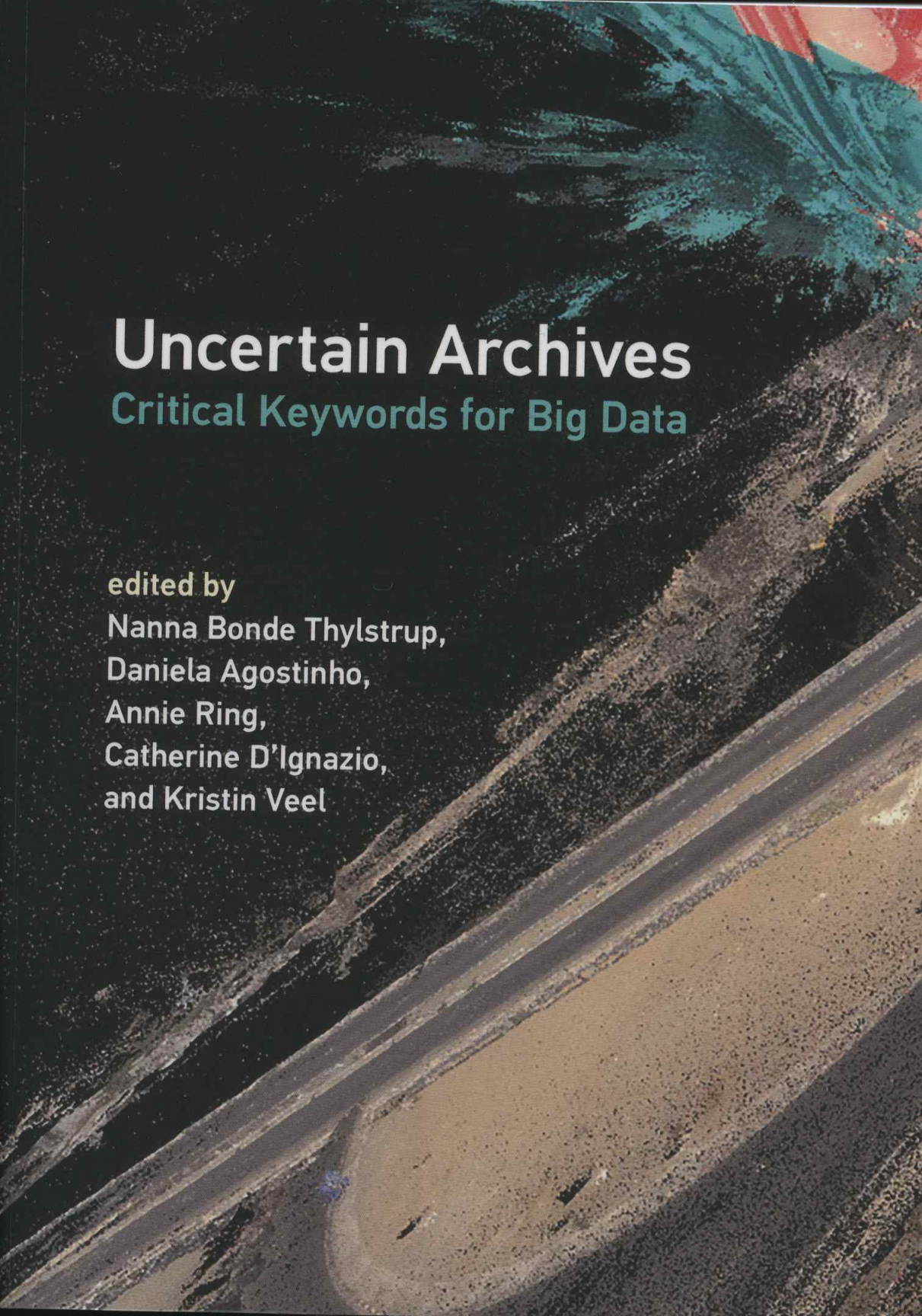
Uncertain Archives
Critical Keywords for Big Data
edited by
Nanna Bonde Thylstrup,
Daniela Agostinho,
Annie Ring,
Catherine D'Ignazio,
and Kristin Veel

43 Prediction

Manu Luksch

We can predict it for you wholesale.

Before You Even Know It

> What do others buy after viewing this item? Well, if you like the sound of that, then you'll love what's coming up next.
>
> —Luksch, Reinhart, and Tode 2015

US Patent No. 8,615,473 B2 of December 24, 2013, "Method and System for Anticipatory Package Shipping," is assigned to Amazon Technologies Inc., a subsidiary of one of the world's most data- and dollar-wealthy corporations. Using *forecast analysis of business variables*—the predicted desire for an item based on a customer's prior online behavior (such as previous orders, searches, wish lists, cart contents, and cursor "loiter time" over items)—the corporation proposes to "speculatively ship . . . without completely specifying a delivery address" to cut down on fulfillment times: the shipped item is held locally and offered to customers with almost immediate delivery.

Prediction Is Difficult

> Prediction is very difficult, especially if it's about the future.[1]

Predictions are claims about specific, well-defined events in the future whose occurrence can be uncontroversially verified. A prediction ought to be sharp (precise) and accurate (have a significantly greater than even chance of being correct).

Of the diverse strategies and approaches to prediction deployed historically, those involving the systematic observation of nature have proved more successful than those based on numerology, the examination of animal entrails, or the interpretation of dreams. Ancient astronomers predicted eclipses based on the periodicity observed in past records; foreknowledge of cosmic events endowed them with immense cultural authority. During the Scientific Revolution, natural philosophers formalized observational and experimental procedures and generated theories with formidable predictive power. Mathematics continues to prove "unreasonably effective" (Wigner 1960) in the natural sciences. (For example, experimentally determined values of a quantity called the electron g-factor agree with theoretical computations to within one part in a trillion.) Despite dramatic epistemological complications arising from twentieth-century developments (Heisenberg's uncertainty principle, Gödel's incompleteness theorems, chaos theory), mathematics and natural sciences set the standards for both deterministic and probabilistic reasoning, most conspicuously as they quantify the uncertainty of their predictions.

What has constrained predictive power in the humanities and social sciences—even in the most mathematized of all, economics, where *Homo sapiens* has been rationalized beyond recognition into a monomaniacal *Homo economicus* (see figure 43.1)? Historically, the obstacles have been located in the very nature of the objects of study. Human and social phenomena arise from the actions of self-interpreting subjects who are susceptible to influence, inclined to rebelliousness and akrasia,[2] resistant to reductionist analysis, and (unlike electrons) heterogeneous. However, the application of machine learning to large data sets appears to have brought about dramatic improvements in the prediction of human behavior. Does this augur the end of individual independence, inspiration, improvisation—and also of mystery, suspense, romance (see figure 43.2)?

Past Performance Is No Guarantee of Future Results

The computed likelihood of an event may be underwritten by different conceptions of probability, as determined by, for example:

(i) Relative frequency in the statistical record of the past. (Although, as financial service providers typically hedge, "past performance is no guarantee of future results.")

(ii) Symmetry. The chance of an event is constrained by logical or physical symmetries of the system, such as those of an unbiased coin. (But a complete model of an unbiased toss would have to include at least the probability of the coin landing on its edge, among other, possibly more exotic outcomes.)

Figure 43.1
Advertisement by investment management company Winton Capital: a depiction of Cinderella annotated with parameters of time, pupil dilation, adrenaline, heartbeat, charm quotient, shoe size, composition of shoe, foot size, subject profile, available transportation, distance to destinations, and current speed. London, 2015. Photo by Manu Luksch.

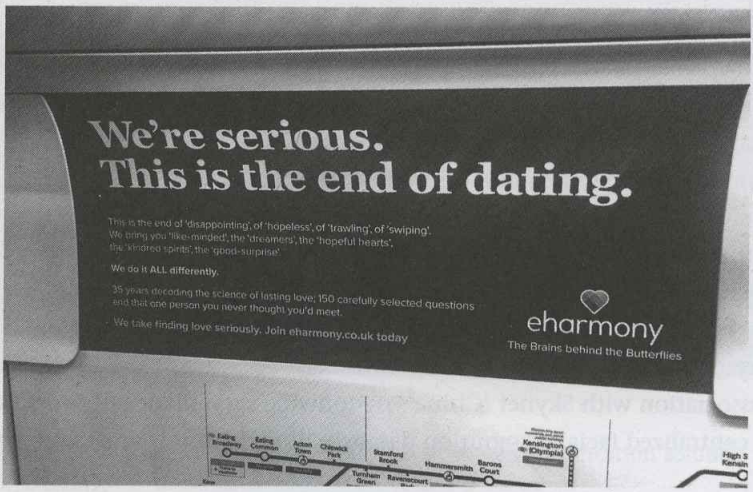

Figure 43.2
Advertisement by online dating platform eharmony: "35 years decoding the science of lasting love. The Brains behind the Butterflies." London, 2018. Photo by Mukul Patel.

(iii) A causal mechanism or model. (Consider an event as an outcome of a random natural process, such as radioactive decay.)

(iv) The degree of belief of an individual that the event will occur (subjective probability).

Prediction in the human sciences is heavily driven by (i) and deeply vulnerable to (iv).

Real-World Counterfactuals: What Could Have Been

To reason counterfactually is to consider what might have been, to speculate about alternative outcomes. But this requires causal understanding, and in messy, complex human domains such as economics, causation is neither simple nor uncontested.

> If, five years after leaving the EU [European Union], the UK economy booms, would this refute the prediction of the Remain Camp that Brexit would ravage Britain's economy? If, instead, after five years the UK is in recession, would this confirm the prediction? In both cases, the answer is no; neither observed outcome proves anything. In order to "prove" our predictions, we would need to establish the counterfactual. But there are too many factors that would need to be accounted, and the outcome quantified under the counterfactual. And that is something we simply are unable to do. (Simon Bishop, RBB Economics, pers. comm.)

The impossibility of the counterfactual arises in another way too. The decision to grant a defendant bail can be assessed *ex post* based on compliance—but not the decision to deny bail. You cannot run a controlled experiment on a society[3]—or on an ecosystem.[4]

The Long and Short of IT

The proliferation of networked sensors and ubiquitous processing makes possible a fully embedded urban information infrastructure—an operating system for smart cities. Trends gleaned through data correlation can be extrapolated into future patterns. Anticipatory algorithmic management will, it is claimed, lead to a resilient, ecologically balanced future free from uncertainty—as you remember it from techno-utopian fictions (Dick 1966).

Just as Joe Chip discovers in a face-off with his smart door in *Ubik* (Dick 1969),[5] even everyday infrastructure is not above suspicion. Closed-circuit television manufacturer Hikvision (40 percent owned by the Chinese Communist Party) has supplied over a million cameras compatible with facial recognition technology to the UK—including to the parliamentary estate and the police. These contracts were awarded with full knowledge of Hikvision's association with Skynet (China's nationwide surveillance network) and its development of a centralized facial recognition database, for which it is amassing data (Gallagher 2019).

In its full expression, the corporate-governed smart city epitomizes a new enclosure movement. Using network surveillance, it builds private databases that fuel prediction products. These products are traded in a behavioral futures market that Shoshana Zuboff (2019) has identified as the mainspring of surveillance capitalism.

Self-Fulfilling Prophecies I: Classification and Profiling

"Behind every customer is an individual," claims consumer credit reporting agency Experian (2016), advertising its Mosaic tool, which enables corporations to "personalise customer experience and increase share of wallet." For useful patterns to emerge from this data, categories must be created. Mosaic classifies consumers into sixty-six types across fifteen groups, using criteria such as home ownership and location, career stage, and ethnicity (see figure 43.3).

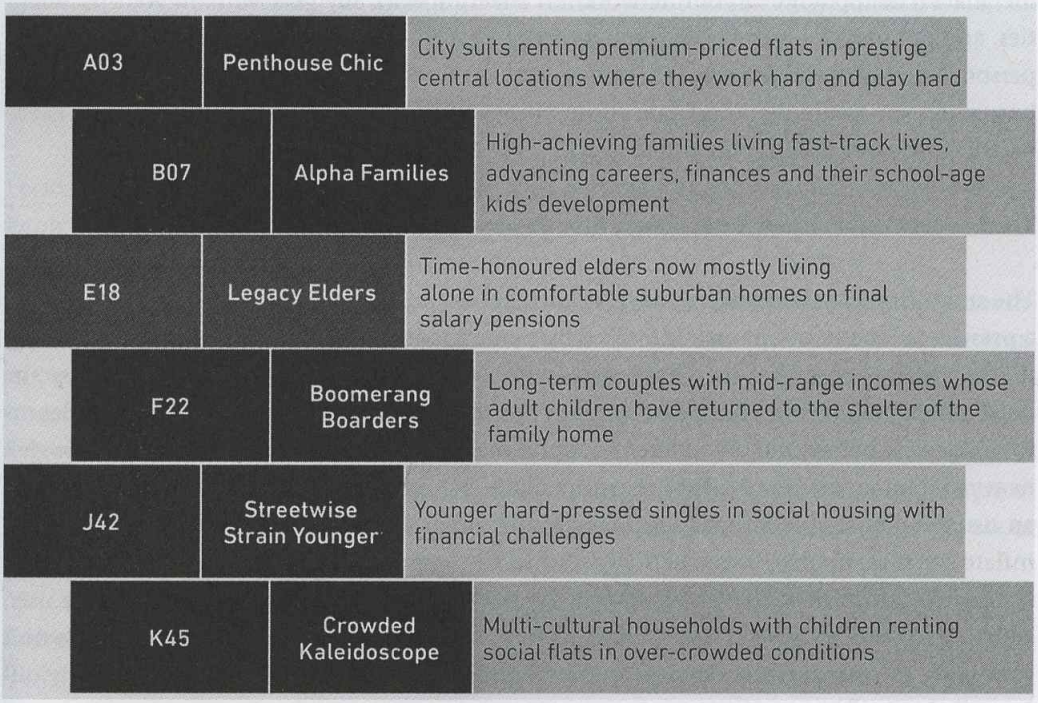

Figure 43.3
Classification examples in Experian Mosaic (2016) "the consumer classification solution for consistent cross-channel marketing."

The classification system is riddled with arbitrariness and value judgments—"affordable but pleasant," "bright young singles," "time-honored elders"—that propagate through decision algorithms that run on the database. Experian markets the tool across the retail and service sectors, including human resources and financial agencies that can have life-changing impact. "Through denying credit or screening career opportunities, negative profiles can haunt an individual across various domains. A person's data shadow does more than follow them; it precedes them" (Kitchin 2015).

Better Than You: False Consciousness, True Intentions

Network intelligence corporations declare, with no sense of hubris, that their extensively correlated data sets "know" you better than you know yourself.[6] They have the capacity to infer your creditworthiness, voting intentions, relationship status, and sexual preferences through a combination of first-order data (such as what your neighbors are like) and second-order metadata (the topology of your network).[7] It is very specifically you, with all your peculiarities, and not just as a member of some category, that these corporations want to know. Deep personal knowledge yields valuable information on how you might behave (what you might buy, when you might die). And data anonymization provides little protection against a motivated, mathematically sophisticated agent.[8]

The Easiest Way to Predict the Future Is to Produce It

The smart city invites citizens to devolve their precious processing to it, allowing it to curate a preselected menu of options. Informed by neuroscientific research on perception and cognition, fed by behavioral data harvested from phones and smart cards, and driven by the need to compete in the attention economy, the agile vendor of products and services learns to nudge user behavior ("if you like the sound of that . . ."). Is the actual reason for Google's heavy secrecy around its YouTube recommendation engine that the algorithms are driven by an unutterably simple-minded agenda—maximizing ad click-through? Reap viewer desire, inflate the filter bubble, but remain agnostic as to content.

Some illusions need to be maintained. Prediction is much less hazardous when the user, confronted with an explicit and limited set of options, feels that she is choosing freely. And seemingly inconsequential changes in the framing of options may lead to dramatically different behavior.[9] Swipe which way?

The act of prediction aspires to be an epistemic act. But a culture governed by predictive algorithms is less a factory of knowledge and more a factory of conformity. A culture

that understands itself to be predictable homogenizes. All prophecies become self-fulfilling (provocations).

A Prediction

If you suspect that anything you think might be held against you—you will self-censor.

The Fast Mirror and the Slow

To protect the Khazar princess Ateh from her enemies, blind men paint letters of a proscribed alphabet on her eyelids each night. Anyone who reads the letters dies. One morning, the princess receives a pair of extraordinary mirrors as a gift, one fast and the other slow. "Whatever the fast mirror picked up, reflecting the world like an advance on the future, the slow mirror returned, settling the debt of the former" (Pavic 1989, 23). Having not yet cleaned the letters from her eyelids, she looks into the mirrors and dies instantly, killed twice in an interblink.

What Wouldn't You Want to Foresee?

Personal: If you could know exactly when you will die—would you want to? If you did know—who else ought to know? Should some predictions not be made, even if they could be? And exactly how would you police that?

Pedestrian: "Every technological intervention that is made with the intention of smoothing out urban experience deprives us of an opportunity to encounter something external to our own will, and so doing robs us of a moment in which we might reflect on the contingency of our own values, choices and beliefs" (Greenfield 2013, location 732–734). Foreknowledge makes life—boring?

Antiexistential: Cassandra foresaw deglaciation, extreme climatic events, and ecosystem collapse but was cursed by sulking Apollo, never to be believed. This tragedy brings about no catharsis.

Correlation Is Enough: The End of Theory?

The spectacular growth of data acquisition and computational capacity, together with the development of algorithms for data correlation and pattern recognition, has given rise to a new class of predictive models of unprecedented accuracy—it is claimed. For corporations

who trade in network intelligence and surveillance, correlation suffices—with the healthy returns produced by pattern-matching algorithms, there is no need to develop causal models. If economic gain can be secured on the basis of privileged knowledge of what will be the case, there is no need to understand why it will be so. Moreover, where the prediction engine is a neural network, it is practically impossible to trace the reasoning behind a particular computation. Only the ends matter—not the means.

Evangelists of machine learning who place short-term efficacy over deep causal understanding contend that science, too, should be conducted under such a program (see, e.g., Anderson 2008). But in the natural sciences, theories are valued not merely for their predictive accuracy; other important criteria for choosing among competing theories that fit the same observations include consistency, simplicity, elegance, generality, and explanatory power.

Correlation is not causation.

What Is It Like to Be a Cat?

Predicting cats: the physics of the falling feline—just how it rights itself midair to land on its feet—exercised nineteenth-century scientists, including James Clerk Maxwell. A rigorous mathematical solution only emerged in 1969 (Kane and Scher 1969), but already in 1894 the French physiologist Étienne-Jules Marey had published empirical studies with photographs that made crucial details of the cat's double-rotational movement visible (see figure 43.4). Marey shot the falling cat with his chronophotographic gun, a twelve-frames-per-second camera that he had developed specifically to study motion. Through geometric analysis of image sequences, Marey not only gained insight into the underlying biomechanics but also saw the possibility of optimization and control. The French military was his first client.

To prevail in the scramble for unconquered territory, an army must tune its soldiers' movements. Like the falling cat, it cannot rely on luck (Luksch, Reinhart, and Tode 2015).

Contemporary predictive analytics is a direct descendant of Marey's work, via the early twentieth-century time-and-motion studies of Frank and Lillian Gilbreth (see figure 43.5).

Cats predicting: biomechanical explanations do not exhaust totemic associations with luck and death: the feline mind remains mysterious to us. Oscar, the resident cat at the Steere House Nursing and Rehabilitation Center, Providence, Rhode Island, has an "uncanny ability to predict when residents are about to die. Thus far, he has presided over the deaths of more than 25 residents on the third floor. . . . His mere presence at the bedside is viewed by . . . staff as an almost absolute indicator of impending death" (Dosa 2007, 328–329). If an artificial agent made such predictions well, would you heed it?

Correlation is not causation, and behavior is not experience.

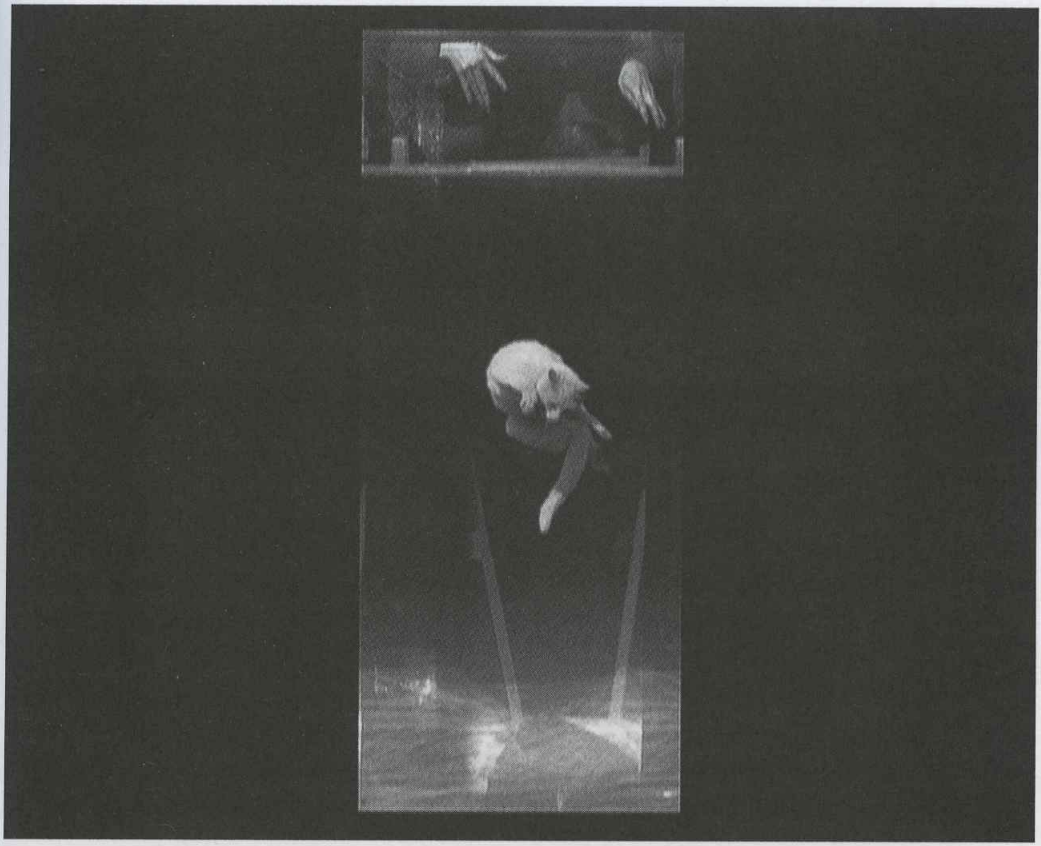

Figure 43.4
Still from *Chat. Chute avec retournement*, by Étienne-Jules Marey, 1890–1904. Cinemathèque Français, FM-CT5. In: *Dreams Rewired* by Luksch, Reinhart, and Tode (2015).

Self-Fulfilling Prophecies II: Cybernetic Failures

"Virtually all predictive policing models . . . use crime data from police departments. But police department data is not representative of all crimes committed; police aren't notified about all crimes that happen, and they don't document all crimes they respond to. . . . If a police department has a history of over-policing some communities . . . predictive policing will merely reproduce these patterns in subsequent predictions" (Isaac and Lum 2018). Similar fundamental errors in conceptualization and specification can be found across other sectors where predictive analytics is hailed as a panacea, including job recruitment and

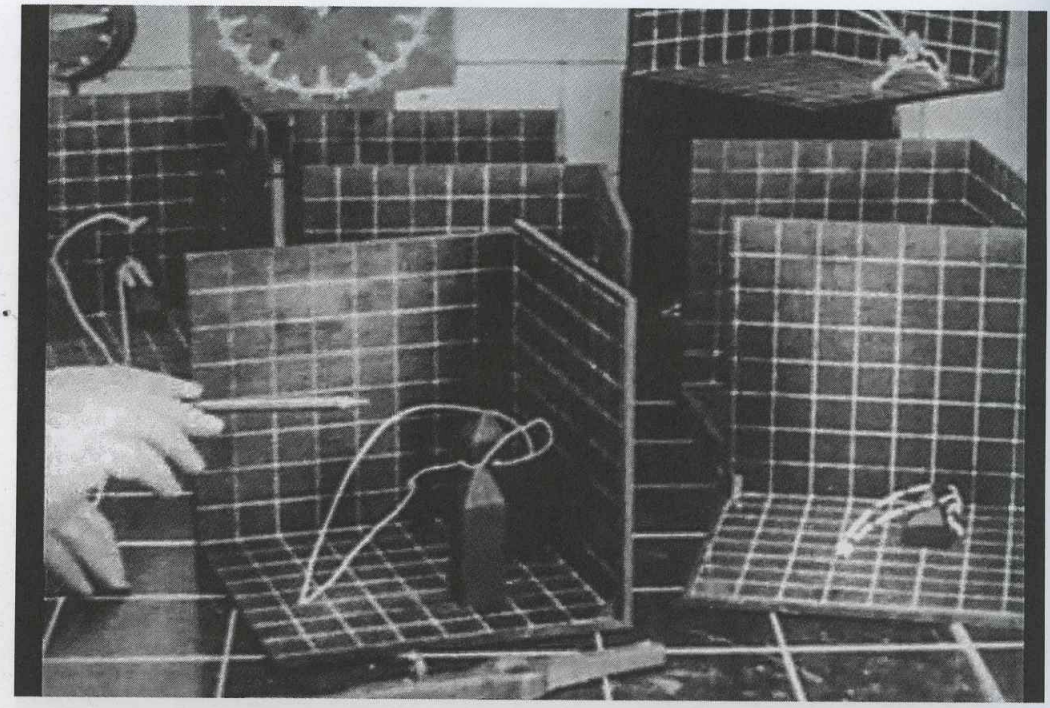

Figure 43.5
Still from *Original Films of Frank B. Gilbreth, 1910–1924* by Ralph Barnes, Lillian Gilbreth, and James Perkins, 1945. Purdue University Libraries, Karnes Archives and Special Collections. In *Dreams Rewired* by Luksch, Reinhart, and Tode (2015).

university admission.[10] In control theory, positive feedback is a source of instability. The amplification of historical bias is an elementary failure of cybernetics.

Enclosure of the Future (Foreclosure)

In its full expression, the corporate-governed smart city epitomizes the contemporary enclosure movement. The enclosure movement of thirteenth-century England took common land into private ownership. The neoliberal transfer to private enterprise of common services, utilities, community knowledge, and space, beginning in the 1980s, constituted a new enclosure of the entire public sphere. Today's enclosure movement, driven by prediction engines (see figure 43.6.), is the most egregious of all—it brings about the enclosure of the future, which is also a part of the commons.

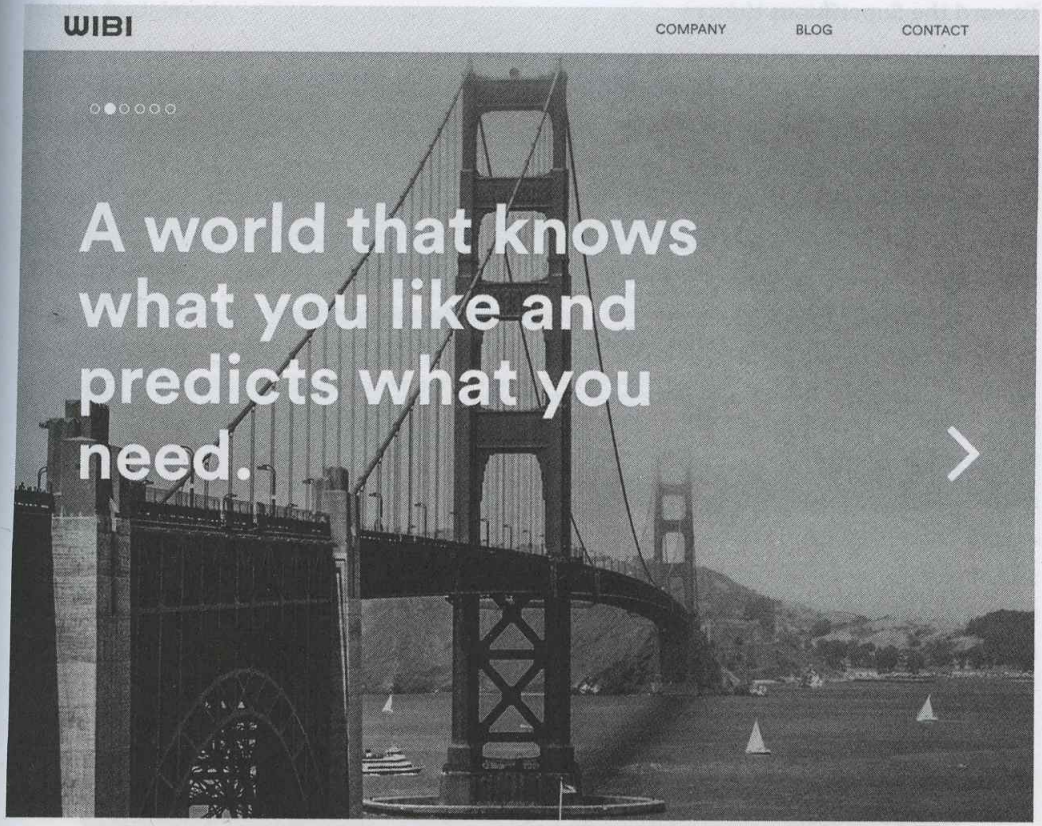

Figure 43.6
Homepage of WibiData, a software company for personalized customer experience. http://www.wibidata.com. Last accessed February 21, 2015.

> There's a more personal world out there.
> A world that knows what you like and predicts what you need.
> Where your experience feels less like it was produced by a machine,
> and more like it's coming from a friend.
> This is the world we're creating. . . . (WibiData, public statement)[11]

Prediction engines warrant close public scrutiny for their impact on human rights—in particular, the rights of privacy, autonomy, and self-determination; the right to information; the rights to due process and freedom from the arbitrary effects of social sorting; and the "right to a future tense" (Zuboff 2019, 20). The opacity of such machine procedures threatens to make our own decisions opaque to us—was that truly an informed choice that we made? And when the very idea of human agency is in flux, the cogency of rights discourse is diminished.

Toward the Superfluous Human

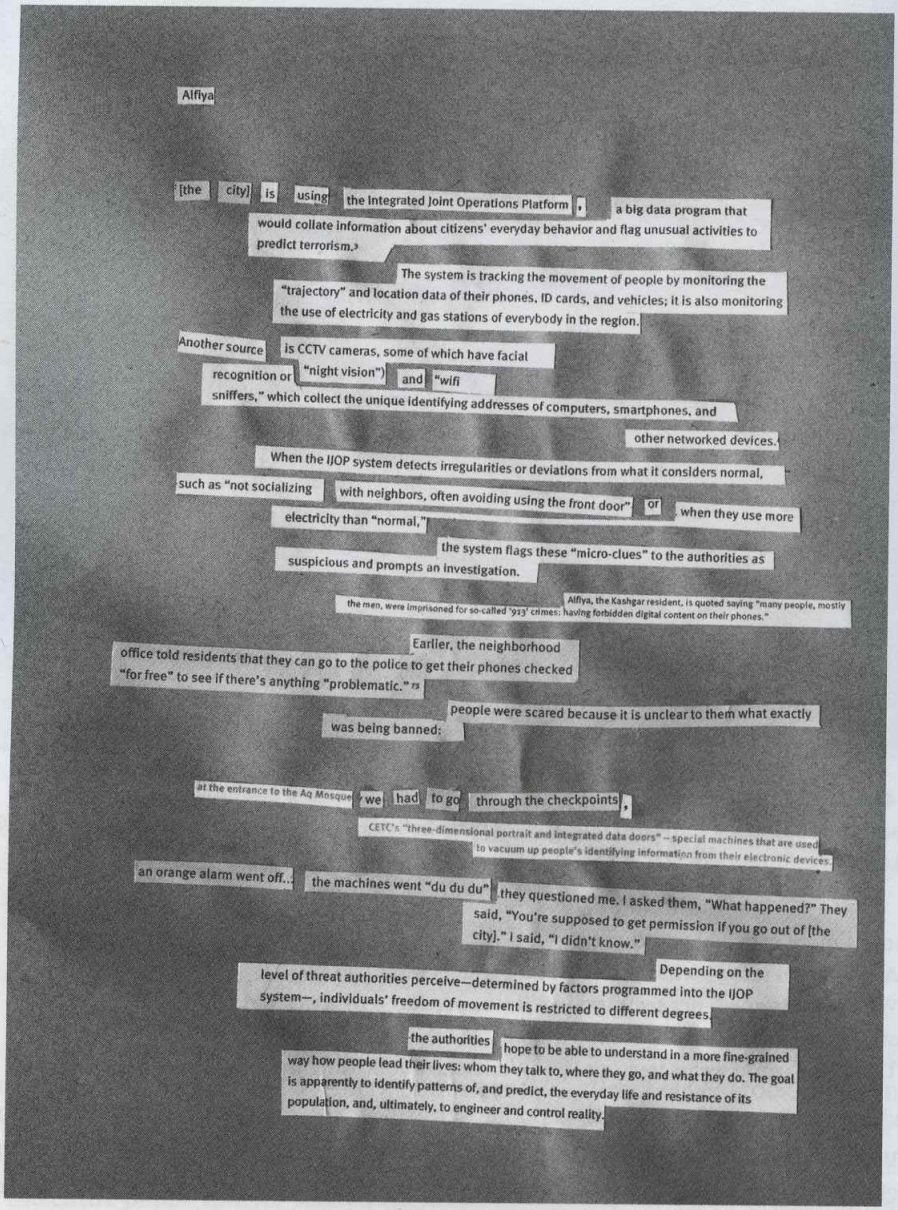

Figure 43.7

Alfiya—A Found Dystopia by Manu Luksch, 2019. Paper collage of text excerpts from *China's Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App* by Human Rights Watch (2019).

Futures Reclaimed

Even as we are objects of prediction and optimization technologies and their algorithmic amplification of arbitrariness, bias, and error, we are further subject to autocratic decisions from centralized power. Competitive advantage and revolutionary surprise are rooted in asymmetries of knowledge; volatility is celebrated equally by Mark Zuckerberg ("move fast and break things") and Donald Trump ("We must as a nation be more unpredictable" [Trump 2016]). There is a widespread belief within the Silicon Valley tech elites that they are "the people who not only can [shape the future], but . . . the only people who can" (Robert Scott, founder of the festival Further Future, quoted in Bowles 2016).

To reclaim the future—perform (*parfournir*: furnish) the unexpected? One suggested mechanism of resistance is the deployment of protective optimization technologies (Overdorf et al. 2018): techniques to collectively nudge or pollute the data and inferential processes of optimization systems; *détournement* for the algorithmic age. Despite various vulnerabilities (unintended consequences, hijacking by special interest groups, the ethical chasm of deception . . .), such tactics of collective action might well prove effective as a first salvo in the battle against centralized predictive power. However, sustainable systemic change requires robust models for accountability and transparency, processes for deliberative justice, and a conception for human and planetary flourishing. Reclamation does not suffice; an envisioning is necessary. Prediction is the uncanny sister of hope.

Notes

With thanks to Mukul Patel for invaluable discussion and support in producing the manuscript.

1. Variously attributed to Niels Bohr, Robert Storm Peterson, and others.

2. Understood as acting against one's better judgment due to weakness of will.

3. See, for example, the discussion on algorithms predicting for recidivism in Fry (2018, 49–78) and also the paper discussed therein (Kleinberg et al. 2017).

4. Temporal factors predominate: When will unintended consequences become apparent? And how will we recognize them as consequences?

5. Chip is held up by his own front door—a door that is aware of its contractual rights.

6. In 2002 statistician Andrew Pole worked for the US chain of Target stores to discover newly pregnant customers. Shopping habits change due to major life events, opening a window for brand switching. The chain was diversifying its stock and wanted to secure new customers through targeted promotions. "As soon as we get them buying diapers from us, they're going to start buying everything else too," Pole claimed to Charles Duhigg (2012). Pole sniffed out pregnant shoppers by looking for changes in

their regular basket—for example, to unfragranced or additive-free toiletries. These consumers received custom mailers with vouchers for new lines of baby goods. An angry father of a teenage girl accused Target of encouraging her interest in babies, only to find that she was already pregnant (he apologized to the store manager, who had called to apologize on behalf of Target).

7. See, for example, Zhong et al. (2015) and Facebook's 2009 patent application US20100257023A1, "Leveraging Information in a Social Network for Inferential Targeting of Advertisements."

8. Given a small amount of auxiliary information, an individual's identity can be inferred purely from the topological features (connections) of their networks (see Narayanan and Shmatikov, n.d.).

9. The classic exploration is Kahneman and Tversky (1981).

10. Cathy O'Neil discusses the discriminatory admissions procedures, perpetuated by algorithms, of a medical school in her *Weapons of Math Destruction* (2016, 115–118).

11. Data analytics company WibiData had major investment from Google's Eric Schmidt.

References

Anderson, Chris. 2008. "The end of theory: The data deluge makes the scientific method." *Wired*, June 23, 2008. https://www.wired.com/2008/06/pb-theory.

Bowles, Nellie. 2016. "'Burning Man for the 1%': The desert party for the tech elite, with Eric Schmidt in a top hat." *Guardian*, May 2, 2016. https://www.theguardian.com/business/2016/may/02/further-future-festival-burning-man-tech-elite-eric-schmidt.

Dick, Philip K. 1966. "We can remember it for you wholesale." *Magazine of Fantasy and Science Fiction*, April 30, 1966, 4–23.

Dick, Philip K. 1969. *Ubik*. New York: Doubleday.

Dosa, David M. 2007. "A day in the life of Oscar the cat." *New England Journal of Medicine* 357: 328–329.

Duhigg, Charles. 2012. "How companies learn your secrets." *New York Times*, February 16, 2012. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

Experian. 2016. "Mosaic: The consumer classification solution for consistent cross-channel marketing." Accessed July 15, 2019. https://www.experian.co.uk/assets/marketing-services/brochures/mosaic_uk_brochure.pdf.

Fry, Hannah. 2018. *Hello World*. London: Doubleday.

Gallagher, Ryan. 2019. "Cameras linked to Chinese government stir alarm in U.K. parliament." *Intercept*, April 9, 2019. https://theintercept.com/2019/04/09/hikvision-cameras-uk-parliament.

Greenfield, Adam. 2013. *Against the Smart City*. New York: Do Projects.

Human Rights Watch. 2019. *China's Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App*. N.p.: Human Rights Watch.

Isaac, William, and Kristian Lum. 2018. "Setting the record straight on predictive policing and race." *Medium*, January 3, 2018. https://medium.com/in-justice-today/setting-the-record-straight-on-predictive-policing-and-race-fe588b457ca2.

Kahneman, Daniel, and Amos Tversky. 1981. "The framing of decisions and the psychology of choice." *Science* 211: 453–458.

Kane, T. R., and M. P. Scher. 1969. "A dynamical explanation of the falling cat phenomenon." *International Journal of Solids and Structures* 5 (7): 663–670. http://doi.org/10.1016/0020-7683(69)90086-9.

Kitchin, Rob. 2015. "Continuous geosurveillance in the 'smart city.'" *Dis*. Accessed July 15, 2019. http://dismagazine.com/dystopia/73066/rob-kitchin-spatial-big-data-and-geosurveillance.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human decisions and machine predictions." Working Paper 23180. Cambridge, MA: NBER.

Luksch, Manu, Martin Reinhart, and Thomas Tode, directors. 2015. *Dreams Rewired*. Film. Austria/United Kingdom.

Narayanan, Arvind, and Vitaly Shmatikov. n.d. "De-anonymizing social networks." Accessed August 1, 2019. https://www.cs.utexas.edu/~shmat/shmat_oak09.pdf.

O'Neil, Cathy. 2016. *Weapons of Math Destruction*. London: Penguin Books.

Overdorf, Rebekah, Bogdan Kulynych, Ero Balsa, Camela Troncoso, and Seda Gürses. 2018. "POTs: Protective optimization technologies." https://arxiv.org/abs/1806.02711v3.

Pavic, Milorad. 1989. *Dictionary of the Khazars*. Female edition. New York: Vintage.

Trump, Donald J. 2016. "Foreign policy remarks of D. J. Trump." *Federal News Service*, April 27, 2016.

Wigner, Eugene. 1960. "The unreasonable effectiveness of mathematics in the natural sciences." *Communications in Pure and Applied Mathematics* 13 (1): 1–14. http://doi.org/10.1056/NEJMp078108.

Zhong, Yuan, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, and Xing Xie. 2015. "You are where you go: Inferring demographic attributes from location check-ins." In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 295–304. New York: Association for Computing Machinery.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism*. London: Profile Books.

This book was set in ITC Stone Serif Std and ITC Stone Sans Std by New Best-set Typesetters Ltd. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Thylstrup, Nanna Bonde, editor.
Title: Uncertain archives : critical keywords for big data / edited by Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio, and Kristin Veel.
Description: Cambridge, Massachusetts : The MIT Press, [2020] | Includes bibliographical references and index.
Identifiers: LCCN 2020004704 | ISBN 9780262539883 (paperback)
Subjects: LCSH: Big data—Social aspects. | Archival resources—Management. | Uncertainty (Information theory)
Classification: LCC CD931 .U53 2020 | DDC 027—dc23
LC record available at https://lccn.loc.gov/2020004704

10 9 8 7 6 5 4 3 2 1