Brown, Trevor Callaghan, Andrea Coravos, Jennifer Chandler, Joshua Corman, Richa Dasgupta, Lilian Edwards, Jen Ellis, Thomas Eovaldi, Mark Geistfeld, James Green, Wendy Grossman, Oona Hathaway, Chris Hoofnagle, Elizabeth Jex, Mark Lemley, Karen Levy, Christopher Marsden, Brian Martin, Terrell McSweeny, Janine Medina, Miranda Mowbray, Alexander Nally, Frank Pasquale, Stephanie Pell, Judith Rauhofer, Martin Redish, James Rule, Suzanne Schwartz, Abigail Slater, Nicolas Terry, Marcia Tiersky, and Beau Woods.

software, and algorithms, IoB will test our norms and values as a society. In particular, it will challenge notions of human autonomy and self-governance. Legal scholars have traditionally considered Kantian autonomy as the paradigmatic lens for legal determinations impacting the human body. However, IoB threatens to undermine a fundamental precondition of Kantian autonomy—Kantian heautonomy. Damaged heautonomy renders both Kantian autonomy and deliberative democracy potentially compromised. As such, this Article argues that safeguarding heautonomy should constitute the animating legal principle for governance of IoB bodies. The Article concludes by introducing the companion essay to this Article, The Internet of Latour's Things. *This companion essay inspired by the work of Bruno Latour offers a sliding scale of "technohumanity" as a framework for the legal and policy discussion of what it means to be "human" in an age where bodies are the "things" connected to the Internet.*

TABLE OF CONTENTS

INTRODUCTION

“[F]reedom of thought ... is the matrix, the indispensable condition, of nearly every other form of freedom.”
—J. Benjamin Cardozo.[1]

“This is your last chance. After this, there is no turning back. You take the blue pill—the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill—you stay in Wonderland and I show you how deep the rabbit-hole goes.... Remember ... all I’m offering is the truth. Nothing more.”
—Morpheus, *The Matrix*.[2]

We are building an “Internet of Bodies”—a hybrid society where computer code and human corpora blend and where the human body is the new technology platform. In November 2017, the Federal Drug Administration (FDA) approved the first use of a “digital pill”[3] that communicates from inside the patient’s stomach through sensors,[4] a smartphone,[5] and the Internet.[6] A year earlier, the FDA

[1] Palko v. Connecticut, 302 U.S. 319, 326-27 (1937).

[2] THE MATRIX (Warner Bros. Pictures 1999).

[3] *FDA Approves Pill with Sensor that Digitally Tracks if Patients Have Ingested Their Medication*, U.S. FOOD & DRUG ADMIN. (Nov. 13, 2017), https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm584933.htm [https://perma.cc/F2VV-RLM7]. The concept of a digital pill had been previously approved by the FDA in 2012. *See, e.g.*, Amy Maxmen, *Digital Pills Make Their Way to Market*, NATURE (July 30, 2012, 9:31 PM), http://blogs.nature.com/news/2012/07/digital-pills-make-their-way-to-market.html [https://perma.cc/FG9U-MYPF]; *see also* Peter Murray, *No More Skipping Your Medicine—FDA Approves First Digital Pill*, FORBES (Aug. 9, 2012, 11:15 AM), https://www.forbes.com/sites/singularity/2012/08/09/no-more-skipping-your-medicine-fda-approves-first-digital-pill/ [https://perma.cc/PR6T-5EKY].

[4] Sensors for monitoring body functions may be as small as 1 millimeter in size. Amelia Heathman, *This 1mm Sensor Could Monitor Your Body in Real-Time*, WIRED (Aug. 4, 2016), http://www.wired.co.uk/article/wireless-sensors-monitor-body [https://perma.cc/FUB6-SD9Q].

[5] The device transmits data to devices the patient (or a doctor) designates. Erin Kim, *‘Digital Pill’ with Chip Inside Gets FDA Green Light*, CNN MONEY (Aug. 3, 2012, 12:39 PM), https://money.cnn.com/2012/08/03/technology/startups/ingestible-sensor-proteus/ [https://perma.cc/LW6H-GGKY] (“The chip works by being imbedded into a pill.”).

[6] Robert Glatter, *Proteus Digital Health and Otsuka Seek FDA Approval for World’s First Digital Pill*, FORBES (Sept. 14, 2015, 8:09 AM), https://www.forbes.com/sites/robertglatter/2015/09/14/proteus-digital-health-and-otsuka-seek-fda-approval-for-worlds-first-digital-medicine/ [https://perma.cc/8JFD-R4GS].

approved the first artificial pancreas—a device for Type 1 diabetics that is hard-wired into patients' bodies and relies on software to calibrate insulin levels on an ongoing basis.[7]

These FDA approvals are a harbinger of the next generation of innovation, one that merges the Internet of Things[8] and artificial intelligence with the human body. This "platformization" of the body holds great promise: it is already leading to groundbreaking changes in healthcare and in lifestyle convenience.[9] However, using the human body as a platform also introduces new categories of possible harm to the confidentiality, integrity, and availability of the bodies used as part of the hardware.[10]

Three months prior to the digital pill's approval, in August 2017, the FDA issued a safety communication warning patients with a particular implanted pacemaker that they should visit their doctors immediately for a firmware[11] update.[12] The notice warned patients that a potentially serious security vulnerability in the code of their embedded medical device might enable a third-party attacker to compromise their pacemaker system and potentially physically harm them.[13] This communication marked a critical moment in the history of innovation: it was the first FDA recall of a device solely for an information security issue.[14]

[7] Susan Scutti, *'Artificial Pancreas' for Type 1 Diabetes Wins FDA Approval*, CNN (Sept. 29, 2016, 6:13 PM), https://www.cnn.com/2016/09/29/health/artificial-pancreas/index.html [https://perma.cc/C4RK-LKHD].

[8] U.S. FED. TRADE COMM'N, INTERNET OF THINGS: PRIVACY & SECURITY IN A CONNECTED WORLD 1-2 (2015), [hereinafter U.S. FED. TRADE COMM'N, INTERNET OF THINGS] https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf [https://perma.cc/W5DL-A4FT] (describing the Internet of Things as the totality of consumer and other devices that connect to the Internet).

[9] *See* Glatter, *supra* note 6.

[10] *Id.*

[11] Firmware is computer code built into a piece of hardware. Margaret Rouse, *Definition: Firmware*, WHATIS.COM (Apr. 2017), https://whatis.techtarget.com/definition/firmware [https://perma.cc/VL6M-R7KB].

[12] *Firmware Update to Address Cybersecurity Vulnerabilities Identified in Abbott's (Formerly St. Jude Medical's) Implantable Cardiac Pacemakers: FDA Safety Communication*, U.S. FOOD & DRUG ADMIN. (Aug. 29, 2017), https://www.fda.gov/medicaldevices/safety/alertsandnotices/ucm573669.htm [https://perma.cc/8LB5-JHJ7].

[13] *Id.*

[14] *See* Evan Sweeney, *FDA Announces Firmware Update to Resolve Cybersecurity Vulnerabilities in Abbott Pacemakers*, FIERCEHEALTHCARE (Aug. 30, 2017, 10:15 AM), https://www.fiercehealthcare.com/privacy-security/fda-rolls-out-firmware-update-to-resolve-

The August 2017 pacemaker security recall was not, however, the first time that computer code put human bodies at risk of physical harm and death.[15] Indeed, a year prior, a patient's heart surgery had been unexpectedly interrupted[16] for five minutes[17] when one of the Internet-enabled machines attached to the patient's body crashed.[18] The machine had unexpectedly performed an anti-malware scan in the middle of the operation[19] and locked up the human interface—the interface upon which the surgeons were relying to keep the patient alive.[20]

This creeping merger of bodies with bits and bytes is also not limited to medical contexts. Employers are throwing "chip[ping] part[ies],"[21] embedding their employees' bodies with chips[22] that

cybersecurity-vulnerabilities-abbott [https://perma.cc/K8UZ-X83P]; *see also* Richard Staynings, *FDA Announces First-Ever Recall of a Medical Device Due to Cyber Risk*, CISCO BLOG (Aug. 30, 2017), https://blogs.cisco.com/healthcare/fda-announces-first-ever-recall-of-a-medical-device-due-to-cyber-risk [https://perma.cc/PSC8-P59R].

[15] *See, e.g.*, Anne Marie Porrello, Death and Denial: The Failure of the THERAC-25, A Medical Linear Accelerator (unpublished computer science paper) (on file with California Polytechnic State University), http://users.csc.calpoly.edu/~jdalbey/SWE/Papers/THERAC25.html [https://perma.cc/5X8L-4ZPN] (chronicling death or severe radiation injury to patients due to software malfunction).

[16] Dan Goodin, *That Time a Patient's Heart Procedure Was Interrupted by a Virus Scan*, ARS TECHNICA (May 16, 2016, 1:58 PM), https://arstechnica.com/information-technology/2016/05/faulty-av-scan-disrupts-patients-heart-procedure-when-monitor-goes-black/ [https://perma.cc/9HE3-D38U].

[17] > [I]n the middle of a heart catheterization procedure, the hemo monitor pc lost communication with the hemo client and the hemo monitor went black. Information obtained from the customer indicated that there was a delay of about 5 minutes while the patient was sedated so that the application could be rebooted. It was found that anti-malware software was performing hourly scans.

MAUDE Adverse Event Report: Merge Healthcare Merge Hemo Programmable Diagnostic Computer, U.S. FOOD & DRUG ADMIN. (Feb. 8, 2016) [hereinafter *MAUDE Adverse Event Report*], https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/detail.cfm?mdrfoi__id=5487204 [https://perma.cc/LQV5-UJPE].

[18] Jacob Brogan, *An Antivirus Scan Shut Down a Medical Device in the Middle of Heart Surgery*, SLATE (May 5, 2016, 4:34 PM), https://slate.com/technology/2016/05/antivirus-scan-shuts-down-merge-hemo-medical-device-during-heart-surgery.html [https://perma.cc/G9VM-VCRB].

[19] Fortunately, the heavily sedated patient survived the operation, but this outcome was not guaranteed. *Id.*

[20] In its FDA incident report, the manufacturer of the equipment blamed the hospital technicians for a misconfiguration, stating that prominent disclaimers existed with the accompanying materials. *MAUDE Adverse Event Report*, *supra* note 17.

[21] Jeff Baenen, *Wisconsin Company Holds 'Chip Party' to Microchip Workers*, CHI. TRIB. (Aug. 2, 2017, 7:32 AM), https://www.chicagotribune.com/bluesky/technology/ct-wisconsin-

connect with other devices[23] and transmit information[24] from employees' bodies.[25] Trucking companies sometimes now expect their drivers to wear clothing or devices that monitor location and alertness[26] and (ostensibly) "improve"[27] job performance.[28] Manufacturers

company-microchips-workers-20170801-story.html [https://perma.cc/3ARQ-L5RY]; *see* James Brooks, *A Swedish Start-Up Has Started Implanting Microchips into Its Employees*, CNBC (Apr. 3, 2017, 12:02 PM), https://www.cnbc.com/2017/04/03/start-up-epicenter-implants-employees-with-microchips.html [https://perma.cc/Z4U2-NJ4C]; Rory Cellan-Jones, *Office Puts Chips Under Staff's Skin*, BBC NEWS (Jan. 29, 2015), https://www.bbc.com/news/technology-31042477 [https://perma.cc/PD8B-M8H9]; Trent Gillies, *Why Most of Three Square Market's Employees Jumped at the Chance to Wear a Microchip*, CNBC (Aug. 13, 2017, 9:00 AM), https://www.cnbc.com/2017/08/11/three-square-market-ceo-explains-its-employee-microchip-implant.html [https://perma.cc/G74T-QC2R]; *Wisconsin Company Three Square Market to Microchip Employees*, BBC NEWS (July 24, 2017), https://www.bbc.com/news/world-us-canada-40710051 [https://perma.cc/UUE7-G8NQ].

[22] Experts expect this practice to become a norm in future employment. Chris Morris, *Wisconsin Company Holds Party to Implant Workers with Microchips*, FORTUNE (Aug. 2, 2017), http://fortune.com/2017/08/02/wisconsin-company-holds-party-to-implant-workers-with-microchips/ [https://perma.cc/5BVF-VRCZ]. ("Noelle Chesley, an associate professor of sociology at the University of Wisconsin-Milwaukee, tells the Chicago Tribune she expects implanting microchips into employees will become the norm in years to come."). Some employees harbor reservations about the chips. Steven Melendez, *Why Would Anyone Let Their Employer Stick a Microchip into Their Body?*, FAST CO. (July 25, 2017), https://www.fastcompany.com/40444110/why-would-anyone-let-their-employer-stick-a-microchip-into-their-body [https://perma.cc/AJ3H-86C2].

[23] Danielle Paquette, *Some Feared Hackers and the Devil. Others Got Microchipped.*, WASH. POST (Aug. 1, 2017), https://www.washingtonpost.com/news/wonk/wp/2017/08/01/some-feared-hackers-and-the-devil-others-got-microchipped/ [https://perma.cc/H5CB-FKDH].

[24] Although current chips generally do not transmit location, the capability is expected in the future. Gillies, *supra* note 21 ("A future version of the microchip could include GPS, and if an employee leaves the company, it won't be removed.").

[25] *Microchipping at Work: US Employees Get Voluntarily Implanted at Staff 'Chip Party,'* ABC NEWS (AUSTL. BROAD. CORP.) (Aug. 1, 2017, 8:54 PM), http://www.abc.net.au/news/2017-08-02/microchip-workers-hold-chip-party/8765934 [https://perma.cc/9PF8-V3QA].

[26] *See* Olivia Solon, *Eye-Tracking System Monitors Driver Fatigue, Prevents Sleeping at Wheel*, WIRED (May 28, 2013), https://www.wired.co.uk/article/eye-tracking-mining-system [https://perma.cc/6WZJ-K56N].

[27] *See* Tim Collins, *The Life-Saving £180 Bracelet that Gives Tired Drivers an Electric Shock if They Begin to Fall Asleep at the Wheel,* DAILY MAIL (July 31, 2017, 8:31 AM), https://www.dailymail.co.uk/sciencetech/article-4746076/Steer-delivers-shocks-drivers-fall-asleep.html [https://perma.cc/U2KL-29ZA].

[28] *See How the Internet of Things Is Transforming Construction*, WHITELIGHT GROUP (Aug. 18, 2014), https://whitelightgrp.com/2014/08/18/internet-things-transforming-construction/ [https://perma.cc/PC4C-G5ZM]. For a discussion of body-attached truck-driving devices, see, for example, Karen Levy, *After the Tornado*, YOUTUBE (Nov. 19, 2017, at 5:27), https://www.youtube.com/watch?time_continue=18&v=6kPjsfYSzp4 [https://perma.cc/346M-KMDN].

of "brain sensing"[29] Internet-enabled headbands[30] encourage "professionals" to use the device to monitor a "client's"[31] brain sensations[32] in real time.[33] Simultaneously, these same companies might encourage consumers to use the headbands[34] to facilitate "meditation,"[35] and developers to build out games and other applications incorporating brain data.[36] Other brain sensing headbands are appearing in classrooms, signaling to teachers and remote parents when children are (allegedly) paying attention in class.[37] Meanwhile, consumers are donning augmented reality devices in

[29] The creators of this product describe it as a type of "brain-computer interface[ ]." *See Muse: The Brain Sensing Headband Changing the Way the World Thinks*, INDIEGOGO (Apr. 24, 2014), https://www.indiegogo.com/projects/muse-the-brain-sensing-headband#/ [https://perma.cc/P6NR-73X2].

[30] *See Technology Enhanced Meditation*, CHOOSE MUSE, http://www.choosemuse.com/ [https://perma.cc/UZC8-C4Q9].

[31] The website of the company in question alternates between using the word "patient" and "client." *What Is Muse Connect and What Are the Benefits of Using It?*, MUSE (Oct. 28, 2018), https://choosemuse.force.com/s/article/What-are-the-benefits-of-using-Muse-Connect [https://perma.cc/W5N4-JEPV] ("Monitor patient progress and improve patient outcomes.").

[32] Specifically, the headband in question monitors "real-time brainwave information to measure states of focus, relaxation, and mind-wandering." *MUSE—The Head Sensing Headband*, ACUPUNCTURE TRADITIONAL CHINESE MED., https://www.acupunctureclinic.ie/wellness-online-store/ [https://perma.cc/S9NQ-BX67].

[33] *See Join the Muse Professional Community*, MUSE PROFESSIONAL, https://choosemuse.com/muse-professionals/ [https://perma.cc/8J7F-K685] ("A personalized dashboard tracks your clients' at-home meditation practice with Muse, so you can view their progress in real time.").

[34] Some IoB helmets also promise to stimulate neurons. Madhumita Venkataramanan, *Neuroelectrics' Wireless Brain Helmet Can Electrically Stimulate Your Neurons*, WIRED (May 4, 2015), http://www.wired.co.uk/article/stimulation-station [https://perma.cc/9U4W-FK5G].

[35] *Muse: The Brain Sensing Headband*, AMAZON, https://www.amazon.com/muse-brain-sensing-headband-black/DP/B00LOQR37C [https://perma.cc/9AAC-GDRK].

[36] *See Muse Developer*, MUSE, http://www.choosemuse.com/developer [https://perma.cc/PR6F-4QMS] ("Receive raw EEG, accelerometer, gyroscope, and battery data [;] [l]everage built-in algorithms for band powers, eye blinks, and jaw clenches.").

[37] *Under AI's Watchful Eye, China Wants to Raise Smarter Students*, WALL ST. J. (Sept. 19, 2019, 5:30 AM), https://www.wsj.com/video/under-ais-watchful-eye-china-wants-to-raise-smarter-students/C4294BAB-A76B-4569-8D09-32E9F2B62D19.html [https://perma.cc/US5T-EAUB].

gaming,[38] and they are purchasing clothes[39] and accessories[40] that connect their bodies to the Internet, sharing corporeal information about themselves in real time.[41] Some consumers are even recreationally implanting chips into their bodies for the sake of convenience,[42] allowing their bodies to perform some of the tasks their phones do now.[43] In short, we are experiencing a creeping transformation where human bodies themselves are becoming connected to and sometimes reliant upon software, hardware, and the Internet for portions of their "default" functionality. This is the Internet of Bodies.

In addition to transforming individual bodies,[44] these Internet of Bodies devices also introduce a new level of peril for society in the aggregate. For the first time in our civilization, computer code will be able to physically damage (civilian) human bodies at scale. In other words, particularly as artificial intelligence becomes incorporated into the Internet of Bodies, the confidentiality, integrity, and availability of some human bodies will inevitably become compromised due to flawed and vulnerable software, either individually or *en masse*: the security compromises that plague our networks, devices, and databases today will shift inside (and physically damage) the human body tomorrow. Yet, the law is currently unprepared to

[38] *See* Jacob Kleinman, *Augmented Reality Glasses: What You Can Buy Now (or Soon)*, TOM'S GUIDE (Feb. 14, 2018, 8:00 AM), https://www.tomsguide.com/us/best-ar-glasses,review-2804.html [https://perma.cc/E2VU-Y9DT].

[39] *See* Michael Sawh, *The Best Smart Clothing: From Biometric Shirts to Contactless Payment Jackets*, WAREABLE (Apr. 16, 2018), https://www.wareable.com/smart-clothing/best-smart-clothing [https://perma.cc/T5ZC-H6EQ].

[40] *See* Michael Sawh, *Put a Ring on It: The Best Smart Rings*, WAREABLE (Jan. 28, 2019), https://www.wareable.com/fashion/best-smart-rings-1340 [https://perma.cc/M2DT-ED2L].

[41] *See* Ananya Bhattacharya, *Bluetooth-Enabled Vibrating Hotpants Are the Dumbest Smart Things at CES 2017*, QUARTZ (Jan. 6, 2017), https://qz.com/878137/bluetooth-enabled-vibrating-hotpants-are-the-dumbest-smart-things-at-ces-2017/ [https://perma.cc/Y2LW-XJX8].

[42] Chips have been used with animal identification for over a decade. Morris, *supra* note 22.

[43] Jefferson Graham, *Who Wants to Get 'Chipped'?*, USA TODAY (Aug. 1, 2017, 12:28 PM), https://www.usatoday.com/story/tech/talkingtech/2017/07/29/wa/520034001/ [https://perma.cc/Q7CH-6LZQ].

[44] For example, the first Cyborg Olympics recently unveiled some of the innovation in progress in IoB technology. Bloomberg (@business), TWITTER (Nov. 17, 2016, 4:50 PM), https://twitter.com/business/status/799414438675632128 [https://perma.cc/C4Y5-9AZ7] ("Welcome to the first cyborg Olympics.").

address these harms and the social transformation that the Internet of Bodies will occasion.

This Article introduces and explains this (already happening) progression of the Internet of Things or "IoT" into the Internet of Bodies or "IoB."[45] As the "meatware"[46] of human bodies blends with software, hardware, and related technologies[47] in the Internet of Bodies era, jurists, legislators, and scholars will be faced with a dual

[45] This author first defined the term Internet of Bodies (IoB) in a legal and policy context in 2016. *See* Andrea Matwyshyn, Northeastern/Princeton/Stanford, The Internet of Bodies, 9th Annual Privacy Law Scholars Conference for Berkeley Center for Law & Technology (June 2, 2016), https://www.law.berkeley.edu/research/bclt/past-events/2016-con ferences/june-2016-the-9th-annual-privacy-law-scholars-conference/program/ [https://perma.cc/YDE3-RM2U]; *see also* Wendy M. Grossman, *Dinosaur Bones*, NET.WARS (June 10, 2016, 6:56 PM), https://www.pelicancrossing.net/netwars/2016/06/dinosaur_bones.html [https://perma.cc/5NZW-FZFL]. Since then, the term has gained resonance with legal and policy audiences. *See Computers, Privacy & Data Protection 2018: The Internet of Bodies*, CPDP2018, https://web.archive.org/web/20180408073819/http://www.cpdpconferences.org/index.html [https://perma.cc/Y2R8-4VMS]. The notion of an "Internet of Bodies" appeared previously on a limited basis in the technology press and forums but without a clear definition or application to legal and policy contexts. *See, e.g.*, Pedro Domingos*, Shall We Have Internet of Bodies (IoB) Similar to Internet of Things (IoT)?*, QUORA, https://www.quora.com/Pedro-Domingos-Shall-we-have-Internet-of-Bodies-IoB-similar-to-Internet-of-Things-IoT [https://perma.cc/MLR9-3XN8]; *Internet of Bodies*, TUMBLR (Feb. 1, 2016) http://internet-of-bodies.tumblr.com/ [https://perma.cc/H2EM-FNUR]; Meghan Neal, *The Internet of Bodies Is Coming, and You Could Get Hacked*, VICE: MOTHERBOARD (Mar. 13, 2014, 2:20 PM), https://motherboard.vice.com/en_us/article/gvyqgm/the-internet-of-bodies-is-coming-and-you-could-get-hacked [https://perma.cc/6TVQ-7D7R]; Avi Zins, *Internet of Bodies (IOB)*, SAÚDE ONLINE, https://saudeonline.grupo midia.com/blog/internet-of-bodies-iob-por-avi-zins [https://perma.cc/7SNP-4XP2]. The term "Internet of Bodies" has also been used by creative professional Ghislaine Boddington in a recent talk about the body as a digital canvas. *See Ghislaine Boddington's TEDx Talk*, WOMEN SHIFT DIGITAL, http://www.womenshiftdigital.com/ghislaines-tedx-talk-on-video/ [https://perma.cc/2U2A-XUWZ]. For work on the Internet of Bodies from other disciplines, see CARLO RATTI & MATTHEW CLAUDEL, THE CITY OF TOMORROW 85-87 (2016) (discussing urban planning); *Ghislaine Boddington*, BODY>DATA>SPACE, http://www.bodydataspace.net/who-we-are/core-team/ghislaine/ [https://perma.cc/LRA9-ZNDQ]. The term has also appeared elsewhere in the context of wearable clothing. *See* Erin Lewis, *Pechakucha Vol. 21 Erin Lewis—Internet of Bodies*, YOUTUBE (July 4, 2013), https://www.youtube.com/watch?v=6OeePCEumUw [https://perma.cc/P75L-M98F].

[46] *See Meatware*, URB. DICTIONARY, https://www.urbandictionary.com/define.php?term=meatware [https://perma.cc/37Q4-PWK4].

[47] In particular, machine learning algorithms and "artificial intelligence" become increasingly common as part of the functionality of Internet of Bodies devices. All of the concerns regarding security articulated in this article extend to the machine learning components of IoB devices. Additionally, machine learning introduces a series of other code integrity risks depending on the nature of its functionality. These issues are explored in detail in Andrea M. Matwyshyn, Artifice and Intelligence (unpublished manuscript) (on file with author).

IoB legal challenge. First, they will need to address the unresolved policy and legal quandaries presented by the Internet of Things. Second, they will face a formidable challenge in addressing what a programmer might call the legal "legacy code"[48] problem of software liability more broadly. Just as companies struggle to address the "technical debt"[49] of their systems, the law now faces a somewhat parallel "legal technical debt" challenge. Multiple traditional bodies of law have failed to meaningfully update themselves across time to effectively address changing technology circumstances. As a consequence, resolving this "legal technical debt" will be doctrinally buggy as courts and regulators seek to redress and mitigate bodily harms caused by computer code: crafting suitable methods of redress for both physical and economic IoB harms will implicate a series of sometimes conflicting policy concerns.

Part I introduces the progression of IoT into IoB. Explaining three discrete generations of IoB—body external, body internal, and body melded—Part I locates our current social reality in this progression at Stage 2—body internal. Yet, using patent filings to reveal expected innovation, Part I argues that late second-generation body internal and early third-generation body melded technologies are already being actively developed. Next, Part I articulates four legacy problems of IoT that will impact the nature of future harms caused by IoB—the "better with bacon" problem of gratuitous Internet reliance and connection, the "builder bias" problem of extreme levels of known (but uncorrected) security vulnerability, the "magic gadget" problem of failing to anticipate failure, and the "mandatory soup" problem of diminishing consumer options for self-help. Part I then presents five areas of law where conflicts over IoB will be most pronounced—guidance from regulatory agencies, contracts, tort, intellectual property, and secured transactions and bankruptcy. Finally, Part I offers concrete approaches for building short term innovation-sensitive legal structures of IoB consumer protection.

Part II then expands on the critical difference between IoB and IoT: IoB's propensity to physically damage human bodies and

[48] *See infra* Part I.B.

[49] *See* Ward Cunningham, *Debt Metaphor*, YOUTUBE (Feb. 14, 2009), https://www.youtube.com/watch?v=pqeJFYwnkjE [https://perma.cc/SFL9-MQ9X].

minds. IoB presents the specter not only of negative consequences with respect to physical and psychological autonomy— in a Kantian sense—but also, even more fundamentally, third-generation IoB threatens to potentially erode Kantian *heautonomy*—the necessary *precursor* to autonomy. For these reasons, Part II argues that the touchstone for all regulation of IoB must be the safeguarding of heautonomy. Part II concludes by asking an uncomfortable theoretical question about our underlying assumptions regarding the human body: should the law assume the body to be a "bug" or a "feature"? The companion essay to this article, *The Internet of Latour's Things*, grapples with the question of whether future law will view the corporeality of the human body as worthy of preservation (or elimination) in a society full of IoB bodies. Part III concludes.

I. THE INTERNET OF (HUMAN) THINGS: DEFINING THE "INTERNET OF BODIES"

Morpheus: The Matrix is everywhere. It is all around us. Even now, in this very room. You can see it when you look out your window or when you turn on your television. You can feel it when you go to work ... when you go to church ... when you pay your taxes.[50]

In the 1999 movie *The Matrix*, a computer programmer named Thomas A. Anderson, who uses the handle "Neo," finds out that the physical reality he experiences is actually a computer-generated illusion.[51] After taking a mysterious red pill, he discovers that underneath the superficially-placid exterior of the world he inhabits, there lurks a linked invisible society of machine overlords.[52] The machines are powered by energy extrusions from millions of human bodies that have been physically networked together.[53] This web of bodies—the Matrix—allows the machine overlords to harness and commodify the bodies of humans, turning them into merely the

[50] THE MATRIX, *supra* note 2.
[51] *Id.*
[52] *Id.*
[53] *Id.*

"hardware" that powers both the machines and the software that perpetuates the simulacrum of the human-viewable world.[54]

The Matrix is, of course, just a movie; a majority of scientists do not believe that the world we currently inhabit is merely an illusion generated by a computer program.[55] However, we are unquestionably entering a technological age where the line between the human body and the machine is beginning to blur.[56] Many human bodies will soon become at least occasionally reliant on the Internet for some aspect of their functionality,[57] and the energy of the human body is already being used experimentally to mine cryptocurrency.[58] Just as the Internet of Things has networked our possessions into a "cloud"[59] of shared gadgetry, so too our bodies are slowly becoming networked into an "Internet of Bodies."[60]

[54] *Id.*

[55] *But see* Andrew Zimmerman Jones, *Are We Living in a Computer Simulation?*, PBS (July 8, 2015), https://www.pbs.org/wgbh/nova/article/are-we-living-in-a-computer-simulation/ [https://perma.cc/87RJ-TLUW]; Clara Moskowitz, *Are We Living in a Computer Simulation?*, SCI. AM. (Apr. 7, 2016), https://www.scientificamerican.com/article/are-we-living-in-a-computer-simulation/ [https://perma.cc/NX7M-YGJ7].

[56] *See, e.g.*, Nathan Hurst, *This Digital Prosthesis Could Help Amputees Control Computers,* SMITHSONIAN.COM (Dec. 13, 2016), https://www.smithsonianmag.com/innovation/digital-prosthetic-could-help-amputees-control-computers-180961397/ [https://perma.cc/9R36-P3GA].

[57] *What Is the Pancreas? What Is an Artificial Pancreas Device System?*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/medicaldevices/productsandmedicalprocedures/homehealthandconsumer/consumerproducts/artificialpancreas/ucm259548.htm [https://perma.cc/6PCL-PJSS].

[58] As the Institute for Human Obsolescence has described it: "A single human body at rest radiates 100 watts of excess heat.... The electricity generated is then fed to a computer that produces cryptocurrency." *Biological Labour*, INSTITUTE OF HUMAN OBSOLESCENCE, http://speculative.capital/ [https://perma.cc/Q9SA-7459]; *see also* Daniel Oberhaus, *You Could Mine 1 Bitcoin Per Month If You Harvested the Body Heat from 44,000 People*, VICE: MOTHERBOARD (Jan. 3, 2018, 10:00 AM), https://motherboard.vice.com/en_us/article/vby7ny/bitcoin-body-heat-mining [https://perma.cc/T8SH-57D2].

[59] For a discussion of "the cloud," see, for example, Margot E. Kaminski, *Robots in the Home: What Will We Have Agreed to?*, 51 IDAHO L. REV. 661, 670 (2015). Most robots will share information with third parties for processing purposes or just to store information in the cloud. *Id.*

[60] For examples of devices in the Internet of Bodies, see *infra* notes 88-93 and accompanying text.

A. Three Generations of IoB

Morpheus: The pill you took is part of a trace program. It's designed to disrupt your input/output carrier signal so we can pinpoint your location.
Neo: What does that mean?
Cypher: It means fasten your seat belt Dorothy, 'cause Kansas is going bye-bye.[61]

In an iconic 1960 episode of *The Twilight Zone*, a misanthropic writer becomes convinced that the appliances in his home are conspiring against him, attempting to intimidate him.[62] His escalating tensions with the machines culminate in his typewriter, television, and telephone informing him that he needs to leave and in his electric shaver menacing him.[63] Ultimately, his car "encourages" his untimely exit.[64]

Despite recent reports of home smart assistants laughing maniacally and scaring their owners,[65] today's Internet of Things—meaning the totality of consumer and other devices that connect to the Internet[66]—usually reflects a less menacing version of The *Twilight Zone*'s sentient appliances.[67] According to some estimates, the number of IoT devices is expected to reach 21 billion devices by the year 2020.[68] These devices include everything from toys[69] to

[61] THE MATRIX, *supra* note 2.

[62] *The Twilight Zone: A Thing About Machines* (Cayuga Productions, CBS Television Network, Oct. 28, 1960).

[63] *Id.*

[64] *Id.*

[65] *See* Christina Bonnington, *Alexa Is Creepily Laughing at People for No Reason*, SLATE (Mar. 7, 2018, 6:28 PM), https://slate.com/technology/2018/03/amazons-alexa-is-creepily-laughing-for-no-reason-its-just-the-start.html [https://perma.cc/9F9X-YRZM].

[66] *See* U.S. FED. TRADE COMM'N, INTERNET OF THINGS, *supra* note 8, at 5-6 (summarizing the findings of a workshop held earlier in the year on the topic).

[67] However, a recent first-person account of a technology journalist chronicled her begging her home IoT devices to make her a cup of coffee, and a later "emotional" overreaction from her coffee machine due to her absence. *See* Kashmir Hill & Surya Mattu, *The House that Spied on Me,* GIZMODO (Feb. 7, 2018, 1:25 PM), https://gizmodo.com/the-house-that-spied-on-me-1822429852 [https://perma.cc/5N8S-34SN].

[68] Nathan Eddy, *Gartner: 21 Billion IoT Devices to Invade by 2020*, INFO. WEEK (Nov. 10, 2015, 11:05 AM), https://www.informationweek.com/mobile/mobile-devices/gartner-21-billion-iot-devices-to-invade-by-2020/d/d-id/1323081 [https://perma.cc/FL9A-XWAR].

[69] *See Electronic Toy Maker Vtech Settles FTC Allegations that It Violated Children's*

toasters[70] to cars[71] to hospital respirators[72] to industrial control systems.[73]

According to a recent Federal Trade Commission (FTC) report, our society is merely "at the beginning of this [IoT] technology trend."[74] While asserting that IoT devices potentially offer substantial benefit to consumers in connected medicine and other contexts, the FTC report highlighted the concerning reality that our existing legal paradigms are not optimally suited for the Internet of Things context.[75] In particular, the FTC explained that IoT has created challenges for meaningful consumer consent, privacy, and security.[76]

In a consonant vein, Professor Scott Peppet has argued that in the Internet of Things "the near impossibility of truly de-identifying ... data, the likelihood that Internet of Things devices will be inherently prone to security flaws, and the difficulty of meaningful consumer consent in this context—create very real discrimination,[77]

Privacy Law and the FTC Act, U.S. FED. TRADE COMM'N (Jan. 8, 2018), https://www.ftc.gov/news-events/press-releases/2018/01/electronic-toy-maker-vtech-settles-ftc-allegations-it-violated [https://perma.cc/E3V7-KBYV].

[70] *See* Joel Hruska, *The Internet of Things Has Officially Hit Peak Stupid, Courtesy of This Smart Toaster*, EXTREME TECH (Jan. 5, 2017, 4:11 PM), https://www.extremetech.com/electronics/242169-internet-things-officially-hit-peak-stupid-courtesy-smart-toaster-griffin-technology [https://perma.cc/5PX3-N4QR].

[71] *See* Jonny Evans, *Just Say No to Connected Cars*, COMPUTERWORLD (July 8, 2015, 10:25 AM), https://www.computerworld.com/article/2945367/just-say-no-to-connected-cars.html [https://perma.cc/BC37-QLSJ].

[72] *Philips Hospital Respiratory Care*, PHILIPS, https://www.usa.philips.com/healthcare/solutions/hospital-respiratory-care [https://perma.cc/D75M-Y52S].

[73] *See Internet of Things and Industrial Control Systems*, U.K. CTR. FOR THE PROT. OF NAT'L INFRASTRUCTURE, https://www.cpni.gov.uk/internet-things-and-industrial-control-systems [https://perma.cc/QR9E-KQ8B].

[74] *See* U.S. FED. TRADE COMM'N, INTERNET OF THINGS, *supra* note 8, at i.

[75] "Staff acknowledges the practical difficulty of providing choice when there is no consumer interface and recognizes that there is no one-size-fits-all approach." *Id.* at v.

[76] Professor Peppet also highlighted the problems of consent. *See* Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent*, 93 TEX. L. REV. 85, 140-46 (2014).

[77] Professor Peppet provides the example of an Internet of Things breathalyzer explaining: "the consumer is essentially led to the incorrect assumption that this small black device is merely a good like any other—akin to a stapler or ballpoint pen—rather than a data source and cloud-based data repository." *Id.* at 90.

privacy,[78] security,[79] and consent[80] problems."[81] Other scholars have focused on behavioral impacts occasioned by a world permeated by the Internet of Things. For example, Professor Meg Leta Jones has asserted that the goal of the Internet of Things, which she dubs the "Internet of Other Peoples' Things,"[82] is to enable "ubiquitous connection"[83] and that "[p]erforming the boundary work necessary to managing one's information becomes increasingly difficult as we move deeper into the Information Age."[84] Professor Paul Ohm and Blake Reid have asked what it means to regulate software when everything around us contains software.[85] Meanwhile, Professor Christina Mulligan has argued that as software becomes increasingly present in consumer goods, Internet of Things merchants will use the licenses to the software contained in these devices to undesirably and materially, contractually restrict both the permitted uses

[78] Peppet, for example, argues in favor of data minimization and use constraints: "As a first regulatory step, we should constrain certain uses of Internet of Things data if such uses threaten consumer expectations." *Id.* at 150.

[79] "Internet of Things Devices May Be Inherently Prone to Security Flaws," argues Peppet. *Id.* at 133-36.

[80] Peppet explains the consent problem as follows:

> The technical problem is simple: coupled with Big Data or machine learning analysis, massive amounts of sensor data from Internet of Things devices can give rise to unexpected inferences about individual consumers. Employers, insurers, lenders, and others may then make economically important decisions based on those inferences, without consumers or regulators having much understanding of that process. This could lead to new forms of illegal discrimination.

Id. at 118.

[81] *Id.* at 85. Peppet advocates four approaches to regulating the Internet of Things:

> (1) broadening existing use constraints—such as some state law on automobile EDRs—to dampen discrimination; (2) redefining "personally identifiable information" to include biometric and other forms of sensor data; (3) protecting security by expanding state data-breach notification laws to include security violations related to the Internet of Things; and (4) improving consent by providing guidance on how notice and choice should function in the context of the Internet of Things.

Id. at 149.

[82] Meg Leta Jones, *Privacy Without Screens & the Internet of Other People's Things*, 51 IDAHO L. REV. 639, 639 (2015).

[83] *Id.* at 641.

[84] *Id.* at 645.

[85] Paul Ohm & Blake Reid, *Regulating Software When Everything Has Software*, 84 GEO. WASH. L. REV. 1672, 1673-74 (2016).

and resale or transfer of devices.[86] Finally, Professor Irina Manta and David Olson have argued that while Internet of Things "[p]rice discrimination can increase total market welfare in some cases, especially in comparison to monopoly pricing; it also can decrease total market welfare if the pricing is done in such a way as to decrease allocative efficiency."[87]

Building on this prior scholarship, this Article asks what it means for existing legal paradigms and for the next generation of innovation when the "things" that are attached to the Internet are human bodies. In brief, this Article argues that this "Internet of human things" or, more succinctly, this "Internet of Bodies" will cause us to materially reframe our legal conversations when computer code regularly begins to cause *physical* harms to human bodies. But before embarking on this legal analysis, let us define the Internet of Bodies and assess how it mirrors and differs from the Internet of Things.

The Internet of Bodies might be divided into three generations of technologies—body external, body internal, and body melded.[88]

1. First-Generation IoB: Body External

The first generation of IoB devices has already become a familiar fixture in our lives. These devices are seemingly ubiquitous, including everything from "lifestyle" connected fitness tracking devices[89] and "smart" glasses[90] to "smart" exoskeletons,[91] connected breast pumps,[92] and brain-sensing[93] headbands.[94] Specifically, these

[86] Christina Mulligan, *Personal Property Servitudes on the Internet of Things*, 50 GA. L. REV. 1121, 1122-24 (2016).

[87] Irina D. Manta & David S. Olson, *Hello Barbie: First They Will Monitor You, Then They Will Discriminate Against You. Perfectly.*, 67 ALA. L. REV. 135, 157 (2015).

[88] *See infra* Parts I.A.1-3.

[89] *See, e.g.*, FITBIT, https://www.fitbit.com/home [https://perma.cc/PF46-EWL6].

[90] *See* Daniel Van Boom, *China's Police Get Face-Recognizing Glasses Ahead of New Year*, CNET (Feb. 7, 2018, 8:05 PM), https://www.cnet.com/news/china-new-year-police-glasses-ai-cctv [https://perma.cc/DGE3-58DD].

[91] *See* Timothy Burke, *Paraplegic in Robotic Exoskeleton Performs World Cup First Kick*, DEADSPIN (June 12, 2014, 3:19 PM), https://deadspin.com/paraplegic-in-robotic-exoskeleton-performs-world-cup-fi-1590050190 [https://perma.cc/RN2M-YLUH].

[92] *See* Zoe Kleinman, *CES 2018: Willow and Freemie Breast Pumps Offer Mums Freedom*, BBC NEWS (Jan. 11, 2018), http://www.bbc.com/news/technology-42643971 [https://perma.cc/E3EM-9VJH]. For example, Willow, a connected breast pump, syncs with the Willow

first-generation IoB devices encompass three categories of body-external products—IoB "medical devices" approved by the FDA,[95] "general wellness"[96] IoB devices that present "low risk" and promote "healthy lifestyle" (and are, therefore, not regulated by the FDA),[97] and various other non-health enterprise, educational, and recreations body-attached devices that connect to the Internet, directly or indirectly.[98]

First-generation IoB medical devices include devices such as Internet-enabled robotic surgery machines[99] and connected prosthetics[100] that a patient operates from a mobile phone.[101] In comparison, the "general wellness/lifestyle" first-generation IoB category encompasses familiar devices such as fitness trackers,[102] health

app. *See Frequently Asked Questions*, WILLOW, https://www.willowpump.com/faq/ [https://perma. cc/UC3S-NE2M].

[93] These headbands include headbands for patients lacking motor function. *See* Mark Honigsbaum, *Could This $300 Headset Transform the Lives of 'Locked-In' Patients?*, GUARDIAN (July 11, 2014, 6:00 AM), https://www.theguardian.com/technology/2014/jul/11/kickstarter-headset-locked-in-syndrome-communication [https://perma.cc/MAB8-G7UN].

[94] IoB headbands also allow gamers to race drones with their minds. *See* Anthony Cuthbertson, *Watch: World's First Mind-Controlled Drone Race*, NEWSWEEK (Apr. 25, 2016, 8:50 AM), http://www.newsweek.com/watch-worlds-first-mind-controlled-drone-race-451965 [https://perma.cc/RCU3-CK9E].

[95] *See infra* notes 98-100 and accompanying text.

[96] The FDA "defines general wellness products as products that meet the following two factors: (1) are intended for only general wellness use, as defined in this guidance, and (2) present a low risk to the safety of users and other persons." U.S. FOOD & DRUG ADMIN., GENERAL WELLNESS: POLICY FOR LOW RISK DEVICES 2 (2016) (emphasis omitted), [hereinafter GENERAL WELLNESS] https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm429674.pdf [https://perma.cc/A2EN-C728].

[97] *See id.* at 1; *infra* notes 101-05 and accompanying text.

[98] *See infra* notes 110-29 and accompanying text.

[99] These first generation IoB prosthetics are not external, but second generation body-embedded prosthetics are also already in trials and use. *See* Elaine Yau, *Forget Pokemons—In World First, Hongkonger Applies Augmented Reality to Surgery*, S. CHINA MORNING POST (Aug. 25, 2016, 12:00 PM), http://www.scmp.com/lifestyle/health-beauty/article/2008395/hongkonger-uses-augmented-reality-surgery [https:// perma.cc/G9RM-SMRR]; *see also* Homa Alemzadeh et al., *Adverse Events in Robotic Surgery*, PLOS ONE, April 2016, at 2, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4838256/ [https://perma.cc/H356-4GSG].

[100] *See, e.g.*, *'Smart Leg' Makes Engineering Prize Shortlist*, BBC NEWS (May 16, 2016), http://www.bbc.com/news/science-environment-36302784 [https://perma.cc/DJ75-TFH8].

[101] *See, e.g.*, Eric Limer, *Meet the Man with iPhone-Controlled Bionic Arms*, GIZMODO (Apr. 13, 2013, 5:40 PM), https://gizmodo.com/5994603/meet-the-man-with-iphone-controlled-bionic-arms [https://perma.cc/U79P-FPKK].

[102] *See, e.g.*, FITBIT, *supra* note 89.

monitoring tattoos,[103] electronic skin[104] with an organic circuit,[105] and "smart" watches with lifestyle monitoring capability.[106]

But it is the last category of these first-generation IoB devices—enterprise, educational, and recreational devices—that presents perhaps the fastest growing category of first generation IoB devices. For example, connected glasses[107] and helmets[108] regularly offer workers information in real time in enterprise settings, and exoskeleton projects for soldiers offer new fighting capabilities.[109] Brain sensing headbands that rely on external EEG electrodes are now used in some classrooms, seeking to monitor student attention.[110] Recent patent filings indicate that Amazon has developed a wristband that conducts ultrasonic tracking of a worker's hands to monitor efficiency in performance of an assigned task,[111] and providing

[103] *See, e.g.*, Rose Etherington, *Biostamp Temporary Tattoo Electronic Circuits by MC10*, DEZEEN (Mar. 28, 2013), https://www.dezeen.com/2013/03/28/biostamp-temporary-tattoo-wearable-electronic-circuits-john-rogers-mc10/ [https://perma.cc/MYY8-GE36].

[104] *See* John Boyd, *Electronic Skin Can Track Your Health and Fitness*, FORBES (Apr. 16, 2016, 2:20 AM), https://www.forbes.com/sites/jboyd/2016/04/17/electronic-skin-can-track-your-health-and-fitness [https://perma.cc/XQN4-YTYQ].

[105] *See 'Electronic Skin' to Monitor Your Health*, BBC NEWS (Apr. 4, 2017), http://www.bbc.com/news/av/technology-39485527/electronic-skin-to-monitor-your-health [https://perma.cc/7S5Q-7ZNL].

[106] *See, e.g.*, Apple Watch, APPLE, https://www.apple.com/watch/ [https://perma.cc/TP4N-2U5A].

[107] *See* Scott Stein, *Google Glass Returns: This Time, It's Professional*, CNET (July 18, 2017, 9:18 AM), https://www.cnet.com/news/google-glass-2-goes-for-enterprise/ [https://perma.cc/9T8Q-F476].

[108] *See* Jenna McKnight, *Daqri's Augmented-Reality Construction Helmet Aims to "Change the Nature of Work,"* DEZEEN (Jan. 27, 2016), https://www.dezeen.com/2016/01/27/daqri-smart-construction-helmet-augmented-reality-wearable-technology/ [https://perma.cc/L94Y-NAFH].

[109] Neil C. Bhavsar, *Can Science Transform Us Into Superheroes?*, FUTURISM (Mar. 22, 2017), https://futurism.com/can-science-transform-us-into-superheroes/ [https://perma.cc/TE43-EL93] (citing Dan Lamothe, *Meet the Exoskeleton the Navy Is Testing to Make Sailors Stronger*, WASH. POST (Sept. 3, 2014), https://www.washingtonpost.com/news/checkpoint/wp/2014/09/03/meet-the-new-exoskeleton-the-navy-is-testing-to-make-sailors-stronger/ [https://perma.cc/RVT8-4G89]).

[110] *See* WALL ST. J., *supra* note 37.

[111] U.S. Patent Application No. 15/083,083, Pub No. 2017/0278051 (filed Mar. 28, 2016) (published Sept. 28, 2017) (Amazon Technologies, Inc., applicant), http://pdfaiw.uspto.gov/.aiw?PageNum=0&docid=20170278051&IDKey=0E2634BC1119&HomeUrl=http%3A%2F%2Fappft.uspto.gov%2Fnetacgi%2Fnph-Parser%3FSect1%3DPTO1%2526Sect2%3DHITOFF%2526d%3DPG01%2526p%3D1%2526u%3D%2Fnetahtml%2FPTO%2Fsrchnum.html%2526r%3D1%2526f%3DG%2526l%3D50%2526s1%3D20170278051.PGNR.%2526OS%3D%2526RS%3D [https://perma.cc/PV5D-CVXC].

haptic feedback to guide the employee's hands in the correct direction.[112] Clothing company L.L. Bean has announced that it is connecting its coats and boots to the blockchain[113] using sewn-in sensors,[114] becoming the latest participant in the broader fashion trend of connected clothing[115] with human-computer interfaces.[116] The Massachusetts Institute of Technology and Microsoft Research have developed temporary tattoos a wearer attaches to her body, allowing her to control various devices wirelessly as a convenience.[117] Gaming devices such as virtual skin[118] and augmented[119] or virtual reality headsets[120] allow for recreational blending of physical and digital reality. Networked in-ear translators help with live multilingual communication,[121] and eye-mapping applications[122] turn

[112] Ceylan Yeginsu, *If Workers Slack Off, the Wristband Will Know. (And Amazon Has a Patent for It.),* N.Y. TIMES (Feb. 1, 2018), https://www.nytimes.com/2018/02/01/technology/amazon-wristband-tracking-privacy.html [https://perma. cc/8ZDR-SRH8].

[113] For a discussion of blockchain technology and how "the recent development of Bitcoin and blockchain technologies has rekindled excitement about their potential among technologists and industry," see Kevin Werbach & Nicolas Cornell, *Contracts* Ex Machina, 67 DUKE L.J. 313, 313 (2017).

[114] Kim S. Nash, *L.L. Bean to Link Boots, Coats to a Blockchain*, WALL ST. J. (Feb. 7, 2018, 1:27 PM), https://blogs.wsj.com/cio/2018/02/07/l-l-bean-to-link-boots-coats-to-a-blockchain/ [https://perma.cc/6DPN-VE3L].

[115] *See, e.g.*, Rachel Metz, *Your Next Password May Be Stored in Your Shirt Cuff*, MIT TECH. REV. (Oct. 31, 2017), https://www.technologyreview.com/s/609264/your-next-password-may-be-stored-in-your-shirt-cuff [https://perma.cc/CT3Z-H3E8].

[116] For example, connected underwear has been developed to assist workers in lifting tasks as a type of exoskeleton. Maya Dangerfield, *Lab-Created Underwear Could Prevent Back Pain*, CNN BUS. (Aug. 30, 2017), http://money.cnn.com/video/technology/future/2017/08/30/lab-created-underwear-could-prevent-back-pain.cnnmoney/index.html [https://perma.cc/KQ2A-CL3E].

[117] Alice Morby, *DuoSkin Temporary Tattoos Can Remotely Control Devices*, DEZEEN (Aug. 17, 2016), https://www.dezeen.com/2016/08/17/mit-media-lab-researchers-duoskin-temporary-tattoos-control-devices/ [https://perma.cc/RFS5-NASQ].

[118] *See* The Verge (@verge), TWITTER (July 6, 2017, 11:50 PM), https://twitter.com/verge/status/883216517776773120 [https://perma.cc/UVD7-KLWQ] ("This 'wearable skin' makes virtual reality feel way too real.").

[119] *See, e.g.*, Chelsea Gohd, *Magic Leap Shows Off Their New Augmented Reality Headset,* FUTURISM (Dec. 22, 2017), https://futurism.com/magic-leap-shows-new-augmented-reality-headset/ [https://perma.cc/4A6S-3N2Q].

[120] *See* Will Greenwald, *The Best VR (Virtual Reality) Headsets of 2018,* PC MAG (Dec. 5, 2017, 12:13 PM), https://www.pcmag.com/article/342537/the-best-virtual-reality-vr-headsets [https://perma.cc/LLM4-WEHC].

[121] *See* David Pierce, *Doppler's Futuristic Earbuds Sound Great. They Also Speak Spanish*, WIRED (Oct. 19, 2016, 6:56 AM), https://www.wired.com/2016/10/dopplers-futuristic-earbuds-sound-great-also-speak-spanish [https://perma.cc/Y5Y5-MB5E]; *Discover the Technol-*

eyes into a mouse.[123] Similarly, both Facebook[124] and Microsoft[125] have disclosed that each company is currently working on brain-control interfaces that will allow users to operate computing devices with only their thoughts and the help of external thought-sensing devices.[126] Meanwhile, Nissan has announced work on "[b]rain-to-[v]ehicle" technology that will allow drivers to use "signals from their own brain to make the drive even more exciting."[127] These

ogy Behind the System, WAVERLY LABS, http://www.waverlylabs.com/ [https://perma.cc/A4FU-6D2E] ("The Pilot Speech Translation companion app connects the Pilot earbud to our cloud-based translation engine for access to all of our translation features.").

[122] *See, e.g.*, Victoria Woollaston, *We Wore Eye-Tracking Goggles on the Tube, in the Name of 'Science*,*'* WIRED (Oct. 7, 2016), http://www.wired.co.uk/article/exterion-eye-tracking-london-underground [https://perma.cc/DV5W-MHP8].

[123] *See* Jing Cao, *The Man Who Created LeapPad Wants to Turn Your Eyes into a Mouse,* BLOOMBERG (Aug. 26, 2016, 7:00 AM), https://www.bloomberg.com/news/articles/2016-08-26/the-man-who-created-leappad-wants-to-turn-your-eyes-into-a-mouse [https://perma.cc/Y2G9-FESF].

[124] *See* Thomas Claburn, *Zuckerberg's Absolutely Mental: Brain Sensors that Read YOUR MIND at 100 Words a Minute,* REGISTER (Apr. 20, 2017, 12:02 AM), https://www.theregister.co.uk/2017/04/20/facebook_brain_typing/ [https://perma.cc/ZMY3-YVK4]; Jolene Creighton, *Zuckerberg: Facebook Is Working on a Brain Interface That Lets You "Communicate Using Only Your Mind,"* FUTURISM (Apr. 18, 2017), https://futurism.com/zuckerberg-facebook-will-reveal-a-brain-interface-that-lets-you-communicate-using-only-your-mind/ [https://perma.cc/6ADL-XLBB]; Andrew Griffin, *Facebook Secretly Building Technology to Read People's Minds So They Can 'Type Directly from the Brain,'* INDEP. (Apr. 20, 2017, 8:55 AM), http://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-mind-reading-brain-technology-building-8-regina-dugan-pentagon-a7692481.html [https://perma.cc/V967-4854]; Mark Zuckerberg, *Live at F8!*, FACEBOOK (Apr. 18, 2017), https://www.facebook.com/zuck/videos/10103658355917211/ [https://perma.cc/3RRF-29YD].

[125] Microsoft's patent explains that neurological data would "modulate a continuous user interface" and that "[n]eurological data can be gathered through a variety of techniques. One non-invasive technique is electroencephalography (EEG)." U.S. Patent Application No. 15/152403, Publication No. 20170329392 (filed May 11, 2016) (published Nov. 16, 2017) (Keskin et al., applicant), http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20170329392.PGNR [https://perma.cc/Q8ST-ZNUZ]. Microsoft has a second patent application for "changing the state of an application by detecting neurological user intent data associated with a particular operation of a particular application state." U.S. Patent Application 15/152,401, Publication No. 9,864,431 (filed May 11, 2016) (published Jan. 9, 2018) (Keskin et al., applicant), http://patft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=9864431.PN.&OS=PN/9864431&RS=PN/9864431 [https://perma.cc/C3NE-YAMP].

[126] *See* Andrew Orlowski, *Microsoft Wants to Patent Mind Control*, REG. (Jan. 15, 2018, 3:28 PM), https://www.theregister.co.uk/2018/01/15/microsoft_bci_patent_application/ [https://perma.cc/95R4-XT5H].

[127] Some reports state that the driver is required to wear an electrode skullcap. Gareth Corfield, *If You Won't Use Your Brain Our Machine Will Use It for You, Nissan Tells Drivers*,

many examples highlight the reality that IoB is already here and quickly expanding. These examples also portend that our future is one where IoB is likely to be legally and socially transformational, for better or worse.[128]

It is noteworthy that unlike many of the earliest first-generation IoB devices whose stated purpose was "self-archival," i.e., a user's personal data collection for self-reflection and tracking,[129] today's first-generation IoB devices often explicitly disclose that furthering third-party "big data" research[130] is a prime motivator for their data collection.[131] This "big data" motivation in particular often drives IoB products marketed for employment and educational settings.[132] In one case, a brainwave headband company targeted educational institutions,[133] ostensibly to assist with monitoring students' attention levels[134] in educational settings.[135] The company also recently

REG. (Jan. 4, 2018, 6:18 PM), https://www.theregister.co.uk/2018/01/04/nissan_brain_controlled_car_wheeze/ [https://perma.cc/63W9-DMZ9].

[128] *See supra* notes 74-86 and accompanying text.

[129] *See supra* notes 101-05 and accompanying text.

[130] For example, DNA samples are uploaded and available through the Internet allowing for cloud-based user analysis in real time. João Medeiros, *DNA Analysis Will Build an Internet of Living Things*, WIRED (Jan. 8, 2016), http://www.wired.co.uk/article/dna-analysis-internet-living-things [https://perma.cc/GGX9-7UEA].

[131] In a medical context, the United Kingdom's National Health Service has experimented with Google DeepMind's Stream application. Jo Best, *DeepMind and the NHS: What It's Really Like to Use Google's Kidney Health App*, ZDNET (Jan. 10, 2018, 11:00 AM), http://www.zdnet.com/article/deepmind-and-the-nhs-what-its-really-like-to-use-googles-kidney-health-app/ [https://perma.cc/9YTC-4TGU].

[132] As explained by Kate Crawford and Jason Schultz, "it is possible to generate a detailed picture about a person's health, including information the person may never have disclosed to a health care provider." Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 98 (2014).

[133] Reviews of both the science behind the product and its efficacy have been mixed, at best, with one critique calling it "malfunctioning" and "cringeworthy." AJ Dellinger, *This Malfunctioning Brain-Scanning Headband Was the Most Cringeworthy Demonstration at CES 2016,* DAILY DOT (Jan. 14, 2016, 1:21 AM), https://www.dailydot.com/debug/brainco-brain-control-technology-ces/ [https://perma.cc/6JLB-A5WL]; Paige Rogers, *Company to Collect Brain Wave Data on 1.2 Mil Students in the Classroom*, NOQ REP. (Dec. 4, 2017), https://noqreport.com/2017/12/04/company-collect-brain-wave-data-1-2-mil-students-classroom/ [https://perma.cc/793W-ZQYV].

[134] Other brain sensing headband research similarly focuses on attention-level monitoring. Alexandra Simon-Lewis, *This Brain-Imaging Headband Can Reveal How Boring You Are*, WIRED (Feb. 27, 2017), http://www.wired.co.uk/article/brain-imaging-headband-communicate [https://perma.cc/9T8P-DVB5].

[135] Ms. Smith, *Company with No Privacy Policy to Collect Brainwave Data on 1.2 Million Students*, CSO (Dec. 5, 2017, 9:00 AM), https://www.csoonline.com/article/3239969/security/

announced its intention to collect data on over a million students to create "the world's biggest brainwave database."[136] In another case, the "Brainternet" project used external EEG nodes and a Raspberry Pi computer to connect a human brain to the Internet in real time[137] in order to continuously monitor brain activity; its creators hope to build a brain application programming interface[138] with bi-directional inputs and outputs.[139]

Legal scholarship has considered a portion of this innovation in the context of what was initially known as the "Quantified Self" movement,[140] primarily assessing the medical desirability and privacy implications of connected devices with health applications.[141] Professor Nicolas Terry expands this analysis to issues of autonomy and data control, explaining that the Quantified Self movement presents an inherent dichotomy of control—while patient collection of "medically inflected" data is encouraged, the definite copy of a

company-with-no-privacy-policy-to-collect-brainwave-data-on-1-2-million-students.html [https://perma.cc/6RRP-GLM4]. Part of the question underlying such devices, however, is what they are actually measuring and whether the collected metrics, in fact, demonstrate optimal student development. *See* Mark Molloy, *Intelligent People Are More Easily Distracted at Work, Study Claims,* TELEGRAPH (Jan. 19, 2016, 11:54 AM), http://www.telegraph.co.uk/news/newstopics/howaboutthat/12107840/IQ-Intelligent-people-are-more-easily-distracted-at-work.html [https://perma.cc/GTS9-LVRJ].

[136] Smith, *supra* note 135.

[137] Patrick Caughill, *Researchers Have Linked a Human Brain to the Internet for the First Time Ever*, FUTURISM (Sept. 14, 2017), https://futurism.com/researchers-have-linked-a-human-brain-to-the-internet-for-the-first-time-ever/ [https://perma.cc/P7K4-4P78].

[138] Wits University, *Biomedical Engineers Connecting a Human Brain to the Internet in Real Time*, MED. XPRESS (Sept. 14, 2017), https://medicalxpress.com/news/2017-09-biomedical-human-brain-internet-real.html [https://perma.cc/X62T-SDHJ].

[139] Caughill, *supra* note 137 ("Brainternet can be further improved to classify recordings through a smart phone app that will provide data for a machine-learning algorithm. In future, there could be information transferred in both directions—inputs and outputs to the brain.").

[140] *See* Kashmir Hill, *Adventures in Self-Surveillance, A.K.A. The Quantified Self, A.K.A. Extreme Naval-Gazing*, FORBES (Apr. 7, 2011, 11:34 AM), http://www.forbes.com/sites/kashmirhill/2011/04/07/adventures-in-self-surveillance-aka-the-quantified-self-aka-extreme-navel-gazing/ [https://perma.cc/32TX-S8B3]; *The Quantified Self: Counting Every Moment*, ECONOMIST (Mar. 3, 2012), http:// www.economist.com/node/21548493 [https://perma.cc/N9LV-7DCS].

[141] Professor Nathan Cortez explains, "[w]hen viewed more broadly, mobile health is part of broader cultural and technological evolutions, including the march towards more personalized medicine, the 'quantified self' movement, the 'lifelogging' phenomenon, and the rising era of 'big data.'" Nathan Cortez, *The Mobile Health Revolution?*, 47 U.C. DAVIS L. REV. 1173, 1197-98 (2014) (footnotes omitted).

health record will always reside with the medical provider.[142] Similarly, Professor Craig Konnoth explains that the Quantified Self movement positions itself as a way of "knowing oneself,"[143] and Professor Frank Pasquale warns that the Quantified Self combined with Big Data offer "frightening opportunities to cure and exploit human vulnerabilities."[144]

Building on the era of the Quantified Self, the age of the Internet of Bodies presents the next iteration of these concerns: IoB adds legal concerns regarding the *physical safety* and continued functionality of the attached human bodies themselves.[145] It also adds a new autonomy question: the inability to disconnect in some cases. Use of some IoB devices becomes progressively less optional. Perhaps your employer or your school now requires that you wear a location tracking badge or perhaps your medical device manufacturer (mandated by your insurance provider) discontinues all devices without internet connectivity. In other words, IoB impacts legal interests in *physical* safety—the integrity, availability, and functional autonomy of human bodies, not merely legal concerns with respect to the confidentiality of data originating from those bodies.[146]

Thus, IoB transforms legal questions of data commodification into legal questions about the commodification and physical control of the human body itself. Indeed, the newest first-generation IoB devices sometimes invert the relationship between the attached body and the remote machines, using human bodies purely as fungible and rentable commodities for their physicality and energy

[142] Nicolas P. Terry, *Big Data Proxies and Health Privacy Exceptionalism*, 24 HEALTH MATRIX 65, 84 (2014) ("At root such patient curation of health data bespeaks autonomy.... However, it fails to take into account ... the canonical version of the record will remain in the provider's control ... [and] that only the provider-curated copy is protected by HIPAA-HITECH.... A similarly dichotomous result is likely as the medically quantified self develops.").

[143] Craig Konnoth, *Health Information Equity*, 165 U. PA. L. REV. 1317, 1341-42 (2017) ("[T]he 'quantified-self' movement promotes data streams as the best form of self-conceptualization and knowledge. This movement promotes the use of devices that not only 'solve problems related to health,' but also produce data ... as a way of knowing oneself.") (footnote omitted).

[144] Frank Pasquale, *Grand Bargains for Big Data: The Emerging Law of Health Information*, 72 MD. L. REV. 682, 684 (2013) ("An era of 'big data' promises exhilarating and frightening opportunities to cure and exploit human vulnerabilities.").

[145] *See infra* notes 282-92 and accompanying text.

[146] *See id.*

extrusion, a non-data driven purpose. For example, one Japanese inventor developed a way to rent another person's body as a telepresence "robot" to allow someone to attend a meeting both physically and remotely.[147] In another case, research using external caps of brain electrodes enabled gamers to control the conduct of another player's body through the Internet in order to play a question-and-answer game.[148] Finally, a Dutch startup recently developed suits intended to extract heat from the human body and repurpose it for cryptocurrency mining.[149] Referring to these first-generation IoB cryptocurrency mining suits, Professor Mark Lemley recently quipped, "It only took 18 years for us to actually implement the Matrix."[150]

Professor Lemley's dark humor points to an important and perhaps ethically uncomfortable inversion—the human body is now being leveraged as a functional vehicle to power external, Internet-connected processes.[151] Even when the research described above seeks to generate mature interface technologies[152] with tangible safety[153] and other[154] applications, the "thing-ified" nature of the human body implicit in the undertaking may trigger safety and dignitary concerns (and incredulous callbacks to "body snatcher"

[147] Will Knight (@willknight), TWITTER (Jan. 29, 2018, 10:52 PM), https://twitter.com/willknight/status/958231499509149697 [https://perma.cc/6R3C-VNSS] ("'Human Uber,' developed in Japan, provides a way to attend events remotely using another person's body.").

[148] George Dvorsky, *This Gamer Used His Thoughts to Control the Movements of Another Player*, GIZMODO (Nov. 6, 2014, 12:30 PM), https://io9.gizmodo.com/new-brain-interface-allows-for-mind-to-mind-video-gamin-1655415879 [https://perma.cc/W37N-N423].

[149] Camille Charluet, *This Startup Uses Body Heat to Mine Crypto—for When Robots Take Our Jobs*, NEXT WEB (Dec. 12, 2017, 12:21 PM), https://thenextweb.com/insider/2017/12/12/startup-uses-body-heat-to-mine-crypto-for-when-robots-take-jobs/ [https://perma.cc/UB2G-WP74].

[150] Mark Lemley (@marklemley), TWITTER (Dec. 15, 2017, 10:55 AM), https://twitter.com/marklemley/status/941743490316222465 [https://perma.cc/HQL9-UYBQ].

[151] *See* INSTITUTE OF HUMAN OBSOLESCENCE, supra note 58.

[152] *See generally* Rajesh P.N. Rao et al., *A Direct Brain-to-Brain Interface in Humans*, PLOS ONE, Nov. 2014, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.011133 [https://perma.cc/HQ68-UGPL].

[153] Dvorsky, *supra* note 148 ("[F]or example, the brain of a sleepy airplane pilot dozing off at the controls could stimulate the copilot's brain to become more alert.").

[154] George Dvorsky, *New Brain-Link Tech Means We Can Now Play 20 Questions with Our Minds*, GIZMODO (Sept. 25, 2015, 3:00 PM), https://io9.gizmodo.com/new-brain-link-tech-means-we-can-now-play-20-questions-1732991346 [https://perma.cc/BYY5-SS4F] ("The researchers are hopeful, for example, that a similar system could be used by people with Broca's aphasia.").

movies).[155] These uncomfortable questions of third-party processes controlling human bodies become even more pronounced in the context of second-generation IoB—IoB devices that are embedded inside the body.

2. Second-Generation IoB: Body Internal

Second-generation IoB technologies refer to those IoB devices where a portion of the device resides inside the body or accesses the body by breaking the skin.[156] For example, pacemakers have long included digital components,[157] and cochlear implants now include functionality reliant on Bluetooth.[158] Digital pills (already approved for the market by the FDA)[159] rely on a 3D-printed circuit and a transmitter inside a capsule.[160] Along similar lines, several companies[161] are currently racing to bring an IoB artificial "pancreas"[162]

[155] *Invasion of the Body Snatchers*, IMBD, https://www.imdb.com/title/tt0049366/ [https://perma.cc/C73B-YH23].

[156] *See* David Horrigan, *The Internet of Bodies: A Convenient—and, Yes, Creepy—New Platform for Data Discovery*, LEGALTECH NEWS (Jan. 7, 2019, 11:30 AM), https://www.law.com/legaltechnews/2019/01/07/the-internet-of-bodies-a-convenient-and-yes-creepy-new-platform-for-data-discovery [https://perma.cc/SF9Q-W8DF].

[157] Lisa Vaas, *Doctors Disabled Wireless in Dick Cheney's Pacemaker to Thwart Hacking*, NAKED SECURITY (Oct. 22, 2013), https://nakedsecurity.sophos.com/2013/10/22/doctors-disabled-wireless-in-dick-cheneys-pacemaker-to-thwart-hacking/ [https://perma.cc/R24R-UEKM].

[158] *True Wireless™ Accessories*, COCHLEAR, http://www.cochlear.com/wps/wcm/connect/us/home/treatment-options-for-hearing-loss/wireless-accessories [https://perma.cc/4N2R-T6NK].

[159] *FDA Approves Pill with Sensor that Digitally Tracks if Patients Have Ingested Their Medication*, U.S. FOOD & DRUG ADMIN. (Nov. 13, 2017), https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm584933.htm [https://perma.cc/C92K-CJLR] ("Abilify MyCite (aripiprazole tablets with sensor) has an ingestible sensor embedded in the pill that records that the medication was taken.").

[160] After the pill is ingested, it is powered by chlorine ions inside the stomach and relays information through the Internet with the help of a dongle and smartphone. Kelsey Atherton, *Take Two Robots by Mouth*, POLITICO (Dec. 13, 2017, 5:21 AM), https://www.politico.com/agenda/story/2017/12/13/five-drugs-for-the-future-000592 [https://perma.cc/T99W-R683].

[161] *See, e.g.*, Stacy Lawrence, *Medtronic Artificial Pancreas May Hit the Market Next Spring, as Pivotal Trial Nears Final Data*, FIERCE BIOTECH (Apr. 7, 2016, 11:59 AM), https://www.fiercebiotech.com/medical-devices/medtronic-artificial-pancreas-may-hit-market-next-spring-as-pivotal-trial-nears [https://perma.cc/2TJD-8HRY]; *Admetsys to Exhibit Smart Pancreas™ at Massachusetts Institute of Technology (MIT) Enterprise Forum 2015 Startup Spotlight*, PR URGENT (June 15, 2015), http://prurgent.com/2015-06-15/pressrelease387382.htm [https://perma.cc/E7LR-FPAG] [hereinafter *Admetsys to Exhibit Smart Pancreas™*].

[162] Paul Karoff, *Artificial Pancreas System Aimed at Type 1 Diabetes Mellitus*, HARV. GAZETTE (Jan. 4, 2016), http://news.harvard.edu/gazette/story/2016/01/artificial-pancreas-

to market—an implantable Internet-connected, sometimes 3D printed[163] "pancreas" managed by software and a mobile phone app.[164] Indeed, the FDA has already approved the first of these artificial pancreas devices.[165] Sensor-enabled sutures can now collect data on healing wounds,[166] and chips with cameras can report information from inside the heart during surgery.[167]

Prosthetics manufacturers have also embarked upon "smart" product[168] development, announcing that the next generation of

system-aimed-at-type-1-diabetes-mellitus/ [https://perma.cc/Z8ZH-29CA] ("The artificial pancreas is not a replica organ; it is an automated insulin delivery system designed to mimic a healthy person's glucose-regulating function."). Intended as a next generation insulin pump, these devices would engage in continuous monitoring of a patient's glucose levels, releasing insulin into the body when needed. *Admetsys to Exhibit Smart Pancreas*™, *supra* note 161.

[163] 3D printing is also being used with prosthetic limbs. Ian Birrell, *3D-Printed Prosthetic Limbs: The Next Revolution in Medicine*, GUARDIAN (Feb. 19, 2017, 1:59 AM), https://www.theguardian.com/technology/2017/feb/19/3d-printed-prosthetic-limbs-revolution-in-medicine [https://perma.cc/VH7Q-BWPX]; Meghan Neal, *3D Bioprinters Could Make Enhanced, Electricity-Generating 'Superorgans,'* VICE: MOTHERBOARD (June 13, 2014, 2:15 PM), http://motherboard.vice.com/read/3d-bioprinters-could-make-enhanced-electricity-generating-superorgans [https://perma.cc/3SSC-B2RF]. Ultimately, patients may be able to print new prosthetics at home with advancements in 3D printing. Matt Reynolds, *Print Your Own Prosthetic: This Code Can Be Used by Anyone to Create Their Own Bionic Limbs*, WIRED (Nov. 5, 2016), http://www.wired.co.uk/article/samantha-payne-bionic-arm-builder [https://perma.cc/9JXL-E6NP].

[164] Karoff, *supra* note 162 ("The closed-loop system consists of an insulin pump, a continuous glucose monitor placed under the user's skin, and advanced control algorithm software embedded in a smartphone that provides the engineering brains, signaling how much insulin the pump should deliver to the patient based on a range of variables, including meals consumed, physical activity, sleep, stress, and metabolism.").

[165] *The Artificial Pancreas Device System*, U.S. FOOD & DRUG ADMIN., http://www.fda.gov/medicaldevices/productsandmedicalprocedures/homehealthandconsumer/consumerproducts/artificialpancreas/default.htm [https://perma.cc/A235-VJB6].

[166] Patrick Collins, *Researchers Invent "Smart" Thread that Collects Diagnostic Data When Sutured into Tissue*, TUFTSNOW (July 18, 2016), http://now.tufts.edu/news-releases/researchers-invent-smart-thread-collects-diagnostic-data-when-sutured-tissue [https://perma.cc/9WXQ-T4L8].

[167] Eric Butterman, *A Way to Your Heart?*, ASME (June 2016), https://www.asme.org/engineering-topics/articles/bioengineering/a-way-to-your-heart [https://perma.cc/H2XJ-J4E7]. Virtual reality rigs are also recording surgery for training purposes. Gian Volpicelli, *What's Next for VR Surgery?*, WIRED (Apr. 14, 2016), http://www.wired.co.uk/article/wired-health-virtual-reality-surgery-shafi-ahmed [https://perma.cc/P7LJ-6QH3].

[168] For example, one such prosthesis uses "15 different sensors that are measuring different parameters with every step that the person is taking, and that data is being processed by three different onboard computers." Rob Hawley, *Wounded Veteran Among Those Benefiting from 'Smart' Prosthetic Ankle*, CBS N.Y. (Nov. 18, 2015, 2:54 PM), https://newyork.cbslocal.com/2015/11/18/bionx-biom-prosthetic-ankle/ [https://perma.cc/2FGK-F9EM].

prosthetics will be hardwired into patients' nerves[169] and muscles,[170] thereby merging flesh with computer code and hardware.[171] The Department of Veterans Affairs Center for Innovation has also launched programs aimed at creating a series of open-source "smart" prosthetics for wounded veterans.[172] Meanwhile, the Defense Advanced Research Projects Agency (DARPA) has been funding the development of next generation bionic arms,[173] and DARPA's Revolutionizing Prosthetics program successfully fitted a paralyzed woman with two nodes directly on her brain, allowing her to pilot a plane in a simulation.[174] Recent research[175] also demonstrated that with the help of an electrode array implanted in the brain, amputees will be able to move digits on a prosthesis with their thoughts alone, even without extensive training.[176] To wit, a monkey recently

[169] DARPA (@DARPA), Twitter (Oct. 27, 2016, 3:47 PM), https://twitter.com/DARPA/status/791773227190194182 [https://perma.cc/5L2D-HV7H] ("Video: Interface connecting prosthetic hand to nervous system helps amputees feel just how hard to squeeze.... #HAPTIX.").

[170] *See* Andrea Powell, *AI Is Fueling Smarter Prosthetics Than Ever Before*, Wired (Dec. 22, 2017, 12:13 PM), https://www.wired.com/story/ai-is-fueling-smarter-prosthetics-than-ever-before/ [https://perma.cc/9KBT-2B25].

[171] *See* Hawley, *supra* note 168 ("Carignan says his company wants to tie sensors into the existing muscles and nerves of the patient so they could have more active control over how the ankle works.").

[172] *VA to Launch Innovation Creation Series for Prosthetics and Assistive Technologies*, U.S. Dep't Veterans Aff. (May 15, 2015, 11:56 AM), https://www.blogs.va.gov/VAntage/19925/va-launches-innovation-creation-series-prosthetics-assistive-technologies/ [https://perma.cc/N9AY-9YDA]. A generation of young amputees is also currently nudging innovation and optional enhancement in IoB technology. Maria Doyle, *Teachers Design Smart, Connected Prosthesis for Double Amputee*, LinkedIn (Mar. 16, 2015), https://www.linkedin.com/pulse/teachers-design-smart-connected-prosthesis-double-amputee-maria-doyle?trk=portfolio_article-card_title [https://perma.cc/BU2E-32SH] ("Concepts include: Connecting with trail maps and conditions via the Internet [and] [t]racking performance.").

[173] *See* B.J. Murphy, *DARPA Hands Off Bionic Luke Arm to Military Medical Center*, Serious Wonder (Dec. 24, 2016), http://www.seriouswonder.com/darpa-hands-off-bionic-luke-arm-military-medical-center/ [https://perma.cc/YQL6-KLY6].

[174] Abby Phillip, *A Paralyzed Woman Flew an F-35 Fighter Jet in a Simulator—Using Only Her Mind*, Wash. Post (Mar. 3, 2015), https://www.washingtonpost.com/news/speaking-of-science/wp/2015/03/03/a-paralyzed-woman-flew-a-f-35-fighter-jet-in-a-simulator-using-only-her-mind/ [https://perma.cc/2WPU-C3ZG]; *see also About Braingate*, Braingate, https://www.braingate.org/about-braingate/ [https://perma.cc/7TSA-XCYS] (explaining that "micro-electrodes" implanted in the brain can be used to operate external devices).

[175] Guy Hotson et al., *Individual Finger Control of a Modular Prosthetic Limb Using High-Density Electrocorticography in a Human Subject*, J. Neural Engineering, Feb. 2016, at 10.

[176] George Dvorsky, *Brain Implant Will Let Amputees Move Individual Fingers on*

controlled its wheelchair wirelessly using a brain implant and its thoughts,[177] part of research toward the development of brain-controlled robotic exoskeletons for humans.[178]

Indeed, the potential health outcomes from these second-generation IoB technologies may be life-altering for many patients. For example, a brain implant currently in trials is expected to demonstrate the ability to restore sight to the blind,[179] and a different brain implant has already helped a paralyzed man regain his sense of touch.[180] Similarly, recent innovations in brain bypass[181] technologies have allowed quadriplegics to operate their limbs with the assistance of brain-implanted microelectrodes, external machines, and a sleeve.[182] A locked-in sufferer of Lou Gehrig's disease has also successfully tested a brain implant of four sensor strips that wirelessly connected to a computer interface and allowed the patient to type out messages using her eyes and "brain clicks"[183]—the thought

Prosthetics with Thoughts Alone, GIZMODO (Feb. 16, 2016, 3:10 PM), https://gizmodo.com/brain-implant-lets-amputees-move-individual-fingers-on-1759445814 [https://perma.cc/WN69-JKW3].

[177] Sankaranarayani Rajangam et al., *Wireless Cortical Brain-Machine Interface for Whole-Body Navigation in Primates*, SCI. REP., Mar. 2016, at 1, https://www.nature.com/articles/srep22170 [https://perma.cc/8LQW-WTR2].

[178] *See* Loura Hall, *NASA's Ironman-Like Exoskeleton Could Give Astronauts, Paraplegics Improved Mobility and Strength*, NASA (Aug. 7, 2013), https://www.nasa.gov/offices/oct/home/feature_exoskeleton.html [https://perma.cc/J3UE-GRSV]; George Dvorsky, *This Monkey Is Controlling a Wheelchair With Its Mind*, GIZMODO (Mar. 3, 2016, 9:00 AM), https://gizmodo.com/this-robotic-wheelchair-is-being-controlled-by-a-monkey-1762391710 [https://perma.cc/8RZJ-B89C].

[179] Dom Galeon, *A New Vision-Restoring Brain Implant Could Give Sight to the Blind*, FUTURISM (Feb. 13, 2017), https://futurism.com/4-theres-a-brain-implant-that-could-restore-vision-to-the-blind/ [https://perma.cc/NN4N-HEB2].

[180] Jess Vilvestre, *A Paralyzed Man Just Regained the Sense of Touch, Thanks to a Brain Implant*, FUTURISM (Oct. 14, 2016), https://futurism.com/a-paralyzed-man-just-regained-the-sense-of-touch-thanks-to-a-brain-implant/ [https://perma.cc/W936-RYMA].

[181] Neural bypass experiments are expected to yield significant results in the near future. *See, e.g.*, Beth Mole, *Using Synthetic Nervous System, Paralyzed Man Is First to Move Again*, ARS TECHNICA (Apr. 13, 2016, 4:40 PM), https://arstechnica.com/science/2016/04/with-synthetic-nervous-system-paralyzed-man-is-first-to-move-again/ [https://perma.cc/SK5E-258J]; Antonio Regalado, *Reversing Paralysis*, MIT TECH. REV. (Mar./Apr. 2017), https://www.technologyreview.com/s/603492/10-breakthrough-technologies-2017-reversing-paralysis/ [https://perma.cc/SH8R-3ZM7].

[182] George Dvorsky, *Brain Implant Enables Quadriplegic Man to Play* Guitar Hero *with His Hands,* GIZMODO (Apr. 13, 2016, 1:00 PM), https://gizmodo.com/brain-implant-enables-quadriplegic-man-to-play-guitar-h-1770566874 [https://perma.cc/9438-VRWL].

[183] Mariska J. Vansteensel et al., *Fully Implanted Brain-Computer Interface in a Locked-*

of "mov[ing] her hand for approximately 1 second."[184] Many promising second-generation IoB medical devices offer potentially transformational outcomes.

However, just as with first-generation IoB devices, while the earliest second-generation IoB devices have usually been classified as medical devices by the FDA,[185] later second-generation IoB may include devices whose manufacturers may consider them to be "healthy lifestyle" and non-medical devices.[186] Indeed, the FDA considered most first-generation IoB devices to be nonmedical,[187] and the FDA has flagged that a number of relevant guidance documents may evolve in the future because of the 21st Century Cures Act of December 2016.[188] Meanwhile, technologically, the line between first-generation and second-generation IoB "healthy lifestyle"/nonmedical technology is already beginning to blur.

In Patient with ALS, 375 NEW ENG. J. MEDICINE 2060, 2060-63 (2016).

[184] *Id.* The machine activates whenever she thinks about moving her hand for about one second. *Id.*

[185] For example, pacemakers fall into the most regulated category, Class III medical devices. Class III devices pose the greatest risk and, thus, are subject to a rigorous premarket approval (PMA) process. *See* Medtronic, Inc. v. Lohr, 518 U.S. 470, 476-77 (1996); 21 C.F.R. § 870.3680 (2012); 21 C.F.R. § 870.3710 (2011).

[186] The FDA defines a device as follows:

> A device is: "an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is: 1. recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them, 2. intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or 3. intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes. The term "device" does not include software functions excluded pursuant to section 520(o).

Is the Product a Medical Device?, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/ClassifyYourDevice/ucm051512.htm [https://perma.cc/M7MA-5UMC].

[187] GENERAL WELLNESS, *supra* note 96, at 2-5.

[188] *Digital Health*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/medicaldevices/digitalhealth/ [https://perma.cc/QR7Y-WTDZ]. The Cures Act was signed into law on December 13, 2016. *21st Century Cures Act*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/regulatoryinformation/lawsenforcedbyfda/significantamendmentstothefdcact/21stcenturycuresact/default.htm [https://perma.cc/2P68-6FWF].

For example, a team of researchers in Australia is currently working on ingestible electronic capsules that monitor gas levels in the human intestinal tract to track variability driven by food consumption.[189] Ingestible digital pills such as this one are likely to (attempt to) enter the market as a "healthy lifestyle" device.[190] While medical uses are foreseeable, the FDA may also analyze this digital pill as primarily monitoring the effects of selective food consumption in healthy bodies.[191] Therefore, much like a connected fitness tracker, this digital gas monitoring pill may fall outside the definition of a "medical device."[192] Consider also a swallowable "smart" vitamin absorption/sleep tracker that sends information about your body to your phone using Bluetooth, which then in turn uploads the information to the tracker company's cloud.[193] This IoB product would also potentially be deemed akin to a fitness tracker and, therefore, perhaps not necessarily classified as a medical device.[194] As a consequence, it too may fall within the "healthy lifestyle" device categorization and outside the definition of a medical device. But some second-generation IoB devices will fall squarely outside either of these health-related categories and reflect selective, aesthetic human self-augmentation.[195]

[189] Beth Mole, *With Ingestible Pill, You Can Track Fart Development in Real Time on Your Phone*, ARS TECHNICA (Jan. 9, 2018, 7:30 AM), https://arstechnica.com/science/2018/01/with-ingestible-pill-you-can-track-fart-development-in-real-time-on-your-phone/ [https://perma.cc/54AP-A97D] (noting the digital pill is paired with a receiver and mobile phone in order to report gas production conditions inside the human body in real time).

[190] *Id.*

[191] *See id.*

[192] *See* GENERAL WELLNESS, *supra* note 96, at 3, 6-7. The FDA also does not currently review vitamin "supplements" for safety and effectiveness before they are marketed, instead relying on manufacturers to verify their safety. *Dietary Supplements: What You Need to Know*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/food/dietarysupplements/usingdietarysupplements/ucm109760.htm [https://perma.cc/4RWC-RNDY].

[193] *Cf.* Mole, *supra* note 189.

[194] If the FDA chooses to take a similar hands-off approach, this device would fall primarily under the FTC's jurisdiction to police health claims. *See, e.g.*, Press Release, Fed. Trade Comm'n, FTC Issues Enforcement Policy Statement Regarding Marketing Claims for Over-the-Counter Homeopathic Drugs (Nov. 15, 2016), https://www.ftc.gov/news-events/press-releases/2016/11/ftc-issues-enforcement-policy-statement-regarding-marketing [https://perma.cc/N6GQ-P3AT].

[195] Also consider the Circadia, an implantable device that allows for wellness tracking of "biomedical data and transmit[s] it to the Internet via Bluetooth." Dom Benoscek, *NIFTIT Partners with Grindhouse Wetware*, NIFTIT BLOG (Dec. 3, 2013), niftit.com/niftit-grindhouse-wetware/ [https://perma.cc/KAY3-B3YM].

While occupational and recreational self-augmentation using second-generation IoB may seem the stuff of dystopian science fiction or merely the unusual hobby of (overly)enthusiastic computer scientists[196] and controversial artists,[197] this practice is, in reality, no longer limited to fiction and the social avant-garde.[198] Indeed, estimates contend that approximately 50,000 to 100,000 people in the U.S.[199] currently have microchips implanted in their bodies.[200] Employers are encouraging their employees to chip themselves for convenience,[201] repurposing technologies long used safely on animals.[202] Chips can be used to store contact information for emergencies or Bitcoin wallet addresses,[203] and chips can be custom programmed to, for example, place a phone call when tapped to a phone,[204] open or lock a door,[205] or buy a smoothie.[206]

One company already sells a do-it-yourself implant kit for a few hundred dollars[207] which allows for purchasers to modify their

[196] *See infra* Part II.B.

[197] *See* Stuart Jeffries, *Neil Harbisson: The World's First Cyborg Artist*, GUARDIAN (May 6, 2014, 2:59 AM), https://www.theguardian.com/artanddesign/2014/may/06/neil-harbisson-worlds-first-cyborg-artist [https://perma.cc/HM9B-99PZ].

[198] *See* Trevor Callaghan (@trevolafoam), TWITTER (Sept. 17, 2016, 3:24 AM), https://twitter.com/trevolafoam/status/777090730472923136 [https://perma.cc/99PS-XMKD] ("Implant Party! #FutureFest16").

[199] The practice is also gaining in popularity in other countries such as Australia. Emma Reynolds, *Australians Embracing Super-Human Microchip Technology*, NEWS.COM.AU (Aug. 25, 2016, 8:32 AM), http://www.news.com.au/technology/gadgets/wearables/australians-embracing-superhuman-microchip-technology/news-story/536a08003cb07cba23336f83278a5003 [https://perma.cc/QX8D-VAKM].

[200] Yael Grauer, *A Practical Guide to Microchip Implants*, ARS TECHNICA (Jan. 3, 2018, 7:30 AM), https://arstechnica.com/features/2018/01/a-practical-guide-to-microchip-implants/ [https://perma.cc/995P-PN84].

[201] *See* Associated Press, *Companies Start Implanting Microchips into Workers' Bodies*, L.A. TIMES (Apr. 3, 2017, 10:15 AM), http://www.latimes.com/business/technology/la-fi-tn-microchip-employees-20170403-story.html [https://perma.cc/WY2B-59RQ].

[202] *See Microchipping of Animals FAQ*, AM. VETERINARY MED. ASS'N, https://www.avma.org/KB/Resources/FAQs/Pages/Microchipping-of-animals-FAQ.aspx [https://perma.cc/NB6J-R3QB].

[203] Cyrus Farivar, *Man Has NFC Chips Injected into His Hands to Store Cold Bitcoin Wallet*, ARS TECHNICA (Nov. 15, 2014, 11:00 AM), https://arstechnica.com/information-technology/2014/11/man-has-nfc-chips-injected-into-his-hands-to-store-cold-bitcoin-wallet/ [https://perma.cc/D49Q-ZCPN].

[204] Grauer, *supra* note 200.

[205] Farivar, *supra* note 203.

[206] Associated Press, *supra* note 201.

[207] CYBORGNEST, https://cyborgnest.net/ [https://perma.cc/3QP4-GKQC]. As of February

bodies[208] in various Internet-connected ways, such as vibrating whenever the wearer is facing north.[209] For example, one wearer uses an implant to inform her when a seismic movement occurs.[210] Another wearer—the first legally recognized "cyborg" per his UK passport—fused his implant to his brain to have it translate color into musical tones.[211] Informal "biohacking" communities and hackathons[212] are increasingly popular, and formalized conferences and workshops already exist.[213]

Also, as in every Internet context, marketing and "customer experience" data collection is pushing new technology adoption. Indeed, recent patent filings indicate this dynamic has already arrived to second-generation IoB.[214] For example, British Airways has filed a patent with the UK Intellectual Property Office seeking to patent a swallowable "ingestible sensor" to monitor customer experience on flights from the inside of customers' bodies.[215]

In particular, as the preceding examples illustrate, one of the business dynamics visible in the evolution of second-generation IoB technologies is the merger of first and second-generation medical IoB with other existing consumer technologies, creating new recreational (nonmedical) IoB.[216] For example, in medical contexts,

2017, around 1000 people had ordered a north-sensing kit. Adam Popescu, *This $425 DIY Implant Will Make You a Cyborg*, BLOOMBERG BUSINESSWEEK (Feb. 16, 2017, 10:30 AM), https://www.bloomberg.com/news/articles/2017-02-16/this-425-diy-implant-will-make-you-a-cyborg [https://perma.cc/R9XD-72AC].

[208] Generally two small titanium barbells akin to a piercing are implanted in the wearer's chest. *Id.*

[209] *Id.*

[210] *Id.*

[211] *Id.*

[212] Nicole Kobie, *How to Hack Your Senses: From 'Seeing' Sound to 'Hair GPS'*, WIRED (July 5, 2016), http://www.wired.co.uk/article/how-to-hack-senses-see-sound [https://perma.cc/H3NF-R9ML].

[213] *See, e.g.*, *Biohackers at DEFCON*, DEFCON BIOHACKING VILLAGE, https://www.defconbiohackingvillage.org/ [https://perma.cc/7FHW-RVEN]. A particularly engaged community exists in Brooklyn. *See Brooklyn Biohackers*, MEETUP, https://www.meetup.com/Brooklyn-Biohackers/?_cookie-check=4rhghGnRdc7nBd3x [https://perma.cc/S69D-24CK].

[214] Eleazer Corpuz, *British Airways Plans to Monitor Its Passengers with a 'Digital Pill'*, FUTURISM (Nov. 30, 2016), https://futurism.com/british-airways-plans-to-monitor-its-passengers-with-a-digital-pill/ [https://perma.cc/98SM-ZPKT].

[215] U.K. Patent Application No. 1600548.0, 2 1.34, Publication No. 2538339 (filed Mar. 24, 2014) (published Nov. 16, 2016) (British Airways PLC, applicant) (noting the sensor would communicate from the inside of the passenger's body).

[216] *See id.*

ocular lens implants have long been used as a surgical correction for eyes damaged by cataracts.[217] Meanwhile, in recreational contexts, first-generation IoB gaming and other augmented reality devices initially involved glasses[218] and other headgear,[219] but then they started to include wearable IoB contact lenses.[220] Blending these two technology trends—one from medicine and one from consumer and enterprise technology—it perhaps should be unsurprising that augmented reality and other recreational visual products are now creeping inside the eyeball in injected form.[221] In other words, while these lenses were first used for medical reasons,[222] they are now also used for recreational[223] and military[224] purposes. It is

[217] Millions of people receive ocular lens implants yearly as part of cataract surgeries. *See* Richard Lindstrom, *Thoughts on Cataract Surgery: 2015*, REV. OF OPHTHALMOLOGY (Mar. 9, 2015), https://www.reviewofophthalmology.com/article/thoughts-on--cataract-surgery-2015 [https://perma.cc/CDG2-YHQR].

[218] *See* Dieter Bohn, *Intel Made Smart Glasses that Look Normal*, VERGE (Feb. 5, 2018, 8:00 AM), https://www.theverge.com/2018/2/5/16966530/intel-vaunt-smartglasses-announced-ar-video [https://perma.cc/9WK7-27Z8]; Jacob Kleinman, *Augmented Reality Glasses: What You Can Buy Now (or Soon)*, TOM'S GUIDE (Feb. 14, 2018, 8:00 AM), https://www.tomsguide.com/US/BEST-AR-GLASSES,REVIEW-2804.HTML [https://perma.cc/AA9D-RQVH].

[219] *See* Lucas Matney, *RealWear Raises $17M as It Looks to Take a Simpler Approach to Enterprise AR Headgear*, TECHCRUNCH (Feb. 14, 2018), https://techcrunch.com/2018/02/14/realwear-raises-17m-as-itlooks-to-take-a-simpler-approach-to-enterprise-ar-headgear/ [https://perma.cc/J3UK-EVAE].

[220] *See* Nick Statt, *Augmented-Reality Contact Lenses to Be Human-Ready at CES*, CNET (Jan. 3, 2014, 4:00 AM), https://www.cnet.com/news/augmented-reality-contact-lenses-to-behuman-ready-at-ces/ [https://perma.cc/TQM9-TUH5].

[221] For example, Google has patented an injectable implant that corrects and enhances vision and comes with an antenna for connecting to the Internet and recharging using special glasses. *See* Anthony Cuthbertson, *Google Patents a Cyborg Lens that Injects into Your Eyeball*, NEWSWEEK (May 5, 2016, 5:14 AM), http://www.newsweek.com/google-patent-cyborg-smart-lens-injecteyeballs-455824 [https://perma.cc/7CMY-3SL5].

[222] *See* Alexandra Sifferlin, *Google Granted Patent for Smart Contact Lens*, TIME (Mar. 25, 2015), http://time.com/3758763/google-smart-contact-lens/ [https://perma.cc/2LY9-H866].

[223] Sony has filed a patent on new contact lenses that can record video. *See* Clemence Michallon, *Sony Files to Patent New Contact Lenses that Can Record Video, Store It, Play It Back—and Adjust Zoom, Focus and Aperture Automatically*, DAILYMAIL (Apr. 30, 2016, 5:27 PM), http://www.dailymail.co.uk/sciencetech/article-3567402/Sony-patent-application-reveals-new-contact-lensesrecord-video-store-play-adjust-zoom-focus-apertureautomatically.html#ixzz48ngMvB00 [https://perma.cc/7GM3-5CHY].

[224] Implanted augmented reality contact lenses might be useful in the creation of a generation of "super soldiers" according to some proponents of the technology. *See* Sarah Buhr, *Omega Opthalmics Is an Eye Implant Platform with the Power of Continuous AR*, TECHCRUNCH (Aug. 4, 2017), https://techcrunch.com/2017/08/04/ophthalmics-is-an-eye-implantwith-the-power-of-continuous-ar/ [https://perma.cc/4NPM-7CYJ].

this progressive creep that will also transform current brain prosthetics into the third generation of IoB—where body and mind meld with the Internet and remote computing, not only for medical purposes but also as a chosen aesthetic enhancement.

3. Third-Generation IoB: Body Melded

Third-generation IoB devices meld the human mind with external computers and the Internet.[225] As currently conceptualized, these devices primarily involve injected or implanted brain computer interfaces that act in a bi-directional read/write manner.[226] In other words, they functionally extend and externalize portions of the human mind.[227] Thus, one of the goals of third-generation IoB is the (optional) cognitive enhancement[228] of healthy, able-bodied humans with the help of brain-implanted computers and linkages.[229] As described by Elon Musk,[230] the founder of a company researching ways to connect computers directly to brains,[231] the goal is a "merger of biological intelligence and machine intelligence."[232] Entrepreneurs

[225] *See* Olivia Solon, *Elon Musk Says Humans Must Become Cyborgs to Stay Relevant. Is He Right?*, GUARDIAN (Feb. 15, 2017, 3:00 AM), https://www.theguardian.com/technology/2017/feb/15/elon-musk-cyborgs-robots-artificial-intelligence-is-he-right [https://perma.cc/SA8J-H8MG].

[226] *See id.*

[227] *See* Sarah Marsh, *Neurotechnology, Elon Musk and the Goal of Human Enhancement*, GUARDIAN (Jan. 1, 2018, 4:00 AM), https://www.theguardian.com/technology/2018/jan/01/elon-musk-neurotechnology-human-enhancement-brain-computer-interfaces [https://perma.cc/ZV6T-PF9C].

[228] The fear of AI takeover fuels discussion of enhanced brain capacity. *See* Christof Koch, *To Keep Up with AI, We'll Need High-Tech Brains*, WALL ST. J. (Oct. 27, 2017), https://www.wsj.com/articles/to-keep-up-with-ai-well-need-high-tech-brains-1509120930?mod=e2twd [https://perma.cc/647S-2LVN].

[229] *See* Nick Statt, *Kernel Is Trying to Hack the Human Brain—But Neuroscience Has a Long Way to Go*, VERGE (Feb. 22, 2017, 12:36 PM) https://www.theverge.com/2017/2/22/14631122/kernel-neuroscience-bryan-johnson-human-intelligence-ai-startup [https://perma.cc/7DFC-PS2A].

[230] *See* Kristin Houser, *Here's Everything You Need to Know About Elon Musk's Human/AI Brain Merge,* FUTURISM (Apr. 20, 2017), https://futurism.com/heres-everything-you-need-to-know-about-elon-musks-humanai-brain-merge/ [https://perma.cc/8PMV-9LA6].

[231] Musk has voiced his concern that humans will be overtaken by artificial intelligence and turned into the metaphorical equivalent of a "house cat[ ]." Solon, *supra* note 225; *see also* Sebastian Anthony, *Humans Must Become Cyborgs to Survive, Says Elon Musk*, ARS TECHNICA (Feb. 14, 2017, 8:50 AM), https://arstechnica.com/information-technology/2017/02/humans-must-become-cyborgs-to-survive-says-elon-musk/ [https://perma.cc/FU8D-9YN2].

[232] Solon, *supra* note 225.

building these third-generation IoB devices sometimes call them a "direct cortical interface,"[233] and they predict a coming "gold rush" in optional brain enhancement.[234] Indeed, some financial analysts forecast a $27 billion market for these devices by 2023.[235] Third-generation IoB devices are often not framed as a medical correction of a preexisting physical state.[236]

Despite Musk's recent assertions that third-generation IoB human testing will start in 2020,[237] perhaps reassuringly, current third-generation IoB devices are generally believed to be in relatively early stages of development.[238] While the goal of third-generation IoB includes brain enhancement and uploadable knowledge,[239] their current uses are, in fact, primarily in the context of treating medical conditions.[240] For example, brain prosthetic devices[241] with wireless components are currently being tested and prescribed for humans with Alzheimer's, Parkinson's, epilepsy, and other conditions.[242]

[233] *See* Cade Metz, *Elon Musk Isn't the Only One Trying to Computerize Your Brain*, WIRED (Mar. 31, 2017, 7:00 AM), https://www.wired.com/2017/03/elon-musks-neural-lace-really-look-like/?mbid=social_twitter [https://perma.cc/V87V-YLEM].

[234] *See* John H. Richardson, *Inside the Race to Hack the Human Brain*, WIRED (Nov. 16, 2017, 6:00 AM), https://www.wired.com/story/inside-the-race-to-build-a-brain-machine-interface/ [https://perma.cc/MHC7-P6XN].

[235] *Id.*

[236] *See id.*

[237] Stephen Shankland, *Elon Musk Says Neuralink Plans 2020 Human Test of Brain-Computer Interface*, CNET (July 17, 2019), https://www.cnet.com/news/elon-musk-neuralink-works-monkeys-human-test-brain-computer-interface-in-2020/ [https://perma.cc/3DJT-SZH8].

[238] *See* Christopher Mims, *A Hardware Update for the Human Brain*, WALL ST. J. (June 5, 2017), https://www.wsj.com/articles/a-hardware-update-for-the-human-brain-1496660400 [https://perma.cc/L2F3-34CC].

[239] *See* Mark Molloy, *Scientists Discover How to 'Upload Knowledge to Your Brain'*, TELEGRAPH (Mar. 1, 2016, 7:45 PM), http://www.telegraph.co.uk/technology/2016/03/01/scientists-discover-how-to-download-knowledge-to-your-brain/ [https://perma.cc/7MLG-QPVU].

[240] *See* Ian Sample, *Paraplegic Man Walks with Own Legs Again*, GUARDIAN (Sept. 23, 2015, 8:00 PM), https://www.theguardian.com/science/2015/sep/24/paraplegic-man-walks-with-own-legs-again [https://perma.cc/PRA4-WKK6].

[241] *See* Jens Clausen et al., *Help, Hope, and Hype: Ethical Dimensions of Neuroprosthetics*, 356 SCI. 6345, 1338 (2017). Deep brain stimulation devices, by contrast, have generally not included external facing components but for doctor interfaces in proximity, presumably. *See* Tim Urban, *Neuralink and the Brain's Magical Future*, WAIT BUT WHY (Apr. 20, 2017), https://waitbutwhy.com/2017/04/neuralink.html [https://perma.cc/8PPA-QJGF].

[242] *See* Robert Perkins, *Brain Prosthesis Aims to Provide Breakthrough for People Struggling with Memory Loss*, USC NEWS (Sept. 29, 2015), https://news.usc.edu/86658/new-device-aims-to-help-people-struggling-with-memory-loss/ [https://perma.cc/YH9G-GCJJ]. For example, memory prostheses have successfully replaced the Alzheimer's-damaged parts of a

Another possible medical application entails helping soldiers recover from postwar memory loss[243] and traumatic experiences.[244] But slippage into nonmedical uses of third-generation IoB is already visible. For example, third-generation IoB research also assists in the creation of cognitively-enhanced super-soldiers as part of the U.S. Armed Forces.[245] As explained by DARPA Director Arati Prabhakar, "we can now see the future where we can free the brain from the limitations of the human body.... We can only imagine amazing good things and amazing potentially bad things that are on the other side of that door."[246]

Although we have not yet evolved an infrastructure and other technical capabilities[247] that can successfully support the "brain-cloud"[248] ideal of third-generation IoB, some experts estimate the arrival of third-generation IoB technology to be as little as a decade away.[249] A decade may seem a long time to technologists, but in terms of legal evolution, this time frame signals a need for expedited debate and legal preparation.[250]

While the potentially life-changing medical and lifestyle impact of a portion of these technologies is unquestionable, it is also the case that people will inevitably be hurt (and killed) by some of these

patient's hippocampus. *See* Eileen Toh, *USC Researchers Develop Brain Implant to Improve Memory*, DAILY TROJAN (Nov. 19, 2017), https://dailytrojan.com/2017/11/19/usc-researchers-develop-brain-implant-improve-memory/ [https://perma.cc/6LJ9-4ZW2].

[243] *See* Perkins, *supra* note 242.

[244] *See* Matt Burgess, *Scientists Use AI to 'Rewrite' Painful Memories in People's Brains*, WIRED (Nov. 21, 2016), http://www.wired.co.uk/article/brain-fear-decode-erase [https://perma.cc/5CDQ-PWE8].

[245] The creation of super-soldiers is the alleged goal of a research program underway through DARPA. *See* Karla Lant, *DARPA Is Planning to Hack the Human Brain to Let Us "Upload" Skills*, FUTURISM (May 2, 2017), https://futurism.com/darpa-is-planning-to-hack-the-human-brain-to-let-us-upload-skills/ [https://perma.cc/4ZXS-XQEK].

[246] Phillip, *supra* note 174.

[247] *See* Houser, *supra* note 230 ("The company has to deal with the problems of biocompatibility, wirelessness, power, and ... bandwidth.").

[248] For a discussion of the cloud, see generally Stephanie K. Pell & Christopher Soghoian, *Can You See Me Now?: Toward Reasonable Standards for Law Enforcement Access to Location Data That Congress Could Enact*, 27 BERKELEY TECH. L.J. 117, 158 (2012) (discussing the Subcommittee on the Constitution, Civil Rights, and Civil Liberties of the Judiciary House of Representative's ECPA Reform and the Revolution in Cloud Based Computing Hearing).

[249] *See* Houser, *supra* note 230.

[250] *See id.* ("The engineering is only half the battle, though. Like Musk mentioned, regulatory approval will be a big factor in the development and adoption of Neuralink's tech.").

IoB technologies.[251] Consequently, impacted plaintiffs will seek recourse through the courts.[252] In order to begin to craft a coherent and innovation-sensitive legal approach to address these IoB harms, let us first examine four lessons from our experiences with IoT.

B. The "Legacy Code" of IoT

Agent Smith: Never send a human to do a machine's job.[253]

As one Silicon Valley startup explains, the human body is "the next big innovation platform."[254] While this statement is accurate in heralding the arrival of IoB, it should also serve as a harbinger of looming harms and legal challenges. Because technology past is generally technology prologue, to anticipate the legal future of IoB, we can turn to an examination of the present state of IoT. Instructively, a series of serious IoT implementation problems have arisen as IoT has gained popularity. Parallel problems are already arising or are likely to arise in IoB.

Four of these IoT implementation problems include the "Better with Bacon" problem of gratuitous Internet connectivity,[255] the "Magic Gadget" problem of failing to plan for failure,[256] the "Builder Bias" problem of shipping without securing,[257] and the "Mandatory Soup" problem of limited self-help, diminished market choice, and obsolescence through adhesion.[258] However, when these four problems manifest in IoB contexts, they will present one critical difference from their IoT incarnation: human bodies may be directly physically harmed.[259]

[251] *See* Urban, *supra* note 241.

[252] *See infra* Part I.C.2.

[253] THE MATRIX, *supra* note 2.

[254] *Nootrobox Is Hiring an Editor-in-Chief*, STARTUP.JOBS, https://startup.jobs/editor-in-chief-at-nootrobox [https://perma.cc/Q2Y6-P8FD].

[255] *See infra* Part I.B.1.

[256] *See infra* Part I.B.2.

[257] *See infra* Part I.B.3.

[258] *See infra* Part I.B.4.

[259] *See* Urban*, supra* note 241.

1. *The Better with Bacon Problem: Gratuitous Internet Connectivity*

An April Fools' Day joke from 2013 touted the launch of the Toaster.io—a toaster connected to the Internet.[260] At the time, the idea of a toaster being connected to the Internet seemed ridiculous to an average consumer.[261] In hindsight, of course, the Internet of Things exploded shortly thereafter, and Toaster.io was merely a preview of actual products soon to arrive on the market.[262]

The seemingly unrelenting "cybering" of all the consumer things[263] that occurred in the IoT marketplace calls to mind an admonition from Professor Siva Vaidhyanathan. Professor Vaidhyanathan has argued that perfunctory innovation may be supplanting the idea of progress, stating that "[p]rogress is out-of-fashion."[264] He argues that "innovation differs from progress in many ways. Innovation lacks a normative claim of significant betterment. It emerges from many small moves ... [and] does not contain an implication of a grand path or a grand design of a knowable future."[265] Indeed, our model of "innovation" often appears to involve relentlessly connecting consumer products to the Internet, even when a product's functionality is not necessarily materially enhanced by the Internet connectivity.[266]

This argument lies at the heart of what might be called the "Better with Bacon" problem.[267] Just as some restaurants seem to

[260] See Zack Whittaker, *The World's First Social Toaster*?, ZDNET (Apr. 1, 2013), https://www.zdnet.com/pictures/april-fools-2013-the-best-techy-pranks-of-the-day/3/ [https://perma.cc/SH33-RKB2].

[261] *See id.*

[262] *See* Roberto Baldwin, *The World Now Has a Smart Toaster*, ENGADGET (Jan. 4, 2017), https://www.engadget.com/2017/01/04/griffin-connects-your-toast-to-your-phone/ [https://perma.cc/598J-X9G4].

[263] *See* Elizabeth Nolan Brown, *Meme Origins: "All the Things" Tic Spawned by Artist Allie Brosh*, BUSTLE (Aug. 30, 2013), https://www.bustle.com/articles/4393-meme-origins-all-the-things-tic-spawned-by-artist-allie-brosh [https://perma.cc/K8T7-P5X6].

[264] Siva Vaidhyanathan, *The Golden Quarter*, AEON (May 13, 2015), https://aeon.co/users/siva-vaidhyanathan [https://perma.cc/6D39-HNB3].

[265] *Id.*

[266] For example, one might ask whether Internet connectivity meaningfully enhances the experience of a saw. Yet, saws are available in IoT form. *See Rotozip*, THE HOME DEPOT, https://www.homedepot.com/p/Rotozip-5-5-Amp-Corded-1-4-in-Rotary-RotoSaw-Spiral-Saw-Tool-Kit-with-5-Accessories-SS355-10/203408190 [https://perma.cc/873E-EWUB].

[267] One common technology variant of the Better with Bacon problem might be the

erroneously believe that all meals are "better" with an ample sprinkling of (sometimes unexpected) bacon,[268] so too some technology producers and users believe that every gadget is "better" with gratuitous, even if functionally nonessential, Internet capabilities.[269] While for some diners surprise bacon presents an unexpected benefit, for vegetarian diners, surprise bacon may effectively undermine the entirety of the enterprise. And, just as surprise bacon bits are never calorie-free (and sometimes unwelcome), gratuitous technology "bacon" is also never costless. It always comes at the expense of security.

While an Internet-connected toaster that, for example, emblazons the morning weather onto toast[270] might seem like a harmless curiosity for a kitchen or corporate break room, its Internet connectivity adds attack surface and material risk for the security of a network as a whole.[271] For example, a vulnerability in an IoT toaster may be an entree for compromising a company's or a consumer's otherwise protected network.[272] Particularly in sensitive situations with national security or infrastructural implications, the IoT

addition of Bluetooth devices. Bluetooth has been amply demonstrated to create additional vulnerabilities in systems. Chris Merriman, *BlueBorne: Bluetooth Hack Doesn't Require Pairing with Victims Devices,* INQUIRER (Sept. 13, 2017), https://www.theinquirer.net/inquirer/news/3017247/new-bluetooth-hack-doesnt-require-pairing-with-victims-device [https://perma.cc/AE75-73QV].

[268] *See, e.g.*, Mr. B, *Top 10 Most Popular Gifts for Serious Bacon Lovers!*, BACON TODAY (2015), https://bacontoday.com/top-10-most-popular-gifts-for-serious-bacon-lovers/ [https://perma.cc/X6L2-H3XC]. Bacon is a culinarily pleasing food for some diners. However, it does not carry equal utility in all implementation environments. Todd Van Luling & Renee Jacques, *The 17 Dumbest Things Vegetarians Have to Deal with,* HUFFPOST (Dec. 4, 2017, 9:59 AM), https://www.huffingtonpost.com/entry/vegetarians-dumbest-things_n_4177147.html [https://perma.cc/QMP6-B924].

[269] An example of the phenomenon is the idea that all devices are better with Bluetooth. *But see* Merriman, *supra* note 267 (stating Bluetooth is a notoriously vulnerable technology).

[270] *See* Abigail Williams, *This High-Tech Toaster Prints the Weather Report on Bread,* HUFFPOST (Aug. 16, 2016, 9:44 AM), https://www.huffingtonpost.com/entry/toaster-weather-forecast-toasteroid_n_57b30217e4b0a8e1502526a4 [https://perma.cc/VJ6E-FNGJ].

[271] *See* Andrea M. Matwyshyn, *The Big Security Mistakes Companies Make When Buying Tech,* WALL ST. J. (Mar. 13, 2017), https://www.wsj.com/articles/the-big-security-mistakes-companies-make-when-buying-tech-1489372011 [https://perma.cc/U3HG-J6BP].

[272] Indeed, Internet-connected ovens have already been known to suffer serious security vulnerabilities in their code. *See Security Flaw Could Have Let Hackers Turn on Smart Ovens*, PHYS.ORG (Oct. 26, 2017), https://phys.org/news/2017-10-flaw-hackers-smart-ovens.html [https://perma.cc/US9H-BHST].

whimsy-to-unreasonable security risk ratio should swiftly tilt the calculus in favor of choosing the non-IoT device.[273]

Again, IoT history offers a warning: in 2013, the technology press and the security research community accurately predicted that ransomware[274] would soon lock up computers at scale and that botnets would use IoT devices in denial of service attacks. Three years later, in 2016, a botnet of IoT devices committed a successful distributed denial of service attack against Twitter and Reddit, and entire hospital networks were crippled due to ransomware.[275] Today, security professionals are already warning that gratuitously connecting human bodies to the Internet will end even more poorly[276]—with botnets of bodyparts and human bodies immobilized by ransomware.[277] Yet, despite these credible and somber warnings, our overenthusiasm and magical thinking leads us to often gratuitously and unwisely connect devices to the Internet without fully considering the additional security risk. This blind overenthusiasm also begets our next problem—the "Magic Gadget" problem.

2. The Magic Gadget Problem: Failing to Anticipate Failure

In his book *Pinpoint*, author Greg Milner describes how, since the launch of the GPS system in 1980, humans have slid into overreliance and magical thinking about the trustworthiness of the

[273] *See* Robert Cottrell, *Why You Should Be Afraid of a Smart Toaster*, BBC FUTURE (Feb. 16, 2015), http://www.bbc.com/future/story/20150216-be-afraid-of-the-smart-toaster [https://perma.cc/G7VE-N5WG].

[274] *See* J.M. Porup, *Ransomware Is Coming to Medical Devices,* VICE MOTHERBOARD (Nov. 19, 2015, 6:00 AM), https://motherboard.vice.com/en_us/article/jpgxxk/ransomware-is-coming-to-medical-devices [https://perma.cc/7C4W-J3QN].

[275] *See infra* notes 297, 321-23 and accompanying text.

[276] One security professional has warned of the same possibility with IoB heart defibrillators. Chris Wysopal (@WeldPond), TWITTER (Oct. 24, 2016, 11:56 PM), https://twitter.com/WeldPond/status/790809257448972288 [https://perma.cc/C5UN-LMHN] ("What's next in 2017? Heart defibrillators joining in IoT DDoS attacks?").

[277] For example, one security professional warned of IoB breast pumps being compromised and used as part of a denial of service attack. *See* Alfredo Ortega (@ortegaalfredo), TWITTER (Jan. 5, 2017, 9:34 AM), https://twitter.com/ortegaalfredo/status/817031461878562816 [https://perma.cc/B6Y5-Y5SU] ("Botnets will get really weird this year."); *see also* Jeremiah Grossman (@jeremiahg), TWITTER (Oct. 19, 2016, 6:54 AM), https://twitter.com/jeremiahg/status/788739996739969024 [https://perma.cc/2XMC-8CJB] ("[B]ody implants are likely to be in our [near] future, so technically we're personally going to be IoT devices.").

technology.[278] While GPS has generally eased the struggles of mapping, in some cases, it has contributed to the untimely demise of its users—what he terms "death by GPS."[279] As users blindly trust the "magic" gadget in their hand, they sometimes disregard other superior sources of evidence in physical space,[280] even despite ample evidence that GPS can fail[281] or be manipulated by attackers.[282] This type of overly-optimistic IoT thinking might be termed the "Magic Gadget" problem.

Turning to IoB, the adventures of Professor Mann offer a cautionary tale. Professor Steve Mann has experimented with IoB technology through an auto-recording augmented reality "glass eye" technology[283] that is permanently attached to his head.[284] In 2012, Mann's IoB device was implicated in a physical altercation in a Paris restaurant.[285] Allegedly, the restaurant employees decided to aggressively enforce a "no camera" policy and attempted to remove Professor Mann's glass eye by force from his head.[286] This unexpected physical disruption to Mann's IoB device allegedly rendered it inoperable, partially due to a secondary, moisture-related,

[278] *See* GREG MILNER, PINPOINT: HOW GPS IS CHANGING TECHNOLOGY, CULTURE, AND OUR MINDS 112-15 (2016).

[279] *Id.*

[280] *See id.*

[281] Kristen Lee, *These Are Your Worst GPS-Fail Stories*, JALOPNIK (Sept. 12, 2017, 10:55 AM), https://jalopnik.com/these-are-your-worst-gps-fail-stories-1803140713 [https://perma.cc/987G-MXEB].

[282] Elias Groll, *Russia Is Tricking GPS to Protect Putin*, FOREIGN POL'Y (Apr. 3, 2019, 5:19 PM), https://foreignpolicy.com/2019/04/03/russia-is-tricking-gps-to-protect-putin/ [https://perma.cc/V5SC-KCMW].

[283] *See EyeTap: The Eye Itself as Display and Camera*, EYETAP.ORG, http://www.eyetap.org/research/eyetap.html [https://perma.cc/W3NR-THAK]. This IoB device is used by Mann partially to improve his night vision using lasers. *See* Jake Edmiston, *No Shirt, No Shoes, No Cyborgs: Toronto Prof Says He Was Roughed up at Paris McDonald's Over No-Camera Policy*, NAT'L POST (July 19, 2012, 1:23 AM), http://news.nationalpost.com/news/canada/no-shirt-no-shoes-no-cyborgs-toronto-prof-says-he-was-roughed-up-at-paris-mcdonalds [https://perma.cc/SL5F-XEV6].

[284] *See* Edminston, *supra* note 283.

[285] *See* Katie Daubs, *Toronto 'Cyborg' Steve Mann Says He Was Assaulted in Paris McDonald's*, STAR (July 18, 2012), https://www.thestar.com/news/gta/2012/07/18/toronto_cyborg_steve_mann_says_he_was_assaulted_in_paris_mcdonalds.html [https://perma.cc/5YD7-65WC] (noting that before the incident, another employee had accepted Mann as a customer, sold him food, and reviewed his doctor's note).

[286] *See* Edmiston, *supra* note 283.

hardware malfunction.[287] Professor Mann's experience offers us a reminder that catastrophic IoB failures will often happen unexpectedly and that they are not always within our control.

Yet, technology over-trust and the Magic Gadget problem often cause a failure to plan for even catastrophic failures. Indeed, a harbinger of these looming Magic Gadget problems in IoB might be found in the March 2016 ransomware attack that crippled the network of a Maryland hospital chain, impairing the patient care in 10 hospitals and 250 clinics.[288] With apparently no adequately robust crisis management system in place, employees described the attack as creating a "chaotic environment" and a "patient safety issue" that was potentially avoidable.[289] The hospitals had allegedly failed to patch vulnerabilities (that were well-known in the security community since 2007) despite direct prior warnings and the existence of techniques exploiting those unpatched vulnerabilities.[290] The Wannacry ransomware attack similarly paralyzed thousands of the United Kingdom's National Health Service administrative computers,[291] potentially contributing to physical harm of patients who were waiting on emergency surgeries and consultations.[292]

But now consider a version of these ransomware scenarios in which the targeted devices are patients' IoB artificial pancreases

[287] *Id.* ("Like I said, I had to go to the washroom.... But when he pushed me out the door, at some point my pants became a toilet.... Some of the critical items were affected (by water damage).").

[288] *See* John Woodrow Cox, *MedStar Health Turns Away Patients After Likely Ransomware Cyberattack*, WASH. POST (Mar. 29, 2016), https://www.washingtonpost.com/local/medstar-health-turns-away-patients-one-day-after-cyberattack-on-its-computers/2016/03/29/252626ae-f5bc-11e5-a3ce-f06b5ba21f33_story.html [https://perma.cc/E7P2-6JEU].

[289] *Id.*

[290] *See* Sean Gallagher, *Maryland Hospital: Ransomware Success Wasn't IT Department's Fault*, ARS TECHNICA (Apr. 7, 2016, 10:12 AM), https://arstechnica.com/information-technology/2016/04/maryland-hospital-group-denies-ignored-warnings-allowed-ransomware-attack/ [https://perma.cc/5643-HNA9].

[291] For a discussion of Wannacry, see Josh Fruhlinger, *What Is WannaCry Ransomware, How Does It Infect, and Who Was Responsible?*, CSO (Aug. 30, 2018, 6:52 AM), https://www.csoonline.com/article/3227906/what-is-wannacry-ransomware-how-does-it-infect-and-who-was-responsible.html [https://perma.cc/3U8S-96D5].

[292] *See* Owen Hughes, *WannaCry Impact on NHS Considerably Larger than Previously Suggested*, DIGITALHEALTH (Oct. 27, 2017), https://www.digitalhealth.net/2017/10/wannacry-impact-on-nhs-considerably-larger-than-previously-suggested/ [https://perma.cc/6Y8A-7A8Q] ("NHS England put the total number of cancelled appointments at some 19,494, which includes at least 139 patients who had 'an urgent referral for potential cancer cancelled.'").

instead of computer systems. Particularly, when we combine the Better with Bacon problem and the Magic Gadget problem with our next problem, the problem of Builder Bias, physical harm to IoB users becomes unfortunately entirely predictable and likely.[293]

3. The Builder Bias Problem: Shipping Without Securing

Consider the scenario in which a manufacturer with lax code-security practices has released a vulnerable IoB pancreas. The device's rampant security vulnerabilities allow for a remote attacker to disable it, demanding a "ransom" payment to turn it back on. Or imagine a botnet comprised of injected IoB eye lenses that cannot be easily removed. How would an average consumer respond when he learns that his eyeballs might be implicated in a distributed denial of service attack on a critical infrastructure target?

As a wisely-programmed computer once announced in the movie *War Games*, "[t]he only winning [strategy] is not to play."[294] The only viable answer to these IoB security failure scenarios lies in avoiding the problem from the outset—devices must be as secure as possible at the point of shipping. Yet, the lessons of IoT caution us that many builders of IoB will fail to build in line with what the FTC calls security by design.[295] As builders rush to ship code to market, they often fail to prioritize the security and consumer safety of their code.[296] IoT product manufacturers, in particular, have sometimes perceived themselves to have little financial incentive to prioritize security or to disclose and correct flaws,[297] and security errors in their products have frequently gone undetected.[298] For example, IoT

[293] Early adopter consumers seeking out "magic gadgets" may pay a heavier than anticipated price. *See* Charles Fain Lehman, *Experts Say Medical Care Next Big Threat,* FREEBEACON (Sept. 24, 2017, 5:00 AM), http://freebeacon.com/issues/experts-say-medical-care-next-big-cyber-threat/ [https://perma.cc/28RN-AQVP].

[294] *See WarGames (War Games) Quotes*, ROTTENTOMATOES, https://www.rottentomatoes.com/m/wargames/quotes/ [https://perma.cc/Y8S3-MMC2].

[295] START WITH SECURITY: A GUIDE FOR BUSINESS, U.S. FED. TRADE COMM'N, https://www.ftc.gov/tips-advice/business-center/guidance/start-security-guide-business [https://perma.cc/6277-TPX3].

[296] *See, e.g.*, Mike Lloyd, *The Internet of Things that Can Attack You,* FORBES (Feb. 17, 2017, 9:00 AM), https://www.forbes.com/sites/ciocentral/2017/02/17/the-internet-of-things-that-can-attack-you/171zb2dfedda [https://perma.cc/83W6-AX8R].

[297] *See id.*

[298] *See id.*

products are not always built to be updateable,[299] and attempts to report flaws by third parties sometimes result in receiving legal threats instead of thanks.[300] This failure of manufacturers to consider the implementation realities of security for fear it might delay shipping might be termed the problem of "Builder Bias."

Historically, IoB devices—just like IoT devices—have also been notoriously vulnerable to attacks by third parties due to imprudent security design choices such as hardcoded passwords.[301] In other words, the Builder Bias problem is already visible in IoB. For example, in 2012 when an episode of the television drama series Homeland included a plot twist where the sitting Vice President was murdered by a terrorist through a remote computer intrusion into his pacemaker,[302] the possibility of such a compromise was already well-recognized within the security research community.[303] In other words, the knowledge that IoB pacemakers could be remotely compromised by attackers existed years before the recent FDA IoB security recall.[304] Yet, despite this widespread knowledge, the medical device company that manufactured the pacemaker subject to the FDA security recall initially chose to deny the existence of a problem and to sue the security researcher who

[299] *See* Jason Perlow, *All Your IoT Devices Are Doomed*, ZDNET (July 12, 2016), https://www.zdnet.com/article/all-your-iot-devices-are-doomed/ [https://perma.cc/TDT7-2WUL].

[300] Zack Whittaker, *Lawsuits Threaten Infosec Research—Just When We Need It Most*, ZDNET (Feb. 19, 2018, 1:00 PM), https://www.zdnet.com/article/chilling-effect-lawsuits-threaten-security-research-need-it-most/ [https://perma.cc/B2B8-CNN7].

[301] NCCIC, MEDICAL DEVICES HARD-CODED PASSWORDS, U.S DEP'T HOMELAND SEC. (Oct. 29, 2013), https://ics-cert.us-cert.gov/alerts/ICS-ALERT-13-164-01 [https://perma.cc/BH42-MA77].

[302] Barbara Chai, *'Homeland,' Season 2, Episode 10, 'Broken Hearts': TV Recap,* WALL ST. J. (Dec. 2, 2012, 11:00 PM), http://blogs.wsj.com/speakeasy/2012/12/02/homeland-season-2-episode-10-broken-hearts-tv-recap/ [https://perma.cc/5Pe5-48ZV]; *see also* Barnaby Jack, *"Broken Hearts": How Plausible Was the Homeland Pacemaker Hack?*, IOACTIVE (Feb. 26, 2013), https://ioactive.com/broken-hearts-how-plausible-was-the-homeland-pacemaker-hack/ [https://perma.cc/C59K-DQWJ].

[303] *See* Tarun Wadhwa, *Yes, You Can Hack a Pacemaker (and Other Medical Devices Too)*, FORBES (Dec. 6, 2012, 8:31 AM), https://www.forbes.com/sites/singularity/2012/12/06/yes-you-can-hack-a-pacemaker-and-other-medical-devices-too/#6df879826853 [https://perma.cc/F92J-YL7P].

[304] *See* Jeremy Kirk, *Pacemaker Hack Can Deliver Deadly 830-Volt Jolt*, COMPUTERWORLD (Oct. 17, 2012, 1:40 AM), https://www.computerworld.com/article/2492453/malware-vulnerabilities/pacemaker-hack-can-deliver-deadly-830-volt-jolt.html [https://perma. cc/BX6W-WHLW].

identified the flaw in its product.[305] Further, based on security research and warnings issued by the FDA[306] and the Department of Homeland Security,[307] we know that a "[v]ast array of medical devices [are] vulnerable to serious" attacks due to unpatched vulnerabilities and products that ship vulnerable by default.[308] Apart from the predicted concerns of ransomware disabling IoB devices in extortion schemes and botnets of body parts attacking third parties,[309] the Builder Bias problem in IoB has already manifested itself by introducing novel national security risks. For example, inadequate security on the website of an IoB fitness tracking application recently disclosed the location of a previously unknown military base through leaked information about the presence of large numbers of human bodies attached to IoB devices.[310]

Brain interfaces in second- and third-generation IoB, in particular, present opportunities for malicious actors to potentially compromise bodies in order to obtain confidential information, such as passwords,[311] or—even more frighteningly—to corrupt the integrity or availability of the information residing in the brain hardware and, perhaps, even the functionality of the brain itself. Professor Jennifer Chandler and a team of coauthors raise concerns about the use of neuroprosthetic devices and the risk of "brainjacking"—the malicious manipulation of connected brain implants.[312] Similarly, Professors Tamara Bonaci, Ryan Calo, and Howard Jay Chizeck

[305] *See* Charlie Osborne, *MedSec Sued over St. Jude Pacemaker Vulnerability Report*, ZDNET (Sept. 8, 2016, 8:30 AM), http://www.zdnet.com/article/medsec-sued-over-st-jude-pacemaker-vulnerability-report/ [https://perma.cc/FPF3-K38R].

[306] *Cybersecurity*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/medicaldevices/digitalhealth/cybersecurity [https://perma.cc/WR6W-WZEU].

[307] *See* NCCIC, *supra* note 301.

[308] *See* Dan Goodin, *Vast Array of Medical Devices Vulnerable to Serious Hacks, Feds Warn,* ARS TECHNICA (June 13, 2013, 4:54 PM), https://arstechnica.com/information-technology/2013/06/vast-array-of-medical-devices-vulnerable-to-serious-hacks-feds-warn/ [https://perma.cc/8BZP-3RX6].

[309] *See supra* notes 273-76 and accompanying text.

[310] *See* Richard Pérez-Peña & Matthew Rosenberg, *Strava Fitness App Can Reveal Military Sites, Analysts Say,* N.Y. TIMES (Jan. 29, 2018), https://www.nytimes.com/2018/01/29/world/middleeast/ strava-heat-map.html [https://perma.cc/54TC-6YRR].

[311] *See* Tom Simonite, *Using Brainwaves to Guess Passwords*, MIT TECH. REV. (May 5, 2017), https://www.technologyreview.com/s/604293/using-brainwaves-to-guess-passwords/ [https://perma.cc/RYR9-4DLB].

[312] Clausen et al., *supra* note 241.

warn of the risk of "brain malware"[313] and the need for an interdisciplinary approach to addressing the development of attacks on brain-computer interfaces.[314] Through the eyes of a security professional, these compromised brains are not an "if," they are a certainty—a "when."[315] While these concerns may still be a few years away; lessons from IoT security remind us that the pace and severity of attacks generally escalate and outstrip our preparedness to address them.[316]

The three prior problems introduced above—the Better with Bacon problem of gratuitous connectivity, the Magic Gadget problem of the failure to anticipate failure, and the Builder Bias problem of shipping without securing—all converge to exacerbate the fourth problem—the problem of "Mandatory Soup."

4. The Mandatory Soup Problem: Diminishing Market Choice and Obsolescence Through Adhesion

In the opening episode of *Battlestar Galactica*, a war rages in the galaxy.[317] All of the most advanced military spaceships have been compromised by the enemy because they have been networked together and, therefore, are vulnerable to remote attack by the enemy.[318] Only one ship remains viable—Galactica.[319] It had been "airgapped"[320]—intentionally kept off the grid and disconnected from

[313] Victoria Turk, *How Hackers Could Get Inside Your Head with 'Brain Malware'*, VICE MOTHERBOARD (Aug. 3, 2016, 7:50 AM), https://motherboard.vice.com/en_us/article/ezp54e/how-hackers-could-get-inside-your-head-with-brain-malwareups [https://perma.cc/2Q3H-AN29].

[314] Tamara Bonaci et al., *App Stores for the Brain: Privacy & Security in Brain-Computer Interfaces*, IEEE TECH. & SOC'Y MAG., June 2015, at 2.

[315] Greg Conti (@cyberbgone), TWITTER (Mar. 27, 2017, 4:35 PM), https://twitter.com/cyberbgone/status/846505878824128512 [https://perma.cc/2PCL-JE6S] ("An entire neural malware & anti-malware field is waiting to happen. Imagine #ransomware & the RSA vendor floor then.").

[316] George Dvorsky, *How Will We Stop Hackers from Invading Our Brains Once We're Cyborgs?*, GIZMODO (June 29, 2017, 2:00 PM), https://gizmodo.com/how-will-we-stop-hackers-from-invading-our-brains-once-1796520628 [https://perma.cc/PUL8-PQY8].

[317] *Battlestar Gallactica: Episode #1.1*, IMDB, http://www.imdb.com/title/tt1699275/plotsummary?ref_=tt_ov_pl [https://perma.cc/RJ45-2VUJ].

[318] *Miniseries, Night 1*, FANDOM: BATTLESTAR GALACTICA WIKI, https://galactica.fandom.com/wiki/Miniseries,_Night_1 [https://perma.cc/DTW9-Q52W].

[319] *Id.*

[320] *See, e.g.*, Unknown Lamer, *Is Analog the Fix for Cyber Terrorism?,* SLASHDOT (Mar.

the other ships as an information security measure by its astute captain, Adama.[321] This plotline from a science fiction television show teaches us an often neglected but basic lesson about technology: the mere existence of a newer technology does not automatically make that new technology the better choice for a particular challenge.[322] This principle that the most connected device may not be the most appropriate device for a particular task might be termed the "Adama Principle."[323]

The Adama Principle is perhaps illustrated best by a famous IoB pacemaker story from 2007 involving former Vice President Dick Cheney. Six years before the compromised pacemaker episode on Homeland[324] aired, then-Vice President Dick Cheney was concerned that attackers would attempt to compromise his implanted defibrillator and kill him.[325] As a consequence, he asked his doctor to disable the device's wireless functionality.[326] But Cheney's leveraging of the Adama Principle is not the norm. Most consumers lack the necessary information regarding potential vulnerability of IoB to be able to make similarly informed choices about their bodies.[327]

Instead of the Adama Principle, what prevails in consumer markets is closer to what might be called the Mandatory Soup problem. Consider a guest at a set-menu wedding dinner. As hardworking servers distribute substantially identical meals to each diner, the opportunity for customization is minimal. As a consequence, a diner sometimes finds herself trapped behind a bowl of unwanted soup for

18, 2014, 12:01 AM), http://it.slashdot.org/story/14/03/18/021239/is-analog-the-fix-for-cyber-terrorism [https://perma.cc/6CV9-FZJ3].

[321] *See Miniseries, Night 1*, *supra* note 318. Captain Adama knew that the enemy—the cylons—were masters at disabling battlestars by breaking into networks via wireless networks and then using them to disable the whole ship and as a consequence, he ordered that his ship never be networked. *See id.*

[322] Michael C. Bodson, *The Latest, Shiniest New Technology Isn't Always the Best*, WORLD ECON. FORUM (Jan. 15, 2018), https://www.weforum.org/agenda/2018/01/why-the-latest-shiniest-tech-isn't-always-best/ [https://perma.cc/MPX7-4B7K].

[323] For a different but related version of this idea, see Raza Panjwani (@occamsraza), TWITTER (June 27, 2018, 7:40 AM), https://twitter.com/occamsraza/status/1011982726113759232 [https://perma.cc/68D7-8GWV].

[324] Chai, *supra* note 302.

[325] Bob Fredericks, *Cheney Feared Terrorists Would 'Hack' Pacemaker*, N.Y. POST (Oct. 19, 2013, 4:11 AM), http://nypost.com/2013/10/19/cheney-feared-heart-gizmo-hack-attack/ [https://perma.cc/4G2Q-NUZK].

[326] *Id.*

[327] Horrigan, *supra* note 156.

a period of time. While other people may want the soup, she does not, and she experiences negative consequences because it has been foisted upon her. For example, the soup blocks her ability to use the plate beneath the bowl, and it inhibits her streamlined access to the wine in the middle of the table. It also places her at unnecessary risk of soup-related sartorial catastrophe.

The consumer marketplace is becoming flooded with a bevy of "Mandatory Soup" IoT products, often making less-connected versions of those same products nonexistent or hard to find.[328] Rather than maximizing competition on *degree* of connectedness as a differentiating factor within individual product lines, we instead see a progressively impoverished marketplace with consumer products tending to default in their evolution to the maximum degree of connectedness.[329] The Adama Principle of selecting the less connected option when appropriate becomes functionally impossible without extraordinary effort in this type of impoverished artificially constrained marketplace. For example, examining the new car market, finding a new car without multiple accompanying end user license agreements, always on location tracking, and several million lines of code is quickly becoming an impossible task.[330]

Indeed, not only is market choice becoming impoverished on the degree of product in connectedness, but the "real" price of competing goods with the same level of connectedness is becoming incomparable to an average consumer at the time of purchase. Material differences in future obsolescence and data stewardship are usually not disclosed at the time of purchase and not predictable for consumers.[331] Thus, the actual cost of product ownership across time for consumers of an IoT device is frequently not accurately calculable

[328] *See* David Roe, *7 Big Problems with the Internet of Things*, CMS WIRE (Feb. 7, 2018), https://www.cmswire.com/cms/internet-of-things/7-big-problems-with-the-internet-of-things-024571.php [https://perma.cc/P6EB-KSYZ].

[329] *See id.*

[330] *See* Julie A. Steinberg, *Fifty Billion Connected Devices Bring Tort, Software Law Clash*, BLOOMBERG L. (Feb. 26, 2016), https://www.bna.com/fifty-billion-connected-n57982067832/ [https://perma.cc/N4QC-TE7W]. Cars are now functionally IoT devices on wheels—whether consumers desire this extreme connectivity and code-reliance or not. *See* Andrea M. Matwyshyn, *CYBER!*, 2017 BYU L. REV. 1109, 1141-42 (2017).

[331] *See* David Gewirtz, *Revolv is Dead. Google Killed It. Long Live Innovation*, ZDNET (June 20, 2016, 3:19 PM), http://www.zdnet.com/article/revolv-is-dead-google-killed-it-long-live-innovation/ [https://perma.cc/X9GA-UHM4].

at the point of purchase.[332] A consumer might choose away from one particular product knowing that the expected life is five years shorter than that of another (superficially) competitively-priced product. Thus, the Mandatory Soup problem exposes consumers to the undisclosed price terms of (planned and unplanned) unilateral manufacturer obsolescence determinations and data handling changes—rights reserved in the terms of the accompanying (and evolving) end-user license agreements (EULA).[333] This dynamic might be called "obsolescence through adhesion."[334]

IoT history again provides warnings about the hidden costs of Mandatory Soup and, in particular, obsolescence through adhesion. In 2014, Nest acquired an IoT start up called Revolv, a company that made a smart home hub intended to control devices such as lights, alarms, and doors.[335] Allegedly because of an "allocat[ion] [of] resources,"[336] Revolv announced that its service would shut down and customers' applications would no longer work.[337] In short, customers who had purchased the Revolv hub were informed, much to their surprise and dismay, that they would be left with a "bricked" device, regardless of what the company's promises or customers' reasonable expectations were at the time of purchase.[338] This Revolv

[332] *See id.*

[333] As such, it might be argued that in egregious cases, undisclosed hidden costs of planned obsolescence amount to an unfair trade practice under Section 5 of the Federal Trade Commission Act, warping competition in the marketplace of IoT products. *A Brief Overview of the Federal Trade Commission's Investigative and Law Enforcement Authority*, FED. TRADE COMM'N (July 2008), https://www.ftc.gov/about-ftc/what-we-do/enforcement-authority [https://perma.cc/44NZ-N37A].

[334] "Obsolescence through adhesion" refers to the combination of technical and contract-based measures in technology products that functionally hide the actual cost of ownership and can discretionarily force a consumer to discontinue use of a particular product, attempting to nudge the consumer into a new purchase. *See* Andrea Matwyshyn (@amatwyshyn), TWITTER (Mar. 7, 2018, 10:39 AM), https://twitter.com/amatwyshyn/status/971424989063925760 [https://perma.cc/S2VP-YHX4].

[335] Nick Statt, *Nest Is Permanently Disabling the Revolv Smart Home Hub*, VERGE (Apr. 4, 2016, 3:40 PM), https://www.theverge.com/2016/4/4/11362928/google-nest-revolv-shutdown-smart-home-products [https://perma.cc/Q9RU-b95W].

[336] *Id.*

[337] Alex Hern, *Revolv Devices Bricked as Google's Nest Shuts Down Smart Home Company*, GUARDIAN (Apr. 5, 2016, 5:04 AM), https://www.theguardian.com/technology/2016/apr/05/revolv-devices-bricked-google-nest-smart-home [https://perma.cc/CQN8-YL3G].

[338] Chris Hoffman, *What Does "Bricking" a Device Mean?*, HOW-TO GEEK (Sept. 26, 2016, 5:36 PM), https://www.howtogeek.com/126665/htg-explains-what-does-bricking-a-device-mean/ [https://perma.cc/L9P6-7GF3].

IoT incident highlights that consumers may not realize that IoT products are now functionally software products still tethered to the manufacturer through remote updates, despite their physicality.

Now, let us turn to the IoB context. Particularly when the obsolescence through adhesion dynamic relates to IoB security and future patches, consumers will find themselves in a dangerous lose-lose scenario that puts them at increased risk of physical harm.[339] Imagine that your eyeball-injected IoB contact lens provider informs you that (per the terms of the contract on which you clicked "yes" when you downloaded your lens software), it has decided that it will no longer support the version of the software your lenses run and that it will no longer push out security patches for your eyes. None of your options are good in this scenario. You can get your lenses removed, risking physical harm and absorbing the cost. You can keep your lenses, knowing they are no longer supported, which, in turn, exposes you to different physical risks through malfunction or security compromise. Alternatively, you can buy "upgraded" lenses, absorbing those associated risks and costs. In all cases, the IoB manufacturer has contractually and technically forced an "upgrade" onto the body of the consumer.

While each of these four problems—the Better with Bacon problem, the Magic Gadget problem, the Builder Bias problem, and the Mandatory Soup problem—is independently a point of concern, when taken together in the context of some second- and third-generation IoB, the risks they present transform into a significant threat in the aggregate—the threat of physical harm to human bodies.[340] Indeed, second-generation IoB presents obvious corporeal risks,[341] while third-generation IoB presents the risk not only of losing control over our own bodies but also our cognitive processing.[342] Put another way, third-generation IoB impacts our functional freedom of thought, and, as a consequence, it presents the threat of potentially losing control over the deliberative individual processes on which we implicitly rely not only for governance of our bodies but

[339] Andrea M. Matwyshyn, *The 'Internet of Bodies' Is Here. Are Courts and Regulators Ready?*, WALL ST. J. (Nov. 12, 2018, 11:19 AM), https://www.wsj.com/articles/the-internet-of-bodies-is-here-are-courts-and-regulators-ready-1542039566 [https:// perma.cc/XAD8-TJYM].

[340] *See id.*

[341] *See supra* Part I.A.2.

[342] *See supra* Part I.A.3.

also for self-governance in a democratic society.[343] In light of the gravity of these risks, let us consider the current state of IoT and IoB regulation and the desirable directions for its evolution.

C. The Future of Corporate Software Liability and IoB

Morpheus: What is real? How do you define 'real'? If you're talking about what you can feel, what you can smell, what you can taste and see, then 'real' is simply electrical signals interpreted by your brain.[344]

At present, third-generation IoB's risks are (mostly) not yet upon us,[345] but the challenges presented by second-generation IoB are already present and escalating.[346] While last century's legal approaches to technology were animated by a principle of avoiding the imposition of software liability in the name of innovation,[347] IoB forces a recalibration of this default. As human bodies become regularly physically harmed by computer code, consumer and market trust in technology will wane without buttressed legal baselines of consumer protection.[348] Indeed, recent survey data warns that this consumer trust breakdown is already in progress: growing numbers of consumers are doubting whether the Internet has been a mostly positive development for society.[349]

Bolstering consumer trust in technology and constructing an innovation-sensitive legal framework for IoB begins with correcting the legacy problems of IoT identified in Part I.B—the residual deficits in consumer protection from IoT that are already transferring

[343] *See infra* notes 565-70 and accompanying text.

[344] *The Matrix*, *supra* note 2.

[345] *See supra* notes 246-49 and accompanying text.

[346] *See supra* Part I.A.2.

[347] *See supra* pp. 117-18.

[348] For one novel model of creating corporate duties of care in technology conduct, see Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1186, 1209 (2016) (arguing "many online service providers and cloud companies ... should be seen as information fiduciaries toward their customers and end users" because by virtue of "their relationship with another," they have assumed "special duties with respect to the information they obtain in the course of the relationship").

[349] *See* Pew Research Internet (@pewinternet), TWITTER (May 24, 2018, 7:32 AM), https://twitter.com/pewinternet/status/999659285406765057 [https://perma.cc/XA74-HHJ6].

URBAN OMNIBUS

Digital Frictions

Disruption at the Doorstep

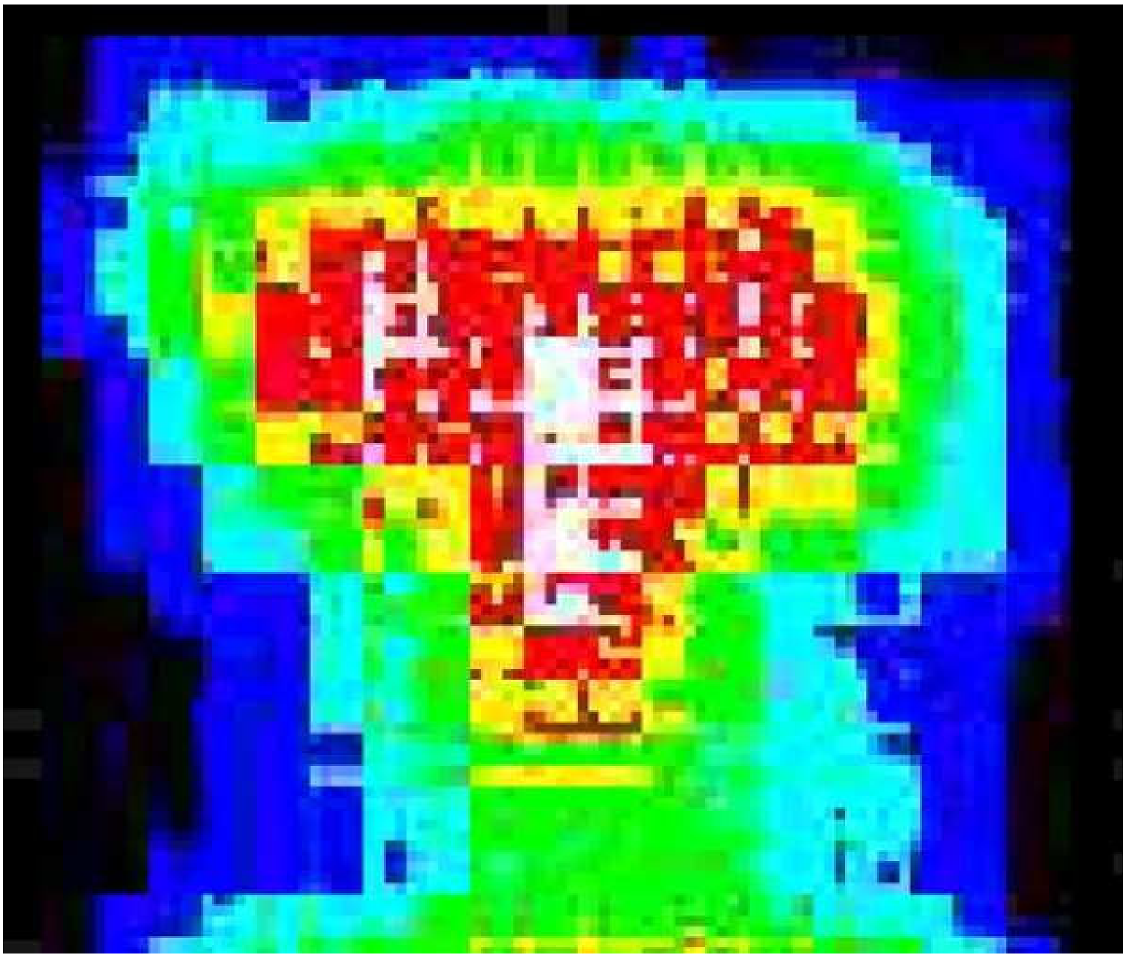
Erin McElroy

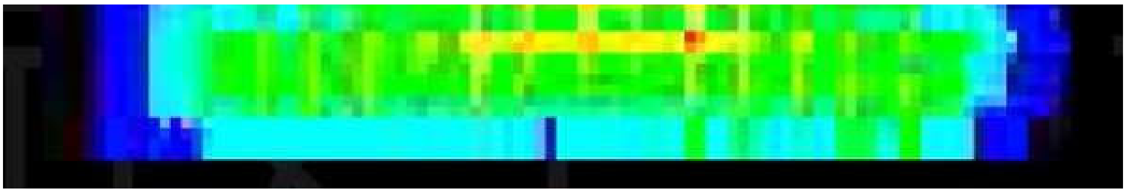
Nov 06, 2019

The entrance to Atlantic Plaza Towers. Photo by Amy Howden-Chapman

A new assemblage of digital tools promises to reduce friction in the management of urban real estate — to help prospective renters virtually experience potential units without the hassle of trekking all over the city; to help landlords ensure that those potential renters will be good neighbors and responsible tenants; and to enable those new renters, their arms encumbered with moving boxes, to glide right through their apartment buildings' secure entries. Property technology, or "proptech," promises easy passage through, and close monitoring of, buildings networked via sensors and cameras to biometric databases and screening platforms. Yet as **Erin McElroy** demonstrates, proptech's promises of security and convenience tend not to apply to poor and working-class tenants of color, who are instead finding themselves targeted by what are fast becoming new instruments of surveillance and harassment in housing complexes across New York City. – SM

In 2018, many of the rent-stabilized tenants at the two-building, 718-unit Atlantic Plaza Towers received notice from their landlord, Nelson Management, that their wireless key-fob entrance system would be replaced with biometric facial recognition technology. Even though they had been required to submit photos of themselves in order to obtain fobs in the first place — and despite the presence of other surveillance systems throughout the complex, including multiple CCTV cameras — tenants were informed that this new system would ensure their safety by keeping keys out of the hands of "the wrong people." Marketed as the True Frictionless™ Solution, this new facial recognition system was developed by the Kansas-based company StoneLock, which serves up to 40 percent of Fortune 100 companies, along with several government entities. This is the first publicly known instance of StoneLock endeavoring to deploy its facial recognition product in a New York City housing complex, though biometric technology, developed by other companies, has already been installed in residences throughout the city, such as the Knickerbocker Village affordable housing complex in the Lower East Side. Not coincidentally, StoneLock's first foray into residential facial recognition will be put to use in surveilling predominantly Black women tenants, many of whom have questioned Nelson Management's decision to test the company's product in in Brownsville, Brooklyn, as opposed to one of the landlord's other properties in a more affluent area of the city.

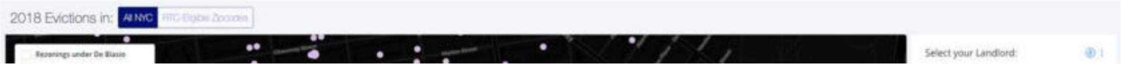


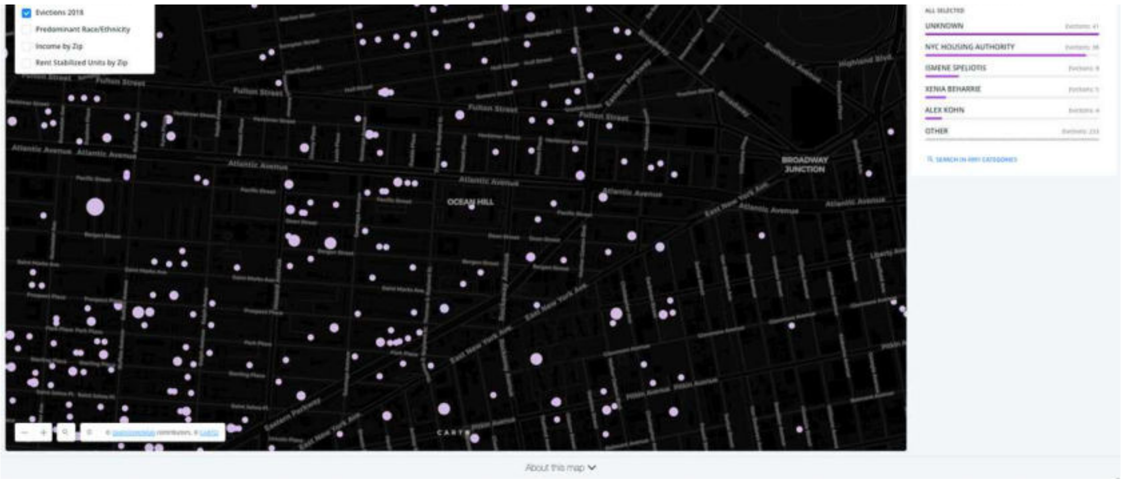


Screenshot of a proprietary, facial recognition "Identity Heatmap" created by Stonelock's True Frictionless™ Solution. The product uses "near-infrared wavelengths… invisible to the eye… to generate unique biometric metadata" displayed as an anonymized multi-colored figure.

Like dozens of other surveillance systems being rolled out in multifamily residences, commercial buildings, and industrial complexes globally, StoneLock's True Frictionless™ Solution is part of a burgeoning class of property technology, or proptech for short (also called real estate technology, or realtech). The last several years have seen a proliferation of proptech companies and platforms reshaping multiple domains of urban life, including the provision, consumption, and management of residential space. Often, proptech entails some configuration of artificial intelligence (AI), the Internet of Things (IoT), machine learning, user dashboards, software, data harvesting, and hardware. It can be difficult to categorize its multiple genres, particularly as many are combined, but proptech can be roughly taxonomized as rental housing management (tenant screening, payment, and maintenance), smart home development, keyless entry surveillance systems, sharing economy platforms, virtual reality-based home sales and rentals, tenant matching, and property database platforms. While aligned with "smart city" rhetoric, proptech makes explicit that private property relations are at the heart of its technological innovations, with companies in this sector catering to landlords (both private and public) who seek to automate the management of their portfolios.

My interest in these emerging technologies, and their often-negative impacts, comes out of longstanding tenant organizing efforts that I have been involved with in the San Francisco Bay Area and in Romania. It is also inspired by research undertaken by a digital cartography collective, the Anti-Eviction Mapping Project (AEMP), which I cofounded in 2013 in San Francisco (and which now maintains chapters in New York City and Los Angeles as well). Across multiple cities, I have witnessed and analyzed how real estate and technology platforms often work in conjunction to displace and target poor and working-class tenants of color. Proptech extends this tendency, while also signaling the merger of two leading global industries, Big Tech and Real Estate, that hinge upon the accumulation of property — data and land, respectively. Proptech collapses these two property regimes, leading to the heightened dispossession of people long targeted by both.

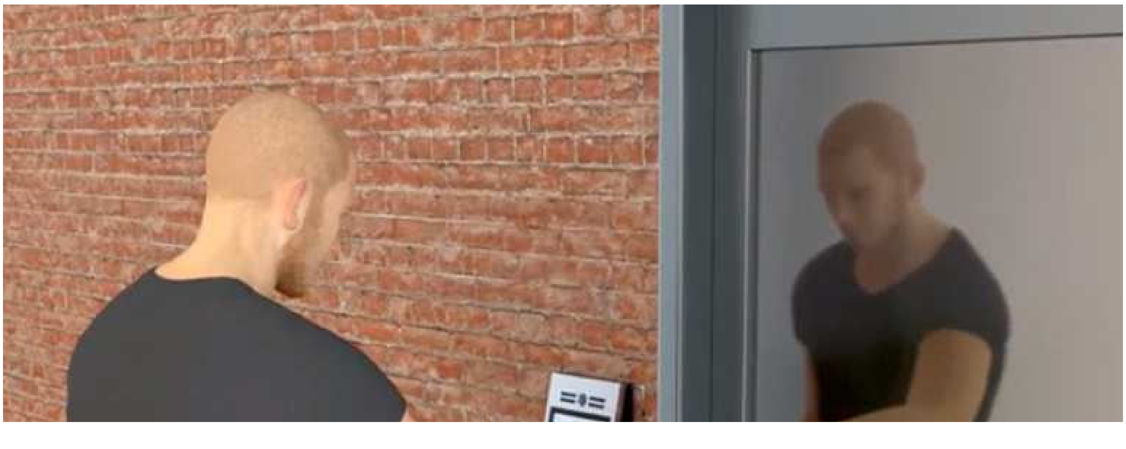




Detail of Brownsville from the New York City's Worst Evictors Map produced by the Anti-Eviction Mapping Project in partnership with The Right to Counsel NYC Coalition and JustFix.nyc.

Frictionless Fictions

At a recent proptech conference I attended near Wall Street, aficionados of the technology used the word "friction" a dozen times, always likening it to a slowness or hindrance to be overcome through technical means. "Frictionlessness," on the other hand, implies ease, cost cutting, and the ability to capitalize upon the consumer desires of young, affluent people — for instance, smart buildings, fast internet, and integration with delivery services. As Robert Nelson (the owner of Nelson Management and a self-described "tech geek") proclaimed to his tenants by way of a flyer: "Your daily access experience will be frictionless, meaning you touch nothing and show only your face. From now on the doorway will just recognize you!" And as StoneLock advertises, their products (including, of course, the True Frictionless™ Solution) will provide users with "frictionless access," so that they "just GO!" A key corollary to "frictionlessness" in proptech parlance is "safety." Ari Teman, a former comedian and engineer who created the "virtual doorman" system GateGuard, which utilizes facial recognition, has defended proptech by advocating that "surveillance can make you feel safe," and suggested that his product enables tenants to keep illegal subletters and unwelcome people from entering their building.



Screenshot from an animated promotional video for GateGuard.

Yet proptech's "frictionless" and "safe" qualities come at the cost of accountability to its users and the public. Beyond causing discomfort and concerns about privacy, proptech presents novel threats to the safety and stability of vulnerable tenants. GateGuard has been installed in roughly 1,000 residential buildings throughout New York City, although Teman (like other proptech developers) has refused to publicize these locations. Likewise, during a City Council hearing I attended in October 2019, city agencies (including the Department of Information Technology, Department of Buildings, and Department of Consumer and Worker Protection) claimed to have neither any knowledge of where residential facial recognition is installed, nor the capacity to map it. And though the purpose of this hearing was to discuss requirements for businesses and residences to disclose their use of "biometric identifier technology," and to provide physical keys to tenants if requested, it has become clear that while proptech maintains an opaque public profile, most people monitored by these technologies don't get the option of consenting to being a test subject.

The Knickerbocker Village housing complex, which adopted a facial recognition entrance system in 2014. Photo by jqpubliq, via Wikimedia Commons

In 2014, entrances throughout the twelve-building Knickerbocker Village complex — regulated by New York State Homes and Community Renewal (HCR), the state's affordable housing agency — were outfitted with FST21 SafeRise, a facial recognition product created by former Israel Defense Forces Major, General Aharon Farkash. During the City Council's hearing, one Knickerbocker Village tenant named Christina Zhang recounted stories of how this system has been implemented in her building. For one, the technology often simply fails to work properly, forcing tenants to line up and dance in front of the cameras, hoping that their movements will inspire recognition. But of greater concern, the complex's tenants — 70 percent of whom are Asian, and many of whom are immigrants — have no idea what the data being collected from them is used for, and have expressed fears for their biometric information ending up in the hands of the NYPD or ICE, both of which are known to use facial recognition and surveillance technologies to identify suspects and track undocumented people.

Tranae Moran, a tenant activist from Atlantic Plaza Towers, at the City Council's October 2019 hearing on facial recognition and biometric data collection. Photo by Erin McElroy

Several tenants from Atlantic Plaza Towers testified at this hearing as well, voicing unease about proptech's harvesting of biometric data. Yet efforts to organize against the use of facial recognition had been brewing in the Brownsville complex for over a year. In 2018, when Nelson Management announced the installation of the True Frictionless™ Solution via paper mailings, many tenants were left in the dark. A prior stipulation had required residents be photographed in order to receive mailbox keys, and some had refused this request. In response, five tenants, all Black women, convened in the lobby of one of the buildings on an October morning to inform their neighbors in person about StoneLock's system. Soon after, these women received notice from Nelson that they had been recorded by the lobby's half-dozen, 360-degree cameras, and were incorrectly informed that their "loitering" was illegal and would have to stop. A follow-up announcement of the impending facial recognition system was then made public, listing the name of every tenant and their apartment number in the building. One tenant named Anita, who had been flyering in her building, decried this incident during the City Council hearing: "Privacy be damned!"

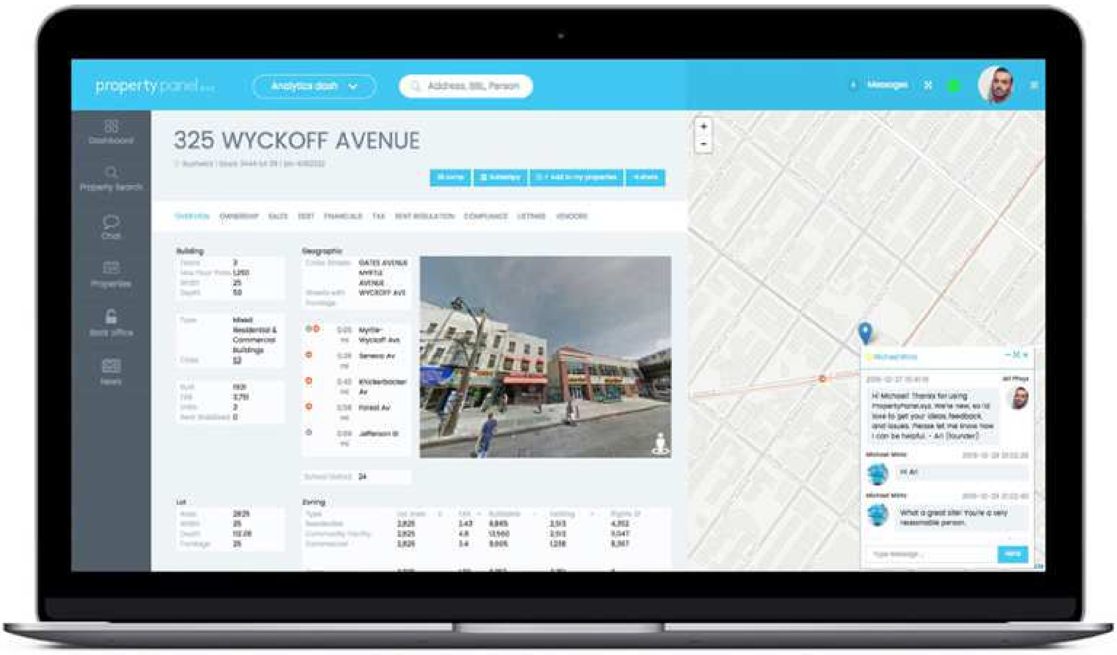
Resident Fabian Rogers, who has been on the frontlines of the campaign against facial recognition, also described his experience living with the technology: "I had many concerns as a tenant, and security was not one of them. It was the landlord's concern, and it was imposed on me. I already feel well enough surveilled with all the cameras and key fobs that exist. I kind of feel like a criminal even though I pay my rent on time." Beyond taking issue with the security theater of facial recognition, tenants expressed little faith in Nelson's claim that their biometric data would remain secure, and speculated that, if facing threat of eviction, this information could be potentially used against them in housing court. For these reasons — and because StoneLock's system will give the company access, without tenants' consent, to nearly 5,000 new faces with which test its algorithms — Atlantic Plaza Towers tenants have been advocating for a ban on facial recognition in the city altogether.

Fabian Rogers and other Atlantic Plaza tenants testifying at the City Council's hearing on facial recognition and biometric data collection. Photo by Erin McElroy

These tenants realize the technological changes being imposed upon them are not for their benefit, but for the "frictionless" experiences and "safety" of future gentrifiers yet to arrive. Atlantic Plaza Towers was built for middle-income families as part of the state-run Mitchell-Lama program in the 1950s, and it remains relatively affordable today, having just gained rent stabilization status two years ago. Yet with the ongoing gentrification of Brownsville and increased number of evictions, affordable housing and services are increasingly harder to come by. As Anita observed, "Tenants have so many issues that need to be addressed, but now we're dealing with this . . . So poor people like me can't live here anymore. I'm pissed at what's going on. So many people in the neighborhood are being pushed out . . . Please consider this a tragedy waiting to happen."

The Property of Data

While the struggle at Atlantic Plaza Towers has drawn public attention to how proptech can amplify tenant insecurity, the database systems that proptech hardware feeds into and supplies with new information, biometric and otherwise, remain obscure. For instance, Ari Teman's GateGuard has the ability to integrate with one of his company's other products, PropertyPanel.xyz, a proprietary database and dashboard platform for New York City landlords to use in acquisitions and property management. PropertyPanel.xyz allows purchasers to gather an array of information about buildings, and to "target" properties based upon value, debt, rent stabilization, ownership, air rights, size, and other criteria. Purchasers can also obtain alerts to violations and complaints, communicate with building staff, screen vendors, and are given the option to integrate PropertyPanel.xyz with yet another Teman product, SubletSpy, which monitors Airbnb tenants for potential infractions. Upon purchase of GateGuard, landlords and property managers consent to Teman accessing "any property of yours, digital or real world, in any method, for any purpose," including for the purposes of plugging data into PropertyPanel.xyz. No other clarifying information is given as to how this data may be used, or how it might facilitate the training of biometric algorithms.



Hi Michael! Thanks for using PropertyPanel.xyz. We're new, so I'd love to get your ideas, feedback, and issues. Please let me know how I can be helpful. - Ari (founder)

Hi Ari

What a great site! You're a very reasonable person.

Screenshot of a PropertyPanel.xyz dashboard demo.

Before inventing this suite of interconnected proptech products, Teman first created the startup Friend or Fraud Inc., which developed software to verify internet users' identities through video-analyzing machine vision, replete with breath and heartrate monitors. Today, he employs a team of workers across the US, Israel, and Eastern Europe who assist 100 landlords and property management companies in New York City, Miami, Chicago, and Los Angeles. Teman first invented SubletSpy in 2014 following an Airbnb experience in which, after renting his New York apartment on the platform, he returned to find the remnants of a well-attended sex party. Teman filed a complaint to Airbnb, but rather than resolving the matter, this action landed him on a "bad tenant" database, making it nearly impossible to find a new apartment in the city.

These inscrutable databases are often compiled by third-party "data brokers," who supply a vast number of individuals' personal information to landlords, marketers, and government agencies. In some instances, these brokers operate public platforms such as MyLife.com™ which gathers information from Facebook, LinkedIn, Twitter, Gmail, Yahoo, AOL, Outlook, school yearbooks, Ancestry.com and more to assign "reputation scores" to a claimed 325 million "verified identities." Other prominent data brokers such as Oracle, Experian, and Equifax buy and sell personal information related to a renter's credit history. Recently, it was revealed that Experian offered to raise users' FICO scores in exchange for credit card passwords, allowing the company to scan a user's purchase history into their databases and sell this information to third parties. Not only is credit reporting often discriminatory (particularly in regards to mortgage lending and rental payment history — an issue amplified during the 2008 subprime crisis), but like information brokerage at large, it alienates and reduces individuals into disaggregated data points. There have also been numerous instances of personal and biometric data being sold (and occasionally hacked) by third parties without consent or even the knowledge of the individual supplying the data.

Proptech companies such as Avail and Cozy (both marketed to small-scale, individual landlords) have entered into this space as well, developing digital products and platforms for the express purpose of screening tenants. CoreLogic, headquartered in California, has developed one of the most comprehensive residential database and tenant screening systems, with records spanning 50 years, 145 million parcels, and 99.9 percent of US property records. Their access to arrest records spans over 70 percent of the US's population centers, and interfaces with law enforcement agencies throughout the country. Updated every 15 minutes, this system includes over 80 million booking and incarceration records from roughly 2,000 facilities. CoreLogic also sources and returns data from the FBI and other federal agencies, promising to enable landlords in identifying "terrorists." Furthermore, the company's Registry CrimSAFE product bundle advertises its ability to seamlessly implement landlord policies, and "optimize" Fair Housing compliance. Yet, since releasing CrimSAFE, CoreLogic has been faced with a lawsuit over its algorithm which, according to the Connecticut Fair Housing Center, "disproportionately disqualifies African Americans and Latinos." Meanwhile, much of the data being mined by CoreLogic, especially that maintained by law enforcement, is plagued with inaccuracies and racial biases.

While it is unclear exactly which database system Teman appeared on, it is clear that, as a wealthy white man, he is not the type of person generally profiled by proptech database bundles. But in response to his own blacklisting, rather than getting involved in racial or data justice work, Teman chose instead to invent SubletSpy. Thus, Ari Teman, himself once a victim of proptech platforms and databases, weaponized tenant profiling technology for his own personal gain.

The Struggle Against Data-Driven Discrimination

While they may not have Teman's resources, housing and technology justice advocates and their allies have been pushing back against proptech's embedded biases and racist effects. In May 2019, Brooklyn Legal Services (BLS), a legal non-profit representing Atlantic Plaza tenants in their struggle against surveillance, composed a letter to HCR noting that facial recognition indicates a dramatic shift in Nelson's prior practices of landlordism. One BLS lawyer, Samar Katnani, has further argued that "the ability to enter your home should not be conditioned on the surrender of your biometric data, particularly when the landlord's collection, storage, and use of such data is untested and unregulated. . . We are in uncharted waters with the use of facial recognition technology in residential spaces."

Algorithmic Discrimination
"Facial recognition systems are trained with biased data."

A group of tenants from Atlantic Plaza Towers and BLS lawyers at the City Council hearing on facial recognition and biometric data collection. Photo by Erin McElroy

Meanwhile, scholars at New York University's AI Now Institute (an interdisciplinary research center dedicated to studying the social implications of advanced technical systems, of which I am a part) also wrote an expert letter in support of the tenants, describing how facial recognition systems for residential entry are bound to fail in accurately identifying tenants of color, women, and gender minorities. Numerous studies have shown that machine learning algorithms, disproportionately built and trained by white men, discriminate on the basis of gender and race, with women of color misclassified with error rates of nearly 40 percent, compared to one to five percent for white men. AI systems, particularly for facial recognition tools, rely upon machine learning algorithms trained with data. Physiognomic labels related to hair, skin, facial structure, and more are codified into racial and gender classifications, echoing 19th-century, pseudoscientific ideas about race and eugenics. And while facial recognition algorithms have been shown to be largely inaccurate in identifying women and Black people, it is still Black people being targeted most by them, and stopped and subjected to searches in facial recognition databases by police, often resulting in false positive identifications. This has led cities such as San Francisco, Oakland, and Somerville to recently ban the use of facial recognition by government agencies altogether.

The pressure applied by Atlantic Plaza Towers tenants has helped paved the way for Brooklyn-based Congresswoman Yvette Clarke to introduce a bill in the US House of Representatives named "No Biometric Barriers to Housing Act" that would prohibit facial, voice, fingerprint, and DNA identification technologies in public housing. The bill would also require the US Department of Housing and Urban Development (HUD) to report on biometric systems used in federally-assisted public housing in the last five years. Despite these potential legislative gains for public housing tenants, HUD has recently proposed alarming alterations to the 1968 Fair Housing Act (FHA) — an offspring of the civil rights era outlawing housing discrimination against people of color. The Fair Housing Act requires local governments that receive HUD funding to address segregation, disinvestment, and displacement in their communities. But as investigative reporters Aaron Glantz and Emmanuel Martinez write, under new proposed regulations, a company accused of discrimination in housing "would be able to 'defeat' that claim if an algorithm is involved." In this way, the "gentrifier-in-chief," president of the "real estate state," has made himself the new vanguard of racist proptech algorithms.

Following the October 2019 City Council hearing on facial recognition, New York City agencies may implement changes to mitigate facial recognition's impacts. But as the tenants of Atlantic Plaza Towers eloquently made clear, their demand is not for band-aid mitigations, but for a ban on facial recognition in New York City. The presence of already-existing security measures made tenants in Atlantic Plaza Towers feel policed in their homes long before their landlord introduced the possibility of facial recognition. Against the backdrop of gentrification, the insinuation of criminality and evictability are often used to maintain what Brenna Bhandar (citing legal scholar Cheryl Harris) calls the "whiteness of property." According to Bhandar, the ownership of property in places marked by the histories of colonialism and the slave trade is rooted in a "modern racial regime" dependent upon the dispossession of real property and data. Proptech has the potential to accelerate both forms of dispossession through the non-consensual mining of, and capitalizing upon, people's intimate data — what can be described as "data colonialism" — which updates processes of settler colonialism and occupation for the digital age. Yet it also sits upon a thick palimpsest of older property schemas and the information systems supporting these regimes. One could trace proptech's lineage back to the earliest moments of settler colonialism and the technologies it employed to gather data, map land, and dispossess Indigenous populations, or more recently to the redlining of communities of color during the mid-20th century. Given the potential for racist abuses in the future, and the inability of the city, landlords, and proptech companies to make transparent where and how new, "frictionless" tools function, a ban is the only just future — friction-filled as it may be.

> This research has received support from the British Academy's Tackling the UK's International Challenges program. The author would like to thank the brilliant tenants of Atlantic Plaza Towers for their organizing work and analysis, and the AI Now Institute and Rashida Richardson for feedback and support in writing this article.

Erin McElroy is a postdoctoral researcher at New York University's AI Now Institute focusing on proptech and technologies of gentrification. Erin is also a cofounder of both the Anti-Eviction Mapping Project and the Radical Housing Journal. Erin received a PhD in Feminist Studies from the University of California, Santa Cruz for a project on race, technology, and dispossession in postsocialist Romania, and continues to organize with housing justice collectives in both the US and Romania today.

The Guardian

Automating poverty

Digital dystopia: how algorithms punish the poor

https://www.theguardian.com/technology/2019/oct/14/automating-poverty-algorithms-punish-poor

In an exclusive global series, the Guardian lays bare the tech revolution transforming the welfare system worldwide – while penalising the most vulnerable

by Ed Pilkington in New York

Mon 14 Oct 2019 10.00 BST Last modified on Mon 14 Oct 2019 20.25 BST

All around the world, from small-town Illinois in the US to Rochdale in England, from Perth, Australia, to Dumka in northern India, a revolution is under way in how governments treat the poor.

You can't see it happening, and may have heard nothing about it. It's being planned by engineers and coders behind closed doors, in secure government locations far from public view.

Only mathematicians and computer scientists fully understand the sea change, powered as it is by artificial intelligence (AI), predictive algorithms, risk modeling and biometrics. But if you are one of the millions of vulnerable people at the receiving end of the radical reshaping of welfare benefits, you know it is real and that its consequences can be serious – even deadly.

The Guardian has spent the past three months investigating how billions are being poured into AI innovations that are explosively recasting how low-income people interact with the state. Together, our reporters in the US, Britain, India and Australia have explored what amounts to the birth of the digital welfare state.

Their dispatches reveal how unemployment benefits, child support, housing and food subsidies and much more are being scrambled online. Vast sums are being spent by governments across the industrialized and developing worlds on automating poverty and in the process, turning the needs of vulnerable citizens into numbers, replacing the judgment of human caseworkers with the cold, bloodless decision-making of machines.

At its most forbidding, Guardian reporters paint a picture of a 21st-century Dickensian dystopia that is taking shape with breakneck speed. The American political scientist Virginia Eubanks has a phrase for it: "The digital poorhouse."

Listen to governments, and you will hear big promises about how new technologies will transform poverty as a noble and benign enterprise. They will speed up benefits payments, increase efficiency and transparency, reduce waste, save money for taxpayers, eradicate human fallibility and prejudice, and ensure that limited resources reach those most in need. But so often, those pledges have fallen flat.

At a time when austerity dominates the political landscape, millions have had their benefits slashed or stopped by computer programs that operate in ways that few seem able to control or even comprehend. Mistakes have become endemic, with no obvious route for the victims of the errors to seek redress.

This week, the automation of poverty will be brought on to the world stage. Philip Alston, a human rights lawyer who acts as the UN's watchdog on extreme poverty, will present to the UN general assembly in New York a groundbreaking report that sounds the alarm about the human rights implications of the rush to digitalize social protection.

Alston's analysis is based partly on his official UN studies of poverty in the UK and US, and partly on submissions from governments, human rights organisations and experts from more than 34 countries. It is likely to provide the definitive snapshot of where the world lies now, and where it is going, addressing the harassment, targeting and punishment of those living in the rapidly expanding digital poorhouse.

Mistakes have become endemic, with no obvious route for the victims to seek redress

In Illinois, the Guardian has found that state and federal governments have joined forces to demand that welfare recipients repay "overpayments" stretching back in some cases 30 years. This system of "zombie debt", weaponized through technology, is invoking fear and hardship among society's most vulnerable.

As one recipient described it: "You owe what you have eaten."

In the UK, we investigate the secure government site outside Newcastle where millions are being spent developing a new generation of welfare robots to replace humans. Private companies including a New York outfit led by the world's first bot billionaire, are supercharging a process which has spawned a whole new jargon: "virtual workforce", "augmented decision-making", "robot process automation".

The government is rushing forward with its digital mission despite the pain already being inflicted on millions of low-income Britons by the country's "digital by default" agenda. Claimants spoke of the hunger, filth, fear and panic that they are enduring.

In Australia, where the Guardian has reported extensively on robodebt, the scheme that has been accused of wrongly clawing back historic debts through a flawed algorithm, we now disclose that the government has opened a new digital front: using automation to suspend millions of welfare payments. Recipients are finding their money cut off without notice.

The most disturbing story comes from Dumka in India. Here, we learn of the horrifying human impact that has befallen families as a result of Aadhaar, a 12-digit unique identification number that the Indian government has issued to all residents in the world's largest biometric experiment.

New high-tech approaches sweep through social services, work, disability and health, often with minimal public debate

Motka Manjhi paid the ultimate price when the computer glitched and his thumbprint – his key into Aadhaar – went unrecognised. His subsistence rations were stopped, he was forced to skip meals and he grew thin. On 22 May, he collapsed outside his home and died. His family is convinced it was starvation.

The Guardian investigations illuminate the shared features of these new systems, whether in developing or developed countries, east or west. The most glaring similarity is that all this is happening at lightning speed, with hi-tech approaches sweeping through social services, work and pensions, disability and health, often with minimal public debate or accountability.

Within that revolution, the human element of the welfare state is being diluted. Instead of talking to a caseworker who personally assesses your needs, you now are channeled online where predictive analytics will assign you a future risk score and an algorithm decide your fate.

In the new world, inequality and discrimination can be entrenched. What happens if you are one of the five million adults in the UK without regular access to the internet and with little or no computer literacy? What if the algorithm merely bakes in existing distortions of race and class, making the gulf between rich and poor, white and black, college-educated and manual worker, even more pronounced?

There is also a chilling Kafkaesque quality that spans the globe. As Manjhi so tragically discovered, mistakes are made. Machines glitch. If there is no one within reach who sees you as a person and not as a 12-digit number to be processed, the results can be fatal.

Computer says no: the people trapped in universal credit's 'black hole'

Vulnerable claimants already reporting problems, even before further DWP digital transformation

Robert Booth Social affairs correspondent

Mon 14 Oct 2019 10.00 BST Last modified on Tue 15 Oct 2019 15.00 BST

When the universal credit computer says no, fragile lives can quickly crumble.

Lucy Morris, a 32-year-old mother of one in Rochdale, was scraping by on her beauty therapist's wage topped up with UC when she failed to check a box on the benefit's online form and lost a £400 payment. It was enough to torpedo her finances and before long the heating was off, vegetables were dropped from meals and the house grew filthy because she could not afford cleaning products.

"It is designed to make it difficult so that they can get as many people off it as they can," she said.

Mark Abraham, a married father of twins in south London, was denied a month's benefits because an automated system linking salary data from HMRC with the Department for Work and Pensions misreported his previous income from a TV production job. He showed wage slips that proved the pay was reported wrong, but jobcentre staff could do nothing. Food ran low, he broke down mentally and ended up homeless and estranged from his family.

"The DWP staff appeared to be in thrall to the UC computer, allowing it to make all the decisions," he said. "Being able to tackle the computer beast that had made this decision wasn't within their capability."

Mary Blyde, a 61-year-old with incontinence and learning disability, was discovered by a charity worker lying in her unheated home in Gateshead on a urine-soaked sofa after her benefit was cut. She had missed a note on her online account warning her she needed to take action. When a charity worker found her, all that was left in her cupboard were three potatoes, a can of meat and a carton of orange juice.

"Sometimes I get scared the money won't come into my bank," she said.

These are some of the lives that campaigners warn risk being forgotten as the government and its technology industry partners pour millions into automation, artificial intelligence and machine learning in the benefits system.

Cruel and chaotic? No, in the DWP's fantasy land, universal credit is a huge success

The strategy might seem inevitable, welcome even, given its spread to other walks of life. How we use transport, watch TV and use email are all increasingly shaped by AI. Many UC claimants welcome the ability to communicate digitally rather than waiting on hold on the phone. Ministers argue digitisation will make claiming benefits more straightforward, reduce fraud and save money.

The DWP's last annual report claimed investment in digital technology "improved the experiences of people who rely on our services, making us more effective and efficient, and enabling us to personalise delivery for customers and claimants".

The ministry tries to create an atmosphere that will attract programmers to develop technologies for a welfare system that is used by 20 million people. There is an "innovation *dojo*" to investigate new technologies, and its "intelligent automation garage" – an echo of Google's Digital Garage training centres which began in sunny California – is looking for ways to leverage the “DWP data lake” for “improved citizen outcomes”, albeit based in a grey government building in north Newcastle.

But many see the reforms as erecting digital walls around the welfare state which are now only scalable by the computer literate. A further digital transformation with robotic automation and machine learning will only make things worse, they fear.

“It’s massive,” said Shona Alexander, chief executive of Citizens Advice in Newcastle which helps people navigate the system. “They think it’s going to save money but we don’t think it will. The waste from the mistakes is many times more [than the savings]. We are seeing more and more people because they have tried online and can’t do it.”

She recalled a released prisoner who dropped in recently for help signing on for UC, could not cope with the system and said he wanted to go back inside. “He kicked off, the police came and he was arrested,” she said.

Charities are trying to help. Blyde, for example, relies on Gary Fawcett at the Your Voice Counts in Gateshead. He has spent 155 hours on her case, remarking: “It almost broke our project.”

A recent exchange on her online journal was telling. She had already lost almost £1,000 in benefits because she did not know she needed to input more data, so he asked the DWP to tell her about future problems by post.

“Mary CANNOT read or access her journal in any way,” he wrote on her journal. “I can no longer be sorting this out for her.”

The reply came: “We do not communicate by letter … this is an online service.”

Fawcett said again she could not access, read or input into the journal. He asked: “Who is supposed to do this?”

The DWP replied simply that the online system needed updating, signing off: “Please do this ASAP.”

Fawcett is also trying to help Julian Jennings, 65, who cannot read or write, has learning disabilities and does not even know he is on UC.

“I have never used a computer in my life,” he said as he sat at Fawcett’s desk with his UC account screen open. “I used to go to the dole. It was much easier. You talked about it and signed it. If there were any problems they used to sort it.”

There is a message from a UC official: “Hello Julian. Please read the attached letter.”

“I can’t read none of that,” he said, peering at the screen. “How are you supposed to get your money?”

Blyde and Jennings are among 1.5 million people in the UK with learning disabilities and are not alone in struggling with the system. Tears filled the eyes of Danny Brice, 47, in London when he showed the Guardian how difficult he has found negotiating the UC programme with learning disabilities and dyslexia.

“I call it the black hole,” he said. “I feel shaky. I get stressed about it. This is the worst system in my lifetime. They assess you as a number not a person. Talking is the way forward, not a bloody computer. I feel like the computer is controlling me instead of a person. It’s terrifying.”

Nine million people in the UK are functionally illiterate and 5 million adults have either never used the internet or last used it more than three months ago.

And yet many of these people rely on a “digital by default” welfare system.

The DWP said that humans remain available to help: “We continue to invest in frontline colleagues, from phone lines to work coaches to front of house staff,” said a spokesperson. “This means people who struggle with digital services, or are worried about a wrong decision, can get the help they need. And because we know that, for whatever reason, some people don’t want to come into a jobcentre, we are funding Citizens Advice to help support people with their claims.”

Nathalie Nasor, a crisis case worker at Oasis Community Housing in Gateshead, has been helping Gary Warburton, 56, a former industrial cleaner who has been on UC since 2017. They have struggled with mysterious cuts to his benefits which seem to be the result of the system combining different databases.

He recently lost sums which turned out to be repayments for a crisis loan which dated back to 1997, an overpayment of tax credit five years ago, magistrates court fines and council tax arrears.

“It drives you absolutely cuckoo,” he said. “These are things from 1998. They control you and don’t give a hoot what happens to you.”

“His anxiety and mood went to an all-time low,” said Nasor.

The digital system has been very good for fraudsters. Staff have already made 42,000 referrals for fraud, the government has said. Kasim Mahmood, 29, a supermarket worker with Asperger’s in Manchester, was one victim.

In July he was approached on Snapchat by someone offering him extra cash. As asked, he provided his address, national insurance number and driving licence, and £1,525.44 landed in his account. The fraudster had applied for a UC advance payment – essentially a loan – and it had arrived fast. He threatened Kasim with violence if he did not send half. So now Mahmood was in debt and his existing benefits were stopped because the DWP computer thought he was receiving universal credit.

“I would struggle to get any benefit, yet there are people out there who get it at the click of a finger as long as they know their way around the digital system,” said his mother, Rucksana Mahmood. “There are some things we have to move forward digitally and others keep in the old fashioned way.”

Last year the UN rapporteur on extreme poverty, Philip Alston, warned that the postwar British welfare state was disappearing “behind a webpage and an algorithm” and that the impact on the human rights of the most vulnerable would be “immense”. The then chancellor Philip Hammond said his report was “nonsense”. Later this week Alston will deliver a separate report on the global rise of digital welfare to the UN general assembly in New York.

Benefits of ‘welfare robots’ and the need for human oversight

Simon McKinnon of the DWP, **Tom Symons** of Nesta and **Pat McCarthy** respond to articles on the use of artificial intelligence in managing benefit claims

Thu 17 Oct 2019 18.02 BST

Re your article (March of the ‘welfare robot’ triggers fears for poorest, 15 October), I found it disappointing and surprising that you do not see the benefits of the DWP’s work in artificial intelligence, despite a recent editorial (22 May)praising the benefits of AI in helping patient care in the NHS.

Our “intelligent automation garage” is using technology to improve the experience for claimants, bringing it in line with the service most of us expect and enjoy from our banking apps, shopping websites or utility providers. We are establishing a welfare system for the future.

Far from being “ruled by computer algorithms”, our bots don’t make decisions regarding people’s benefits. Instead they focus on everyday repetitive tasks so our colleagues can spend more time supporting vulnerable claimants face to face. Our use of machine learning is making the system simpler for people – so, for example, it won’t ask for more information than is necessary for straightforward claims such as childcare and housing. Importantly, digital-first is not the same as digital-only – as the Guardian, a digital-first platform since 2011, will understand.

While we proudly invest in our award-winning digital team, we continue to support vulnerable people in our jobcentres every day and ensure there’s still face-to-face support for those who need it.

Simon McKinnon
Chief digital and information officer, Department for Work and Pensions