8

Machine learning

The algorithmic production of knowledge

We've already discussed the massive volumes of data that stand to be collected by a true internet of things, the torrential trains of ones and zeroes coursing through and between all the world's networked devices. We've seen how digital fabrication devices can turn data into material artifacts of the most extraordinary delicacy and precision, how the desire to make shared data incorruptible furnishes the blockchain with its very purpose and reason for being. But we haven't yet paused to reflect on just what data *is*.

Let's start with the thought that whatever set of events life presents us with, we need to situate ourselves in the world, evaluate our circumstances and the possibilities they afford us for purposive action, and then decide among the options we're presented with—and this is true whether we're deciding who to befriend on the first day of kindergarten, choosing the right seeds to plant on the shadier side of our community-garden plot, or wrestling a quarter-billion-dollar fighter jet through a

high-G dogfight with a similarly equipped enemy. A simple way of defining data, then, might be *facts about the world, and the people, places, things and phenomena that together comprise it*, that we collect in order that they may be acted upon.

But before we can act upon any such collection of facts, we have to make sense of it. A commonplace of information science holds that data, information, knowledge and wisdom form a coherent continuum, and that we apply different procedures at every stage of that continuum to transform the facts we observe into insight and awareness. There are many versions of this model, but they all fundamentally assert that we *measure* the world to produce data, *organize* that data to produce meaningful, actionable information, *synthesize* that information with our prior experience of the world to produce knowledge, and then—in some unspecified and probably indescribable way—arrive at a state in which we are able to apply the things we know with the ineffable quality of balanced discernment we think of as wisdom.

The various versions of this model all generally assume that data itself is neutral and objective. Whenever we say "data," however, what we're really referring to is that subset of the world's infinite aspects that have been captured by some instrument or process of measurement. (In fact, the French word for "sensor," *capteur*, directly reflects this insight, and some of the more thoughtful observers of information-processing technology have argued that in English the word "capta" would be a more accurate way of describing that which is retained.)[1] For our purposes, what is vital to remember is that there's no such thing as "raw data." Whatever data we measure and retain with our sensors, as with our bodily senses, is invariably a selection from the far broader array available to us; perception itself is always already a process of editing and curation.

As we saw in our discussion of the blockchain, the conventional way of deriving actionable information from large bodies of data was to apply structure to it, by storing it in the linked cells of a relational database. For example, such a database might record each of Amazon's user accounts sequentially,

with cells containing the account holder's first name, last name, delivery address, the various credit cards they have on file, and so on. Retrieving information in this case is a simple matter of submitting a structured query to the database, and even this straightforward process is generally buried beneath the still-simpler front-end interface of a consumer-facing website: for example, we make such a query every time we click on a link that reads "View Your Orders" or "Manage Payment Options."

But managing flows of so-called big data—a buzzword that simply denotes the extremely high volume, velocity and variety of contemporary data production—stresses such conventional techniques to the breaking point. Storage capacity sufficient to cope with the onslaught simply may not be available. Populating a database accurately requires a significant investment of resource and effort, which may not be forthcoming. And in any event, most of the world's data—and virtually all of it that's germane to systems that operate in physical space and real time—does not happen to reside in the neat tables or crisply cellular structure of any database, and never will. So the new way of handling such situations is to look for emergent patterns in previously *unstructured* data, like a large body of text, a series of images, or indeed a real-time video feed. In fact, this is what "big data" is all about. There's something uncanny, almost Deleuzian about this process of interrogation: as they are iteratively resolved in ever-higher fidelity, the patterns themselves begin to suggest the questions that might be asked of them.[2]

The way such streams and flows are induced to render up whatever patterns are latent within them is by passing them through an algorithm—or more likely, several of them.

With its faintly exotic etymology and unfortunate near-homophony with a completely unrelated concept in mathematics, "algorithm" is one of those words that does so much to shroud discussions of information technology in an unnecessary aura of complexity. There needn't be any mystery on this count, though: all the word means is a finite, structured, sequential and

highly explicit set of instructions, a procedure for doing *this* to *that*. A well-specified recipe is an algorithm, as is the process of alphabetizing a list of names.

As you may by now suspect, algorithms are everywhere beneath the surface of contemporary life.[3] They govern what songs or films a streaming service will recommend, the price at which a given commodity will be offered to market, where a restaurant will seat its customers, which potential partners will appear in a dating app, and (if you're unwise enough to use a browser without an ad blocker installed) what ads are served to you. In contemporary society, a very great deal of material power reposes in the party that authors an algorithm. They determine your credit-worthiness, your insurability and the priority with which you will receive medical care in a mass-casualty emergency. (This last example generally involves assessing a patient's condition against a set of procedures specified on a laminated card, and it's an excellent reminder that not all algorithms are necessarily executed by software.)

Algorithms also manage processes that are still, for the most part, at the edge of everyday experience, but drawing ever closer: they instruct a bipedal robot how to pick up a package without losing its balance, a drone how fast each of its rotors must spin to maintain level pitch, or an autonomous car how to recognize obstacles in the roadway.

Because of the colossal volume of data that passes through them, changes to any of the more widely relied-upon algorithms can have consequences that ripple through the entire society. Every time Google tweaks its search algorithm, or Facebook the one it uses to govern story placement, certain business propositions suddenly become viable, and others immediately cease to be; more profoundly yet, certain perspectives on reality are reinforced, and others undermined. These particular tweaks, we should be clear, are made manually, invariably in response to some perceived vulnerability or weakness—in Google's case, that content farms and other low-quality sites were rising too high in their search results by gaming its algorithm, in Facebook's the accusation that their trending news feature was

biased against right-wing sources. But not all such algorithmic refinement is manual.

What links all of these situations is their dynamism. Because the circumstances of the world evolve so rapidly, an algorithm that is expected to face up to the challenges of everyday life can only suffer from being static and set in stone. Compelled to make its way in a fundamentally unpredictable and even turbulent operating environment, like any of us an algorithm will ideally be equipped with the ability to learn from its experiences, generalize from what it's encountered, and develop adaptive strategies in response. Over time, it will learn to recognize what distinguishes a good performance from an unacceptable one, and how to improve the odds of success next time out. It will refine its ability to detect what is salient in any given situation, and act on that insight. This process is called "machine learning."

What distinguishes this from "deep" learning, as some would have us call the process through which a machine develops insight? And why does it seem to have become so prominent in recent years?

In the beginning was the program. Classically, using computers to solve problems in the real world meant writing programs, and that meant expressing those problems in terms that could be parsed and executed by machine. Research into artificial intelligence proceeded along these lines for decades, culminating in the so-called expert systems of the 1980s, which attempted to abstract the accumulated expertise of a human diagnostician or trial lawyer into a decision tree built on a series of explicit if-then rules.[4] These systems did work, for some crude approximation of "work," but they were clumsy and brittle, failing completely when encountering situations their programmers hadn't envisioned.

And there was a still-deeper problem with this high-level approach to artificial intelligence. Many of the things we'd like algorithmic systems to do for us—whether recognizing handwriting or natural speech, identifying people and other discrete objects in the visual field, or succeeding at the continuous

exercise of identification we think of as the sense of vision itself—confound explicit articulation, and therefore expression in the form of executable code. These are things our brains do trivially and without conscious thought. But precisely for this reason, because we cannot explicitly reconstruct how we arrive at the decisions involved, we're generally unable to encode them as instructions computational systems would be able to make use of.

In other words, we might be able to imagine the set of principles that allow us to isolate the things we perceive in the visual field, and gloss them as "cat" or "coffee mug" or "Ricky," but we'd have a very hard time making those principles concrete enough to articulate and convey to a machinic system not overfond of ambiguity. What's more, we have an astonishing ability to keep track of the stable identity of things through relatively profound changes in state—we still recognize the cat in bright sunlight, Ricky after he's grown a beard, even the mug after it's been shattered on the floor—and this is still more difficult to account for. We might achieve these tasks with no discernible effort at all, but if machinic systems are ever to have the slightest hope of mastering them, they would have to be provided with some way of acquiring knowledge that does not involve explicit instruction.

Enter the neural network, a way of organizing individual processing units into meshes that mimic the way neurons are interconnected in the human central nervous system. In its basic contours, this idea had been floating around computer science for decades; the first conceptual glimmerings had come in a 1943 paper, and the first "perceptron," or artificial neuron, was built in hardware at Cornell in 1957.[5] Though they may not have borne fruit until much later, neural networks were by no means an intellectual backwater—if anything, they were the staple of artificial intelligence research throughout the 1980s, and were actually deployed at scale in commercial applications like check reading by the late 1990s. But riven by arcane doctrinal disputes, and undermined by the stark limitations of the available hardware, the field languished. It wasn't until the early years

of this century, and the belated refinement of these techniques, that computer science finally began to produce systems robustly capable of learning from experience.

The contemporary neural network is built on a layered model of perception. At its most fundamental level are processing elements called *input neurons*, which work just as the brain's neurons do, firing in response to a specific stimulus. In machine-vision applications, for example, these are tasked with detecting features like edges and corners, and are therefore responsible for the crudest binary figure-ground calculation: is there something in the image, or not?

If the answer is "yes," these primitives will be passed on to a higher layer of neurons responsible for integrating them into coherent features. As neurons in each successive layer fire, a picture of the world is filled in, at first with low conceptual resolution ("this is a line," "this line is an edge"), then with increasing specificity ("this is a shadow," "this is a person standing in shadow"). And then an accumulation of finer and finer detail until the criteria for top-level recognition are triggered, and an *output neuron* associated with the appropriate label fires: this is Ricky standing in shadow. The algorithm has learned to recognize the subject of the present image by attending to statistical regularities among the thousands or millions of such images it was trained on. And so it will be for each of the higher-level objects a neural network can be trained to recognize: they must be built from the bottom up, in a cascade of neural firings.

What gives the neural network its flexibility, and allows for it to be trained, is that the connection between any two neurons has a strength, a numerical weighting; this value can be modulated at any time by whoever happens to be training the algorithm. The process of training involves manipulating these weights to reinforce the specific neural pathways responsible for a successful recognition, while suppressing the activation of those that result in incorrect interpretations of an image. Over thousands of iterations, as the weightings between layers are refined, a neural network will learn how to recognize complex

features from data that is not merely unstructured, but very often noisy and wildly chaotic, in the manner of the world we occupy and recognize as our own.

Stripped of its mystification, then, machine learning is the process by way of which algorithms are taught to recognize patterns in the world, through the automated analysis of very large data sets. When neural networks are stacked in multiple layers, each stocked with neurons responsible for discerning a particular kind of pattern, they are capable of modeling high-level abstractions. (This stacking accounts for the "deep" in deep learning, in at least one of the circulating definitions.) It is this that ultimately gives the systems running these algorithms the ability to perform complicated tasks without being explicitly instructed in how to do so, and it is is how they now stand to acquire the capabilities we have previously thought of as the exclusive province of the human.

The training of an algorithm can happen in one of two different ways. In the more conventional *supervised learning*, an algorithm is offered both training examples and their corresponding labels. The task to be performed might be binary—given this series of transactions, some of which are known to be fraudulent, is this particular one valid, yes or no?—or categorical—given one of a series of images to be identified, which category does the depicted object belong to?—but both types of tasks will be trained by manually reinforcing the pathways that lead to a correct answer.

For example, a thousand files containing training images of three classic American muscle cars of the 1960s might be presented to an algorithm for analysis, along with the respective labels *1968_Camaro*, *1968_Mustang* and *1968_Charger*, and it will be asked to place each image in the correct category. At first, when presented with a fresh image, it will do no better than chance. But as its weightings are tweaked, the algorithm will learn to identify those features which are definitive of each one of the possible options before it: a distinctive grill or fender or hood scoop. Eventually, given an accumulation of such details, it will arrive at a judgment as to which category it thinks the car

in the image belongs to. (Some degree of anthropomorphism is difficult to avoid in describing how this process works; of course, the algorithm doesn't "think" anything at all, but has assigned a weighted score to the image, representing the probability that the car in the image is the one its calculations suggest.)

We might say, then, that the first goal of machine learning is to teach an algorithm how to generalize. A sound algorithm is one that is able to derive a useful *classifier* for something it hasn't yet encountered from the things it has been shown. Perhaps after reviewing its thousand images, for example, our algorithm concludes that a taillight configuration of six vertical rectangles is very highly correlated with the label *1968_Mustang*. When applied to a new data set, this classifier can be graded by the metrics of *accuracy* (were all of the images tagged as Mustangs identified correctly?), *precision* (were all of the Mustangs known to be in the set correctly identified as such?), and *recall* (of the known Mustangs, how many were successfully identified?). Precision is an index of an algorithm's discriminative quality, while recall is correlated with its ability to return a complete set of results.

Note that this particular classifier is a highly reliable discriminator, in that both Camaros and Chargers have round taillights. Nevertheless, it will still fail to capture some, and perhaps many, of the Mustangs known to be present in the training set—those not pictured from an angle at which the taillights are visible, for example. In other words, if used on its own, this classifier will produce high *accuracy* (zero false positives), but low *recall* (many false negatives). As we'll see, given that false positives and false negatives are not equally costly in the real world, this state of affairs may in fact sometimes be desirable; the important precondition to any valid use of learning algorithms will always be to ask careful questions about which metric actually matters most in a given situation.

But the relative weakness of taillight configuration as a classifier in this example also speaks to the difficulty of identifying which features of a data set might lead to a higher degree of confidence in identification. In order to permit ready

discrimination between alternatives—whether those alternatives are objects that might conceivably be present in an image, phonemes in an audio stream, or characters in a body of text—such a feature must be unambiguous, distinctive in some way and independent of any other variable. The effort involved in extracting appropriate candidates from a data set is called "feature engineering," it is still generally done manually, and it remains among the most time-consuming and expensive aspects of training an algorithm.

The problems we associate with bad machine learning are those that arise when an inappropriate feature is used in classification, or the process of abstraction otherwise goes wrong. Broadly speaking, there are two ways in which this can happen: *overfitting* and *bias*. Overfitting means that an algorithm has "memorized" training data rather than learning to generalize from it, which most often happens when the training set sharply diverges from what it experiences in the real world. Perhaps all of the Camaros our algorithm was shown in the training phase happened to be red, and as a consequence it has mistakenly settled on this feature as a definitive classifier, rather than an independent variable. It will therefore have problems with accurate identification when presented with a black Camaro.

Or it could suffer from the opposite problem, bias. In the context of machine learning, bias means that even after extensive training, an algorithm has failed to acquire anything essential at all about the set of target objects it's being asked to identify. An algorithm displaying high bias is basically taking random stabs in the dark, however much confidence it may seem to be mustering in its labeling, and will without hesitation identify outright static as a house, a whale or a chair. (We should be careful to distinguish this sense of the word from its more usual, pejorative sense, in which the implicit prejudices of the party responsible for training an algorithm are reflected in its output—though that happens too, as on the notorious occasion on which a Google Images algorithm identified a picture of black people as "gorillas," apparently because the only training images labeled "people" it had ever been provided had light skin.)[6]

However they might undermine an algorithm's practical utility, or embarrass the software developers involved, errors of bias and overfitting can be corrected. They will eventually yield to patient retraining, involving recalibration of algorithm's internal weightings. At the outside, a human teacher can directly furnish the learning system with heuristic cues, so-called "privileged information"; just as we might expect, these insights sharply accelerate and improve the algorithm's ability to recognize artifacts.

None of this will help, however, if a training set is simply too small or homogeneous to permit generalization to the contents of the real world. An algorithm raised up on such a set will only ever perceive what it has been taught to perceive, whatever should happen to be placed in front of it. It was largely this tendency, coupled to a positive feedback loop, that produced the wildly hallucinogenic images of Google's briefly popular Deep Dream software. The original Deep Dream filter was trained exclusively on the Stanford Dogs Dataset, a body of imagery produced for a competition in which machine-vision algorithms were tasked with distinguishing among 120 different canine breeds.[7] Little wonder, then, that it would see dogs in everything it was shown, especially when same image was passed through the algorithm again and again, sharply amplifying its quality of dogness.

The uncanny ability such an algorithm has to zero in on the all-but-inarticulable essence of something is illustrated by a neural network Yahoo is currently training to automatically characterize images as "not safe for work." When set loose to generate images of its own, it renders a fleshy, Gigeresque landscape that registers as entirely unwholesome, yet never quite resolves into anything specifically identifiable as obscene; whatever writhing, disembodied genitalia you may perceive in the images are artifacts of your own perception.

Done carefully and conscientiously, with whatever hallucinogenic potential it may hold carefully suppressed, supervised machine learning produces impressive results. It has certainly

been embraced at scale by commercial enterprises like PayPal and NBC Universal, who use the technique in training algorithms to identify potentially fraudulent payments, or predict which low-demand media properties can most profitably be moved to cheaper offline storage.

But these contexts are fairly static. They evolve only slowly, and so both an algorithm and its teachers enjoy the luxury of time. Faced with unacceptable results, trainers can tune the algorithm, run material to be classified through it, and tweak it again in response to the output, until they arrive at an outcome they're happy with. The challenge ramp ascends sharply, though, when systems are faced with a dynamic multivariate decision space, like directing a car safely through urban traffic. The software controlling a moving vehicle must integrate in real time a highly unstable environment, engine conditions, changes in weather, and the inherently unpredictable behavior of animals, pedestrians, bicyclists, other drivers and random objects it may encounter.[8] (Now the significance of those reports you may have encountered of Google pre-driving nominally autonomous vehicles through the backstreets of its Peninsular domain becomes clearer: its engineers are training their guidance algorithm in what to expect from its first environment.)

For autonomous vehicles, drones, robots and other systems intended to reckon with the real world in this way, then, the grail is *unsupervised deep learning*. As the name implies, the algorithms involved are neither prompted nor guided, but are simply set loose on vast fields of data. Order simply emerges.

The equivalent of classification for unsupervised learning is *clustering*, in which an algorithm starts to develop a sense for what is significant in its environment via a process of accretion. A concrete example will help us understand how this works.

At the end of the 1990s, two engineers named Tim Westegren and Will Glaser developed a rudimentary music-recommendation engine called the Music Genome Project that worked by rebuilding genre from the bottom up. (The engineers eventually founded the Pandora streaming service, and folded their recommendation engine into it.) Music Genome compared

the acoustic signatures and other performance characteristics of the pieces of music it was offered, and from them built up associative maps, clustering together all the songs that had similar qualities; after many iterations, these clusters developed a strong resemblance to the musical categories we're familiar with. Nobody had to tell the algorithm what a slow jam was, what qualities defined minimal techno, or how to distinguish Norwegian black metal from Swedish death metal: it made these determinations itself, inductively. The program may not have had any names for these categories, but given any specific piece of music, it was capable of assigning that composition to a cluster any human aficionado would recognize as being appropriate, instantaneously and with a very high degree of accuracy.

This is just how unsupervised learning would work when applied to our example. Rather than being assigned to a known category (i.e. *1968_Charger*) by an instructor, an unsupervised algorithm will group together an emergent cluster of all of the images sharing a particular constellation of features: this particular headlamp configuration, this distinctive paint scheme, this wheelbase. Three robust and well-defined clusters should emerge from the thousand images, possibly orbited by a few edge cases and outliers. In this case, the algorithm will probably not be able to identify the images as belonging to anything beyond *Cluster_1*, *Cluster_2* and *Cluster_3*, but provided with the labels for each would immediately be able to identify any newly presented images of cars. (We should also consider the possibility that it is working in synchrony with a text-recognition module capable of reading trunk-lid or fender badges, in which case it may very well be able to arrive at such determinations unaided.)

And this is the hinge at which machine learning meets big data. The more data it has available to train on, the better an algorithm is able to identify features and useful classifiers, and the more robustly defined the clusters it is likely to discern. The ultimate aim here is unsupervised feature learning, in which an algorithm builds up a sense of what is salient in the world

without anyone ever having told it what to look for. Such a system isn't merely developing a picture of the world from first principles, but doing just that from moment to moment. It's as if it were specifically designed to attack the old philosophical question of persistence—just what is it that binds the baby I was to the person I am to the corpse I'll one day be? Does this coffee mug or this chair retain coherent identity across contexts?

Such learning must be "unsupervised" because, again, these questions are no longer abstractions. The algorithms being trained in this way are intended to operate reliably in the real world, as the machinic faculty of discrimination at the heart of autonomous systems charged to operate in the vast unpredictability of the everyday—robots, vehicles, weapons platforms. The consequences of failure here can all too easily be fatal.

In mid-October 2015, the automobile manufacturer Tesla released a new version of the software running their Model S and Model X series cars.[9] Called 7.0, it was an example of the sudden upgrades in capability we are coming to expect from the software-driven objects all around us. This wasn't simply an incremental improvement, though: version 7.0 brought with it a much-anticipated feature that Tesla chose to call Autopilot. This made use of each car's existing suite of onboard cameras and sensors—a forward-looking, long-distance radar system to see through bad weather, cameras equipped with image-recognition software, and a battery of ultrasonic proximity sensors—to achieve a limited degree of autonomous operation.

As far as Tesla was concerned, this capacity was meant to augment, rather than supplant, human guidance. That it *was* limited, however, wasn't always clear from the company's official pronouncements. "The car can do almost anything," enthused CEO Elon Musk, talking up Autopilot at an unveiling event. "We're able to do lane keeping on freeways, automatic cruise control, active emergency braking ... It'll self-park. Going a step further, you'll be able to summon the car, if you're on private property." Anyone enticed by this reeling-off of capabilities—or his earlier brag that a driver could take a Model S from San

Francisco to Seattle "without touching the controls at all"—could perhaps be forgiven for missing the hesitant "almost" with which he hedged the claim.[10]

Musk further touted his product's almost uncanny ability to learn from experience, referring to each Model S owner as an "expert trainer" who could tutor Autopilot simply by driving with the mode engaged. Whatever measurements were captured in this way would then propagate across the Tesla network, furnishing the set of learning algorithms at Autopilot's core with an astonishing training set: one million miles' worth of high-resolution driving data added to the collective repository in this way, weighing up to 10 gigabytes per mile, each and every day.[11] Like no real-world product that came before, the entire fleet of Tesla vehicles running version 7.0 would build on whatever knowledge they derived from their encounter with the world, constantly adapting, constantly sharing, constantly improving.

None of that saved Joshua Brown, who was killed in May 2016 when Autopilot steered his 2015 Model S, at full speed, straight into the side of a tractor-trailer turning across his lane.[12]

Brown was history's first known fatality in a crash involving an autonomous vehicle. But he was also a longtime Tesla enthusiast, well known in the owner community; Musk had actually tweeted a video Brown made celebrating Autopilot to his 4.4 million followers, a little over a year prior to the accident.[13] By the standards of adroit public relations as well as those of common decency, the company was compelled to respond with some kind of statement.

Tesla chose to hang a generic title on its blog post commenting on the crash. It was called "A Tragic Loss," as though the victim deserved not a single syllable more than the absolute minimum gesture prescribed by courtesy; perhaps their attorneys had advised them that anything more specific would be injudicious.[14] The copy in the body of the statement was similarly *pro forma*, explaining that the "customer who died in this crash had a loving family and we are beyond saddened by their loss," as though the presence of a loving, or any, family was the only factor that made Brown's death worth mourning.

A dutiful recitation followed of all of the occasions on which the Model S user interface and documentation informed a driver of the "need to maintain control and responsibility for [their] vehicle" while using Autopilot. Even in context, it reads like something intended as nervous self-reassurance. Tin-eared though this may have been, however, it wasn't yet the oddest aspect of a very odd piece of writing. A single sentence buried halfway through the post delivers this disquieting explanation for the cause of the crash: "Neither Autopilot nor the driver noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied." Horrifyingly, Brown's Model S never slowed, never even thought to slow, because it didn't see anything it might have needed to slow for.

This apparent whiteout gives us some insight into the way a Tesla Model S equipped with Autopilot perceives the world. In failing to detect the outline of a white truck against a white sky, this ensemble of sensors and interpretive algorithms foundered at the most basic task of vision, resolving a figure from its background. That it did so is unsurprising, though, when we consider what Autopilot was actually designed to do: keep the car centered on well-maintained freeways with clear, high-contrast lane markings. In other words, the moment the first driver took to the roads in a car running Autopilot, a gulf opened up between what the function could actually do and its implicit premise—a premise underwritten by everything from Musk's public commentary to Tesla's choice of naming. Joshua Brown accepted that premise at face value, and it killed him.

In the wake of Brown's death, Tesla tried mightily to steer public opinion toward the conclusion that any lapse in vigilance was entirely his own. A followup post on the corporate blog, complaining about negative media coverage of the crash, reiterated the official position that Autopilot is nothing more than "a driver assistance system that maintains a vehicle's position in lane and adjusts the vehicle's speed to match surrounding traffic."[15] It didn't address why, that being the case, they hadn't chosen to call the feature Autolane.

It wasn't just the feature's name, of course, that left Brown with the idea that he was safe in relinquishing control of his car. Autopilot was introduced at a time of significant hype about autonomous vehicles, by no means all of it Musk's own, and if we understand the choice of name as being part of a commercial product-differentiation strategy, we also need to consider what Tesla thought it needed to be differentiated from. The notion that a car might safely operate itself under conditions of highway driving was in the air more broadly in the culture, and the release of Autopilot merely reinforced it.

Those most knowledgeable about the current state of the art in machine learning and autonomous guidance had argued that it was irresponsible of Tesla to release any such feature, whatever its name, until it was capable of delivering on everything it implied.[16] But the company's behavior was unsurprising given the pressures it was exposed to in the contemporary market economy. The idea that a manufacturer might hold off from shipping a feature until it can actually achieve all the things its prospective users have been led to believe it can feels like an artifact from a long-gone age of centralized production, when the entire decision nexus of an industry might feasibly be gathered at a single conference-room table. It has virtually no chance of prevailing in a world where productive capacity is so widely distributed. Some party will always be hungry to claim the first-mover advantage, to benefit from the perception of being first to market with autonomous capability, and will do assuredly do everything this side of the law to make promises that they've achieved it. It will take more than a disclaimer in the licensing agreement, or a few lines of boilerplate tucked away at the bottom of the press releases, to keep people from believing them.

The gulf between what we believe about automated systems are capable of and what they can actually do is only one of the frictions that confronts us as such systems like Tesla's Autopilot become ever more prominent in shaping the circumstances of everyday life. Not one of these systems is simple in itself, and they interact in complicated ways. They subtly alter the ways we see and engage with the world, and in particular they pose

troubling implications for our ability to apprehend the arrangements of power we contend with. But developing some sense of what they do is critical to understanding the deal we strike whenever we surrender control of a situation to the judgment of algorithms.

The tacit bargain that automation offers us is that in exchange for some perceived enhancement of performance, we relinquish discretion, and at least some degree of control over a situation. Sometimes this act of stepping away is trivial—virtually no one other than the individual dispossessed of a job directly suffers when robotized sleds replace human workers in the warehouse—and sometimes it feels more like a vital surrender of judgment, as in the development of that class of systems the United Nations refers to as "lethal autonomous robotics."[17] This suggests a common-sense way of reckoning with the impact of automated systems, the axes of what we might call a matrix of concern: the more people affected by a particular act of automation, the more vulnerable those people are, and the harder it would be to reverse its effects, the more cautious we should be in enacting it. Our task as a society would then be to determine just where in this envelope any given proposed displacement lies. By these lights, we ought to have a great deal of concern when someone is proposing to bring learning algorithms to bear directly to bear on decisions of great public consequence, on a population that is already at risk, with immediate and life-changing consequences.

As we've seen, data analytics is a fourfold process that involves *collecting* large volumes of facts about the world; *sifting* them algorithmically, to reveal whatever patterns are latent within them; *inspecting* those patterns to determine optimal points for intervention; and finally, *acting* on that knowledge to reshape the trajectory of the system being studied, so that its future evolution more closely conforms with desire.

This is a powerful and highly generalizable set of capabilities, and in principle it can be applied to the management of any complex system, from the steering and guidance of a car to the shaping of public policy. When applied to the maintenance of

public order, this capability is called *predictive policing*. The idea is that, equipped with nothing more than a sufficiently rich set of data on past incidents, public safety departments can predict crime hotspots, and even individual criminals, with a high degree of accuracy, far enough ahead of time that they are able to circumvent any actual offense. (Proponents of this approach invariably cite the "precrime" unit of Steven Spielberg's 2002 *Minority Report,* evidently mistaking the film's depiction of dystopian oppression for an aspirational goal.)[18] There are a few different approaches and strategies bound up in the practice of predictive policing, but what they all have in common is that they propose to sit in Olympian detachment, far removed from the play of events, and reach down into all the murk of our affairs to wrest the single salient truth from a whirling storm of confusion.

The simplest tools mobilized in predictive policing efforts, and in a way the most general, are dedicated to geolocating and otherwise parsing the things people say on social media, in the hopes of drawing actionable inferences from them. This is the province of "location-based social intelligence" applications like Snaptrends and SpatialKey, which promise to "identify, isolate and assess" threats, whether direct or indirect. A Snaptrends brochure for prospective customers in the law enforcement sector makes the proposition explicit: "From angry Facebook posts to suggestive Instagram uploads, today's would-be criminals often leave A STRING OF CLUES across social media," and a public-safety agency made aware of those CLUES can deploy its resources in time to preempt the commission of crime.[19]

Such tools use sentiment analysis, a facet of the emerging pseudoscience of "intent recognition," to extract actionable intelligence from utterances.[20] But it's astonishing that anyone takes sentiment analysis seriously in any but the most trivial applications, let alone what is all too often the life-or-death context of a police stop. The algorithms involved are notoriously crude and simple-minded, stumbling when confronted with sarcasm and other common modes of expression. They have trouble with word order, double negatives, ambiguous qualifiers

and inverted sentence structures.[21] In short, they simply cannot be relied upon to distinguish even the most obvious snark from a genuine CLUE.

More insidiously, context necessarily colors the interpretation of otherwise-innocuous utterances, especially utterances swept up by the kind of braindead, single-keyword searches Snaptrends showcases in its promotional literature. When I use the word "heroin" in a social-media post, I could be responding when asked to name my favorite Velvet Underground song, or discussing allegations that the CIA's Air America was involved in drug trafficking in Laos in the late 1960s, or misspelling the gendered term for the protagonist of a work of fiction, but what I am almost certainly not doing is openly offering the Schedule I drug known by that name for sale.

And yet this starkly unlikely scenario is what is being inferred in an act of collection of this sort. It will register against my identity, tagging me as well as anyone unwise enough to communicate with me as people with a known interest in illicit drugs, as we are snared in automated round-ups of "activity clusters" and "relationship networks." Should our activity climb past some threshold of intensity that is never specified by either Snaptrends or its client agencies, we may become the subject of intensified monitoring, or even physical surveillance. (Those worried about the clear potential for mission creep and overreach will surely breathe a sigh of relief at the reassurances of Racine, Wisconsin police public affairs officer Jessie Metoyer, who promised a reporter that, at least in the hands of her department, Snaptrends "would strictly be used for criminal investigation purposes.")[22]

Seemingly miffed by the implication that their product might entrain thoroughly illegal searches and seizures, Snaptrends falls back on a blame-the-victim strategy, arguing that if social media users don't want to wind up in the crosshairs of official attention, they might want to apply a greater degree of prudence in the things they choose to post. (A Snaptrends representative patronizingly explained that there can be "no expectation [of privacy] with open settings.")[23]

Even beyond the distasteful premise of this argument, it's unfair to expect this degree of sophistication from users in a milieu where privacy settings are often deliberately obscured—and for that matter, in which millions of Facebook users don't even understand that they're using the internet.[24] As absurdly, offensively shallow as all of this may be, it is what the automation of administrative awareness actually means. The only quality that distinguishes it from low comedy is the distant nagging understanding that life chances, and lives, hang in the balance wherever such tools are taken at face value by the undertrained, the uncaring or the outright gullible.

The next stage of "intelligent policing" is so-called redboxing, in which predictive algorithms preferentially dispatch police units to neighborhoods or specific locales considered to be at particularly high risk of crime. An analytic package called PredPol, which has been licensed by more than sixty American police departments, including those of Los Angeles and Atlanta, is among the redboxing market leaders—in fact, the company's logo is a stylized red box. PredPol furnishes patrol units with a list of the ten to twenty locations in their assigned area of operation where it believes crime is most likely to occur over the course of their shift.[25]

Different departments have their different ways of making use of this intelligence. The Modesto, California, PD parks their Mobile Overt Surveillance System vehicle, otherwise known as the Armadillo, in redboxed hotspots, as a heavily armored, all-seeing standing deterrent.[26] The LAPD takes an approach rather more suited to our mediated age, and tweets the location of identified redboxes so the civic-minded can keep a watchful eye on them. Still other departments simply "flood the zone" with patrol officers.

PredPol's site sports a variety of claims for the product's efficacy, all well larded with statistics and all worded with the most exquisite lawyerly precision. The company appears happy to leave a casual reader with the impression that use of its algorithm significantly reduces crime without ever actually coming out and making that claim, perhaps because it cannot.

But common sense suggests what will happen whenever a supervisory presence is brought to bear more heavily on one place than another: a higher percentage of whatever crimes that are being committed in that neighborhood will be detected, even if the baseline level of criminality in other places is identical, or still higher. That the obvious logical flaws in predictive approaches like this surrender to a few moments' consideration suggests a few basic possibilities, both disturbing: either their promoters don't know, or they don't care.

As Charlotte-Mecklenburg, North Carolina police chief Rodney Monroe points out, though, "We're not just looking for crime. We're looking for people."[27] And that is precisely the aim of what is without doubt the most notorious of current-generation predictive policing initiatives, the Chicago Police Department's so-called Heat List.[28]

This is an algorithmically compiled index of the 1,400 Chicagoans the city's police department considers most likely to commit, or suffer, homicide at some unspecified point in the future. It is not a matter of idle interest: having identified these individuals, CPD teams actually go out and visit them at their homes and street corners and places of work. Again, so far as the Police Department is aware, the named individuals have not yet committed any known crime of interest—and yet there they are, knocking on the door and asking to be let in for a brief chat. (I had promised myself that I wouldn't use the clichéd description "Orwellian" for any of this, but the official name for these visits, "Custom Notifications," really does seem to demand it.)[29] Whether this can be squared with the Fourth Amendment to the United States Constitution, which nominally protects citizens against unreasonable search and seizure, and the Fifth and Fourteenth, which guarantee them the right to the due process of law, is not yet a settled question. Concerns have also been expressed that the program amounts to a reintroduction of racial profiling through the back door—in other words that the Heat List is indeed, and in so many words, a *Minority Report*.[30]

But there's no way any of us can know for sure whether or not this is the case. Beyond generalities, its operators refuse to

discuss the most basic questions about this tool, like how you get on the list, how you get off it, who has access, how long it persists and how its use is regulated.[31] The most they're willing to admit is that prior arrests and conviction records are heavily weighted in its algorithm, with the by-now usual implication that those with nothing to hide have nothing to fear. You don't have to believe each and every last person on the list is a model citizen in order to be wildly disturbed by this—and it's still more disturbing to contemplate this capability in the hands, specifically, of the Chicago PD, a force with a documented penchant for rogue operations,[32] a record of having literally maintained black sites,[33] and a thoroughgoing culture of impunity.[34] (A similar program was piloted in London in 2014, the fruit of a collaboration between consultancy Accenture and the Metropolitan Police; it remains to be seen whether the Met's interpretation is any less problematic.)[35]

You might ask what the problem is with programs like Snaptrends, PredPol or Chicago's Heat List, if they keep innocent people safe and free from harm? What's so offensive about algorithmically mediated interdiction, if it keeps young black men especially from being drawn (further) into the clutches of a prison-industrial complex all but certain to grind them to dust?

One way of answering might be to point out that PredPol and similar redboxing tools don't so much criminalize behavior as they criminalize the simple fact of physical presence. Suspicion is shed upon you not for anything you've done, or even for anything you *might* do, but simply because you happen to occupy an area of interest. (As one skeptical criminal-justice scholar characterized the insinuation it presents to a patrol officer, "I go in this box, and everybody's Michael Brown.")[36]

In response, PredPol's chief of research and development Jeffrey Brantingham is quick to reply that "this is about predicting where and when a crime is most likely to occur, not who will commit it."[37] But this is sophistry. Brantingham's is a distinction without a difference—as he certainly would have known, given that the geosocial model prevailing in the field, "travel-to-crime,"

which asserts that most offenses are committed within a relatively short distance of a criminal's home or base, was developed by his parents, Patricia and Paul, in the 1980s.[38]

Another way of considering the question might call attention to the salience of that which is *not* being measured by these systems. In economics as in physics there is a property called "path dependence," which is the tendency of a dynamic system to evolve in ways that are determined by decisions made in its past. That system might have taken any number of different developmental paths at its outset, but once embarked on a particular course, the choice of trajectories it will enjoy as time unfolds is strongly constrained by the choices that came before. There is a very real danger of path dependence in the use of predictive analytics, based as they are on the notion that meaningful inferences about the future can be drawn from a consideration of the prior distribution of events.

Predictive policing may seem to be concerned with the future, in other words, but the future in question is one oddly entangled with the past. A neighborhood in which a statistically significant spike in felony assault has taken place may find itself the focus of intensive patrolling moving forward, leading to new citations for felony assault being issued at a rate far above the citywide average, and therefore new cycles of police vigilance. A teenager who was once tagged in a Facebook picture alongside friends throwing gang signs may be swept up a social network analysis, find her whereabouts, activities and patterns of association tracked, and eventually be cited for some trivial offense that anyone else (or even she herself, prior to the descent of this watchfulness) might have gotten away with.

She will, of course, thereafter have a criminal record; given that predictive algorithms are known to weight prior offenses heavily as predictors of future run-ins with the law, she will show up sooner and higher on all such rankings, even as other people—equally or perhaps far more inclined to criminal behavior—slip through the weave and evade detection. And this is even before considering the impact of those many varieties of crime that are corrosive of a community's trust, insulting to its

hopes and injurious to its fortunes, yet aren't measured by any kind of algorithm at all.

We have a word for all of this, and it's *bias*.

None of the operations of these tools are in any way free from human discretion, however much those responsible for engineering them might want us to believe otherwise. Heat List developer Miles Wernick, by training and experience a specialist in medical imaging, defends his creation against charges of racial, or any other, bias by claiming that the algorithm is intended "to evaluate the risk of violence in an unbiased, quantitative way." A representative of the organization sponsoring Wernick's work, the National Institute of Justice, expands on his point: the individuals named on the Heat List "are persons who the model has determined are those most likely to be involved in a shooting or homicide, with probabilities that are hundreds of times that of an ordinary citizen."[39]

To be sure. But we constantly need to remind ourselves that somebody designed that model—if not Wernick himself, then some other specific, identifiable actor, operating inside history. Somebody selected its sources, devised its features and weightings, or at the very least validated that some attribute happened upon by an automated feature-extraction process was indeed a likely signifier of criminal intent. At every step of the way, human judgments were folded into the ostensibly neutral operation of the algorithm.

Proponents argue that these tools transcend the fallible knowledge, the profoundly situated experience and the variable training of the individual public-safety officer, and supplant it with a cool and dispassionate collective intelligence derived from a million points of data. But what is that intelligence other than a distillation of the way we've chosen to order our societies in the past?

The choices we make in designing an algorithm have profound consequences for the things that are sorted by it. Even the choice of weighting applied to a single variable can lead to different effects in the application of an algorithmic tool.

Let's say that as a municipal administrator concerned with the maintenance of good order and the protection of the citizenry, you want to flag and neutralize as many potential murderers as possible, before they're able to do any harm. Your review of the data offers you only a few selectors to work with, but you eventually determine that of the seven individuals charged with homicide in your district in the past six months—*charged with*, mind you, not *convicted of*, because a separate agency holds that set of data, and you don't have access to it—100 percent of them are males from single-parent households, between the ages of eighteen and thirty, who lack anything beyond a high-school education. Each of them naturally has other qualities, life experiences and attributes, but this is the only set of features they all share. And so this becomes the cluster of features around which you develop your predictive algorithm. You have chosen to optimize for *recall*.

In the six months that follow, every time someone who matches these criteria comes into your field of awareness, in the course of a traffic stop or an unrelated investigation, his file is flagged for intensive follow-up. This means not seven, not eight or nine potential murderers have been diverted into your intervention program, but hundreds of them, each of whom precisely conforms to the contours of your model. And you can get in front of the press and the public, and tell them with a clear conscience that your model is *clean*. It never once mentions race, or anything like it. It is as limpidly neutral as can be.

Is it, though? Is there any way in which your set of sorting criteria might strongly correlate with other features of the target set—features that no ethical designer could ever legitimately consider, like race or income? Not that you would intentionally choose your selection criteria as a proxy for those features, of course, but it will be very hard for you to argue that you are entirely free from the mire of the past.

And just as important, are those factors in any way meaningfully predictive of a propensity to commit homicide? There is in principle no way of measuring the frequency of events that have failed to happen. But let's, for the sake of

argument, assume that your algorithm didn't miss a single one of the residents of your district who would have gone on to pull the trigger on someone in that six-month period. Either what you developed is just preternaturally accurate or, what is far more likely, some false positives have been folded up into its assessment of likely criminality. In this case, that anodyne technical term—*false positive*—means that entirely innocent people have been saddled with the identity "criminal" and swept up into your dragnet, with everything that implies for their life chances. And in the United States, anyway, this is clearly illegal: it blatantly violates the Constitutional guarantees that all citizens in principle enjoy equally.

So in the United States, if the law is to be observed, recall *can't* be the criterion that is emphasized in the design of a predictive policing algorithm. It has to be accuracy. A successful system, by these lights, would necessarily tolerate some false negatives to ensure that it doesn't entangle any false positives. Making the terms of this bargain explicit: some actual bad actors will escape your net of computational awareness—and presumably go on to cause harms that theoretically could have been prevented—because the alternative is Constitutionally and ethically intolerable. Other societies could, of course, arrive at just the opposite determination: that sweeping the occasional innocent into the clutches of an algorithmic gill net is the regrettable, but eminently acceptable, cost of full assurance. Thankfully, that's not the society we happen to live in at the moment. You, as the party responsible for the design of a predictive algorithm, can choose to do otherwise.

We are told that the Heat List works—that 70 percent of the people shot in Chicago during the first six months of 2016 were already on it, and 80 percent of those arrested in connection with these incidents.[40] But that definition of "working" is difficult to square with the reality that the number of homicides committed with a firearm have continue to rise in the city since the List's introduction and the advent of Custom Notifications. Chicago police superintendent Eddie Johnson explains this as a shortfall in execution, rather than conception: "We are targeting

the correct individuals, we just need our judicial partners and our state legislators to hold these people accountable."[41]

And more pressing still is the question of what using a tool like this does to us. If a tool like the Heat List "works," what was the cost of that efficacy?[42]

The promise of preventing some future harm seems to justify just about any action taken in the present. It's hard to argue with this when the future harms imagined involve a level of everyday violence no one should ever be asked to become used to. But the very first thing we learn when we evaluate systems like PredPol and the Heat List is that the consequences of adopting them cannot in any way be said to break over us equally. The geographer Ben Anderson makes this uncomfortably plain in his account of the way these systems work: "Certain lives may have to be abandoned, damaged or destroyed in order to protect, save or care for life" that is considered to be more valuable.[43]

This is an explosive thing to admit, especially at a time when the Black Lives Matter movement is bringing sustained attention to bear on issues of structural injustice, reflexive overpolicing of communities of color, state violence, and impunity for state actors implicated in that violence. We can be sure that in no society will the terms of this bargain ever be spoken aloud by the parties proposing it, and certainly not in so many words. But we shouldn't fool ourselves as to what's actually happening when we embrace tools that claim to magick away centuries of discrimination.

And this speaks more deeply still to the question of automation, and all the contexts in which it might be welcomed for its supposed rationality, objectivity and neutrality. The evidence presented to us by the current generation of algorithmic tools suggests that this is a fool's errand, that there can and will be no "escape from politics" into the comfort of governance by math. What we will be left with is a picture of ourselves, a diagram of all the ways in which we've chosen to allocate power, and an unforgiving map of the consequences. Whether we will ever summon the courage to confront those consequences with integrity is something that no algorithm can decide.

What happens when pattern-recognition systems disclose uncomfortable truths to us, or at least uncomfortable facts?

We hardly lack familiarity with the conscious introduction of uncomfortable facts into public debate. Self-delighted pop contrarians like Malcolm Gladwell and the *Freakonomics* team of Steven D. Levitt and Stephen J. Dubner have built careers on observing seemingly counterintuitive correlations that turn out to have a reasonable amount of explanatory force—for example, claims that the observed downturn in violent crime in the United States following 1991 can be traced to the more liberal access to abortion that American women had enjoyed starting twenty years earlier.[44] Their arguments tend to take the form "everything you think you know is wrong," and despite what might appear to be a slap-in-the-face quality, they're easily assimilated by the mainstream culture. If anything, the factoids dispensed by observers like these often become part of the conventional wisdom with astonishing rapidity.

But the reason why these narratives get adopted so quickly has a great deal to do with their inherent conservatism, the ways in which they can be wielded to support prejudices with existing potency in the culture. What if an algorithmic trawl through the available data surfaces a significantly more abrasive pattern of facts, something that's harder to square with the way we'd prefer to present ourselves and our institutions? For example, what if a multidimensional analysis conducted for a big-city police department revealed, with absolute statistical certainty, that hiring mildly overweight white male veterans of the US armed forces between the ages of twenty-five and forty-five, who purchase domestic beer in cans and consume mixed-martial arts media, is overwhelmingly correlated with use-of-force incidents and subsequent liability claims? Is this the kind of interruption of conventional wisdom that would easily be tolerated?

We've seen that a coffee mug or a curb can be identified by machine-vision systems with relative ease. But most of the objects and other features an algorithmic system will be tasked with characterizing have identities and meanings somewhat

more charged than that—and this, of course, is where things start to get complicated. You can teach an algorithm to recognize a table readily enough, based on its characteristics and the ways in which it relates to the world's other contents. It might be able to identify, with successively finer degrees of precision, a *vehicle*, a *car*, a *police car*, a *New York City police car*. That's straightforward enough. But how do you teach it to recognize *poverty*?

Or *do* you teach it to recognize any such abstraction in the first place? We assume that if an algorithmic system is to have effective agency in public affairs, it must respond to the same categories we do. What is more likely, however, is that an unsupervised learning system will have no *a priori* notions of "person" or "community" at all, let alone "taxpayer" or "citizen" or "grievance." Where we might think of ourselves as working class, or Scots Irish, or Sikh, or a San Diegan, or a Republican, none of those categories mean anything to a learning algorithm, except possibly as tags for closely correlated syndromes of human behavior.

A learning algorithm will derive the categories that are salient to it, building them from the bottom up. Here the rhetorical function of data in the sense we're accustomed to—as something marshaled in support of an argument—is inverted, as the patterns and syndromes of fact disclosed to us instead begin to *suggest* arguments that might be made about the state of the world.[45]

And as a result, such systems may, just like a child, innocently come up with fairly pointed and uncomfortable questions. Why *does* this group of people not receive as many resources as those others, when it clearly limits their ability to act in the world? Why *are* service calls in this district responded to so much more quickly than those originating in this other neighborhood? So long as such questions do not appear to originate from some tacit bias within the algorithm itself, I believe (to paraphrase Brian Eno) that they ought to be honored as a hidden intention wherever they arise[46]—a gift from the collective unconscious, and a rare hint that the most effective way of solving the problem

at hand might involve frontally engaging a set of circumstances we ordinarily prefer to ignore.

Sometimes this will involve asking pointed questions about the nature and intended purpose of the sensemaking tools we are offered. The premise of algorithmic technologies is not merely that they detect patterns, after all, but that they help us *recognize* them, and this in turn implies that there is something semantically meaningful to us in that which is identified. Why is this object of interest? What does our interest in it imply?

In learning to question what motivates the design of our sensemaking tools, we might want to ask, for example, what desire is being spoken to when machine-vision engineers devise an algorithm that sorts people passing through the gaze of a camera by gender. The justifications underlying the development of such an algorithm range from ends most of us would be likely to endorse—the automated characterization of images circulated online by child pornographers, for example, to aid in the protection of the children involved[47]—to others we might be far less comfortable with, many of which are founded in the fact that women and men have differential value as audiences for advertising.[48] We should be attentive to the reasons why a specific party proposes to deploy a specific technology in a specific context, and sometimes the answers to such questions will indeed tell us everything we need to know.

But there's an additional factor complicating our evaluation of algorithms belonging to this particular class, and it's independent of any justification that may be offered for their use. At the current state of development, when an algorithm proposes to "determine gender," it does so by retrieving measurements of facial structure from an image—jawbone length, distance between the eyes, and so on—and comparing these values to the ones associated with the label *male* or *female* in whatever set of images it was trained on.[49]

Biology may not be destiny, in other words, or gender itself anything but a performance,[50] but you wouldn't know any of that from reading the descriptions of systems like these, in which advanced methods like genetic algorithms and support

vector machines are marshaled to render a simple binary decision.[51] Whatever the justification behind deploying a system based on such methods, all questions of identity, fluidity, multiplicity or an individual's right to construct the way they are perceived by the world are here foreclosed, while a certain degree of misgendering is automated.

Not every algorithm is going to face complexities of this exact type, of course, but here bias (an incomplete or inadequate view of the world held by the algorithm's designers) and legibility (the sifting of a set of facts so as to render the patterns within it available for inspection) combine to produce an effect I think of as *overtransparency*. This is a surfacing of some state of affairs, whether based in fact or "fact," that causes a significant degree of social friction or harm, and it is bound to be a routine property of our broader embrace of algorithmic orderings. Preventing the emergence of situations like this, keeping automated systems from drifting back toward the inscription of received social categories, will require constant vigilance, some degree of technical sophistication and the mobilization of opinion—and all of these things require the exertion of energy. What I worry is that those with the most to lose from overtransparency may have the least energy available with which to counter it.

The burdens of overtransparency, perhaps unsurprisingly, will weigh particularly hard on the poor and the powerless. But some portion of that burden will fall on every one of us, whatever our status or situation. For example, when walking down a city street, we still tend to nurture the unconscious assumption that we are somehow insulated in our privacy by the others surrounding us. But the advent of powerful facial-recognition algorithms, and particularly the escape of those algorithms from their original context, threatens our ability to remain anonymous in this way—and by extension, our ability to assemble in public, demonstrate collective grievances and assert popular power.

This is the lesson of the recent Russian application FindFace, which lets users upload a picture of someone unknown to them, and compare it to those shared to the Russian-language

social network Vkontakte by its roughly 200 million users. FindFace's primary innovation isn't so much its raw pattern-matching ability—so far, the matches it comes up are accurate only around 70 percent of the time—but its speed; developer Alexander Kabakov brags that "[w]ith this algorithm, you can search through a billion photographs in less than a second."[52]

It didn't take users much longer than a second to figure out what they wanted to do with it. By the time FindFace had been in the wild for a month, it had already been used by a photographer to identify hundreds of random strangers riding the St. Petersburg metro,[53] and by an organized cabal of misogynist trolls to out and otherwise harass women working in the sex industry.[54] What may have seemed like an amusing party trick when described in the abstract begins to look a lot more serious when packaged as an app and made available to a broad public. The stakes get higher still when that capability is grasped by the state: Kabakov and his partner are currently concluding an agreement with the Moscow city administration to furnish the municipality's 150,000 CCTV cameras with their face-recognition algorithm.

This story epitomizes so many of the more troubling aspects of our encounter with algorithmic tools, all at once. It demonstrates the modularity of technology, how easily an algorithm developed in one context can be ported to another. It demonstrates how a developer's commercial interest so often overwhelms any concern they may have preserved for ethical behavior, or the fortunes of anyone affected by the tools they bring into being. It surfaces and makes plain the violence that has always been implicit in the power to see and the power to sort. Most specifically, it demonstrates how assumptions that have framed urban experience since human beings first gathered in cities are being undermined by newly emergent technical capability.

There are, of course, other ways in which the advent of overtransparency threatens freedom of assembly. So-called "group event detection" algorithms applied to the real-time analysis of video allow police forces to determine when a group

of two or more has formed.[55] Simulations of crowd behavior are used to better understand how social disturbances arise, pinpointing the "catalyzed space-time clusters of rebellion" unrest ripples out from, and identifying how those clusters can be disrupted.[56] Still other algorithms determine how many troops will be needed to suppress outbreaks of disorder, built right into municipal management systems in so many words, as a dropdown menu option available to administrators.[57] As applied to the city, the language of algorithms is that of "anticipatory surveillance," "scalable anomaly detection"[58] and preemptive control. The kind of conclusions that drop out of this body of work ("to quell the riot, you have to arrest 40 percent of the rioters") chill the blood, especially when coupled with the ability to identify specific individuals of interest as they move within the surging crowd.

But equally important is that virtually none of the algorithmic tools used in crowd control were originally developed for this set of applications. A learning algorithm that has outstripped the baseline of human cognitive performance may be of little enough concern in the lab, or even as part of a trade-show demo, so long as those things are self-contained and inaccessible. The moment it escapes from that context, though—whether its source code is uploaded to the GitHub repository, published in an academic journal, patented and made available for licensure, or simply reverse-engineered by another party—it is in the wild, and can be folded into any number of other systems, advancing ambitions arbitrarily remote from any it was developed to serve. And so it is that code leaps from one platform to another, like a plasmid swapped between organisms in the shallow primordial seas.

Kabakov may have intended FindFace as a diversion, or possibly as a pretext to flirt with women he wouldn't have dared to approach otherwise. The authors of the group-detection software were probably sincere about deploying it in the context of group homes and eldercare. But these technologies have transparently obvious political implications for people who live in places where the freedom of assembly is not guaranteed.

And you could not hand an authoritarian government a more perfect tool for the application of draconian hygiene than something capable of alerting the secret police that a knot of potential dissidents has formed, and identifying them by name. (Sometimes, indeed, little repurposing is required, especially when market actors work in close concert with the state. When the *MIT Technology Review* reports that Chinese search giant Baidu is able to use map searches to "determine, up to three hours in advance, when and where a dangerously large number of people might congregate," it's not at all hard to imagine who their prime customer might be.)[59]

Security expert Bruce Schneier is eminently correct to remind us that "many of these technologies are nowhere near as reliable as claimed."[60] But again, just as with Tesla's Autopilot, the meaningful question isn't whether these technologies work as advertised. It's whether someone *believes* that they do, and acts on that belief. In the end, the greatest threat of overtransparency may be that it erodes the effectiveness of something that has historically furnished an effective brake on power: the permanent possibility that an enraged populace might take to the streets in pursuit of justice.[61] In this light, these algorithms should be seen for what they really are: a series of technical counters to liberty, and steps toward the eclipse of freedoms we have enjoyed since the dawn of the modern public.

Among the most disconcerting aspects of the world we are building is that we will never know the reasons underlying a great many of the things that happen to us in the course of our lives. Already a literal and uninflected description of daily life sounds like nothing so much as the conspiracy theory of a paranoid schizophrenic: we're surrounded by invisible but powerful forces, monitoring us from devices scattered throughout our homes, even placed on our bodies, and those forces are busily compiling detailed dossiers on every last one of us. They pass the contents of these dossiers onto shadowy, unaccountable intermediaries, who use everything they learn to determine the structure of the opportunities extended to us—or, what may be

worse, not extended. We'll be offered jobs, or not; loans, or not; loves, or not; cures, or not. And the worst of it is that until the day we die, we'll never know which action or inaction of our own led to any of these outcomes.

This account is enough to stir up visions of Kafka, Borges and Philip K. Dick huddled up in some damp and miserable afterlife, plotting their hundredfold revenge on humanity for some long-forgotten transgression. You wouldn't necessarily want to repeat it word-for-word to a cop, or an intake counselor, or anyone else you needed to convince of your stability and levelheadedness. But there it is, laid out in schematic: the terms under which we now live out our lives.

As the examples of PredPol and the Heat List demonstrate, our ability to inspect the way in which algorithmic power is exerted in the world is already complicated by the impenetrability of the systems involved. They're proprietary business secrets, or the details of their construction aren't being shared with us because some shadowy bureau has determined that their disclosure "would endanger the life or physical safety of law enforcement personnel or any other person." Or it's simply that their guts are lying open before us—every line of code commented with the greatest conscientiousness, the name of every register plain as day—but the whole utterly taxes our ability to comprehend.

As legal scholar Frank Pasquale points out, algorithmic systems are the proverbial "black boxes," in that they produce material effects in the world without necessarily revealing anything about how they did so.[62] This profound murk hampers our ability to evaluate whether or not we feel that the algorithms operating on us are acting in ways consonant with our values.

Whether wielded by a market actor or an institution of state, then, the reasoning behind the judgments rendered by such black boxes is often unavailable for inspection—and this is most likely intentional. Among Pasquale's fundamental points that the structure of what we do and do not know about the way these algorithms work is a site of the most intense interest. Quite simply, some parties derive advantage from the fact

that we don't understand the tools used to rank and order us. And this results in a pronounced and troubling asymmetry in the world, when the actors in a position to determine our lives know far more about us than we know, or will ever be able to find out, about them.

The circumstances that are determined in this way aren't simply which songs a streaming service will choose to play you, which restaurant you'll be steered toward upon arrival in an unfamiliar city, or which driver Uber will send to pick you up at the tail-end of a Friday night on the town. They're far more consequential decisions than that—life-altering, even. We've already seen how an HR manager equipped with a workplace-analytics suite can use it to decide questions as laden with import as who to hire and who to let go, how the exercise of the law and the operations of the criminal justice system are equally shaped by the use of algorithmic assessment tools. Similar processes guide the apportionment of financial resources, enhancing or undermining our ability to function as independent actors in the economy.

Of the four major ways in which households in the developed world are sustained economically—via formal employment, the extension of credit, capital gains (at the high end) and government transfers (at the low)—algorithms already condition access to three, and will certainly determine the choice of products you are offered should you be fortunate enough to require the services of a wealth manager. Those algorithms are developed by parties who answer to no one other than their clients or employers, and the tools they produce are almost never assessed on any criterion other than the minimal one that they are broadly seen to work. We should understand this as what it is: an unprecedented intervention by a small set of private and unaccountable actors in the structure of opportunity, and the distribution of life chances.

Among the financial circumstances that are determined in this way, the one with the most pervasive reach is credit score. Via an entirely unsurprising process of mission creep, a narrow and algorithmically determined creditworthiness has become

an index of reputational worth that affects your fundamental ability to participate in a fully formalized economy.[63] Just as the Social Security Number was pressed into service as a *de facto* national identity number, so too has credit score been deployed as a selection criterion in contexts it was never intended for and never designed to function in. An individual's credit score affects their employability—a 2013 report prepared by the Demos public-policy research organization concluded that nearly half of all employers used the index to determine hiring decisions for some or all positions[64]—their access to housing, even their access to that most vital of contemporary utilities, a mobile-phone service plan. (If you doubt this last, try signing up for a mobile plan in a country where you lack a credit history.) And all of this, of course, becomes still more important in a time when the state has broadly retreated from the provision of benefits.

Once someone is past the age of majority, moreover, their lack of a credit history is not a neutral fact. It's a charged lacuna, something that can be interpreted as a positive suggestion that one's financial activities are informal, offshore or otherwise illicit—or simply, and perhaps more damningly, that they just aren't reliable in the ways our society constructs reliability. (I think of the "Credit Poles" in Gary Shteyngart's mordantly dystopian *Super Sad True Love Story*, lampposts topped with LED signboards that display pedestrians' credit ratings in real time, and blink a damning red when one's score falls below the threshold.)

The extension of credit operates obscurely, in ways that seem designed to confound oversight and to route around regulations on the way in which creditworthiness is calculated. Consider an algorithm currently being used to assess creditworthiness, based on "subtle patterns of behavior that correlate with repayment or default"[65]—in this case, patterns of mobile-phone usage. This algorithm has been developed by a startup named Branch.co, that seeks to extend financial services to the same market of "the unbanked" we encountered in the context of cryptocurrency technologies. Branch uses both data—the

content of text messages and emails—and metadata—the frequency and duration of calls—to build a character model of its subscribers, even weighing whether or not you've bothered to furnish the contacts in your address book with last names. The price of noncompliance with their model of good character is punitive: the interest rate such low-scoring borrowers are assessed literally doubles.

Branch, like many institutions in similar situations, presumably keeps the precise composition of its risk assessment algorithm secret for two main reasons. The first is simply that they derive value from its being a proprietary trade secret, or believe that they do. They think that it gives them a competitive advantage, and they don't want rivals nullifying that advantage by copying it. That part is straightforward enough. But the second reason is that, like all such metrics, these stats can be juked: Branch's algorithm is subject to Goodhart's Law, the principle that "when a measure becomes a target, it ceases to be useful as a measure."[66] In other words, they believe that if it became more widely known just how their algorithm arrived at its determinations, it would be easier for unreliable people to act in ways that would fool it into classifying them as trustworthy.

On the surface, then, this is the same reason that Google holds the precise composition of its search algorithm closely: to prevent it from being gamed by interested parties. But altering a web page so that it might rise higher in a ranking of search results is relatively uncomplicated. By contrast, performing good citizenship in the way Branch's algorithm would require is exhausting; considering the number of separate factors it weighs, and the semi- or even subconscious level at which some of them operate, it may not even be possible. Who, after all, is capable of maintaining conscious control over all the signals we broadcast through our behavior, at the level of data and metadata both?

Here a judgment is being made about what it is that makes a specific human being reliable, in a very narrow context, and has encoded that judgment in a numeric value. That score thereafter serves as a global representation of that person's character. Having developed such a representation, Branch, like any other

party in its position, can either license it to other companies as a stand-alone index of reliability, or provide it through an API so it can be folded into some other machinic weighting. And so the judgement once made spreads across the network, and shows up in any number of remote contexts, very much including ones it may never have been intended for.

To recap: *we don't know if the information on which a determination of creditworthiness was founded is correct.* The parties that develop such scores almost never take responsibility for founding prejudicial decisions on bad data—at best, perhaps, they delete that data, rarely with so much as an apology tendered. By the time they do make this correction, though, it may be too late; the information has already cascaded onto other commercial partners, data brokers or other third-party service providers, either in itself or as bundled into an aggregate score.

As well, *we don't know if the algorithm complies with the relevant law.* In the United States, for example, the Federal Trade Commission's inventory of Equal Credit Opportunity Rights explicitly "prohibits credit discrimination on the basis of race, color, religion, national origin, sex, marital status, age, or because you get public assistance," and this is intended to protect certain classes of people who have historically been denied access to financing.[67] Without access to an algorithm, there is no way of knowing whether it observes those provisions—or, perhaps more worryingly, whether the behaviors it weighs transparently serve as proxies for factors that lenders are specifically forbidden by law to consider in their provision of credit.

And finally, without access to its composition, *we can't reconstruct whether the conclusions an algorithm arrives at bear even the slightest relationship to someone's actual propensity to repay a loan.* Like any other sorting algorithm, the ones used in the determination of creditworthiness always direct our attention to a subset of the information that is available. That information may have less bearing on someone's trustworthiness than other facts which might well be more salient, but which by their nature are less accessible to the lender. The mathematician

and alternative-banking activist Cathy O'Neil has documented, for example, that lenders systematically refuse credit to borrowers on the basis of "signals more correlated to being uneducated and/or poor than to the willingness or ability to pay back loans," and these signals can be as arbitrary as the fact that they exclusively used capital letters in filling out their loan application.[68]

There might very well be other information that casts a specific individual's reliability in a much better light, but simply isn't available to the lender in numerical form, or available at all. Perhaps behavioral models will improve, as lenders sweep up ever-larger bodies of correlated fact; one German provider claims to use 8,000 data points in determining borrower reliability.[69] But perhaps these models won't actually get any more accurate—and the point is that, in the absence of any right to inspect them, there will be no way any of us will ever know for sure. As long as the systems are "working"—that is, they are producing net benefit and a positive return on investment—any concern for mistaken results, whether it involves the production of false positives or false negatives, can be waved away as a quibble. As things stand now, there is little to no incentive for anyone to fix the situation, and this is especially distressing when that same credit score conditions access to so many of the other goods produced by our society.

For many years now, ever since it first became clear that control over so many of our life chances had passed into the hands of parties equipped with tools like these, concerns about the obscurity of their functioning have prompted calls for "algorithmic accountability." This effort has notably picked up momentum in recent months, culminating in the framing of measures like the European Union's new General Data Protection Regulation, scheduled to take effect in April 2018.[70]

The law has two major provisions. The first is intended to protect vulnerable people from the consequences of automated decisions, and it articulates a series of categories still more comprehensive than the one enunciated by the US Federal

Trade Commission: "racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership... data concerning health or data concerning sex life or criminal convictions and offenses."

The desire to protect the vulnerable is, of course, entirely laudable. We've already seen, though, what the problem is with articulating lists of protected categories like this, which is that certain kinds of innocuous data can be used as proxies for factors that developers are forbidden to use in crafting an algorithm. If the law prevents you from using household income as a determination factor in choosing whether to offer someone a loan, you can just use their postal code, which will after all tend to be strongly correlated with income; if health status and medical history are off limits in making a decision about insuring someone, use their browser history, and mine it for its predictive value.[71] If a regulation bans the use of specific items of sensitive data, it leaves open the possibility that proxy values can be found that produce precisely the same discriminatory result. Conversely, as we cannot even in principle specify ahead of time what kinds of correlations might emerge from the analysis of a sufficiently large data set, the only way to prevent all such correlations from being used with discriminatory intent is to ban data capture in the first place—and that's obviously off the table in any technologically advanced society. As Oxford researchers Bryce Goodman and Seth Flaxman point out, then, the EU regulation is either too narrowly written to be effective, or so broadly interpretable as to be unenforceable.

This suggests that it isn't so much the obscurity of any specific algorithm that presents would-be regulators with their greatest challenge, but the larger obscurity of the way in which sorting algorithms work. And this impression is reinforced by the law's second major provision, which aims directly at the question of algorithmic opacity. Its Articles 12 and 13 create "the right to an explanation," requiring that anyone affected by the execution of some algorithmic system be offered the means to understand exactly how it arrived at its judgment, in a "concise, intelligible and easily accessible form, using clear and plain language."

Again, laudable—and again problematic, on two grounds. The implicit logic operating here is that once we are furnished with an explanation, we will be able to act on it in some way. But this places the burden of responsibility on the person the law refers to as the "data subject," who is required to seek out an explanation, and then exercise prudence in their choices once it's been provided to them. The right enunciated here is thoroughly consonant with the neoliberal practice of governmentality, which tends to individualize hazards and recast them as issues of personal responsibility or moral failure, rather than structural and systemic issues. It's a conception of good governance that conflates transparency with accountability: if the information is available, you're expected to act upon it, and if you don't, it's nobody else's fault but your own. This clearly relies entirely too much on the initiative, the bravery and the energy of the individual, and fails to account for those situations, and they will be many, in which that individual is not offered any meaningful choice of action.

Furthermore, this sort of accountability is ill-suited to the time scale in which algorithmic decisions take place—which is to say, in real time. Explanation and redress are by definition reactive and *ex post facto*. The ordinary operation of a sorting algorithm will generally create a new set of facts on the ground,[72] setting new chains of cause and effect in motion; these will reshape the world, in ways that are difficult if not impossible to reverse, long before anyone is able to secure an explanation.

It's evident that the authors of this well-intended regulation either haven't quite understood how algorithms achieve their effects, or have failed to come up with language that might meaningfully constrain how they operate. Their perplexity goes to a deep feature of the way in which predictive algorithmic systems work. Predictive analytics is all about discovering reliable correlations between two seemingly unrelated patterns of fact—for example, between a person's propensity to fill out a loan application in uppercase letters, and the likelihood that they will eventually default on that loan. But as Goodman and Flaxman point out, there's never any concern

for causal reasoning involved in making this correlation, nor any attempt to work out how or why these two observations might be related to one another. Nobody's arguing that someone's idiosyncratic spelling practices *caused* their shaky financial situation; in fact it's highly unlikely that there's any direct connection between the two to speak of. Both are epiphenomenal of some deeper syndrome of behavior. And while that syndrome might well be of interest to a psychologist or a social worker, it's completely immaterial to a prospective lender. From their perspective, it's enough simply to note that a correlation exists, and that it's sufficiently robust to permit the presence of the one to serve as a predictor of the other. So much for the right to an explanation.

And this begins to gesture at the ultimate complication with laws designed to produce algorithmic accountability. It's one thing to feel like you're in the grip of someone else's agenda, that you don't (and won't ever) know how selecting one or another among the options you're being presented with might serve the shadowy ends of another; still worse is the fear that there is no overriding logic at all to the decisions that shape our lives, that these systems behind them exercise their considerable power in an arbitrary and capricious manner. Perhaps worst of all, though, is the fear that there *is* a logic behind such decisions, but that it resides on a plane of complexity permanently inaccessible to the human mind. This is the realm of "opaque intelligence."[73] Who can say, in a layered, cascading, probabilistic model of behavior, what originally triggered a determination that someone is trustworthy, insurable or reliable?

This is not hypothetical. It is affecting the choices we are being presented with right now. Many of the systems we already use every day work in ways that are not fully understood by their designers. On Facebook, for example, "there is no way to know with any certainty why any specific [news item] is included or missing from" the ticker of Trending News stories.[74] The algorithm that makes that determination has already breached the threshold of incomprehensibility. As internet researcher Christian Sandvig told *The Intercept*, the reason that a particular

story or controversy appears or does not appear in that list "may not be recoverable." The one that governs the appearance of Trending News is far from the only such algorithm out there, sorting, ordering and classifying as you read these words, and doing so in a way that no human being alive will ever be able to account for.

Calls for algorithmic accountability face the most severe impediments when the computational models in question might have evolved on their own, without the involvement of any human programmer. This is the principle behind the development of so-called genetic algorithms; the technique is not universally applicable, but can often result in stunningly effective designs.[75] And when it does, no human mind will ever be able to account for the decisions it has made. Who would be so unwise as to claim authorship of any such thing? And what does accountability even mean in this context?

And if the logic of any one algorithm is indecipherable, try to imagine how hard it would be to reconstruct the logic behind a given decision when multiple algorithms mesh with one another to produce an outcome, the entire interaction unsurveilled by any human eye. In our tightly coupled, hyperlinked economy, there are any number of circumstances where our fortunes are shaped by such complex multiway interactions. It might not be possible for anyone to determine afterward, even in principle, whether a decision resulted from any particular process of reasoning, or was simply produced by a poorly buffered algorithm interacting with other automated systems in unforeseen, nonlinear ways.

The idea that we can somehow force these black boxes open, then, and demand that they render up their secrets in the name of accountability, simply isn't tenable. While Pasquale's call for a move "toward an intelligible society" is entirely welcome, any such thing would require a well-coordinated combination of technical, organizational and regulatory measures. It is not at all clear who would be responsible for articulating those measures, who would have the incentive to undertake them, or how they might be enforced.

The question of incentive naturally prompts some reflection as to just who it is that benefits most from the unfathomable obscurity at the heart of algorithmic systems. And indeed—as Pasquale points out at length in his book, and as we've seen from the examples of Google and Branch—there are certainly circumstances in which some party's interest is advanced by our inability to determine how they arrive at their judgments.

But there's a more distressing possibility, which is that no human party may derive any benefit from it at all. This may simply be the price of invoking systems that operate at higher orders of complexity than any our organic minds can encompass. The kind of opacity we've considered here may therefore simply be the pilot wave of a deeper transition rolling through our societies, as algorithmic decision processes take hold in most spheres of life. In the world we are building, we may well contend with patterns of advantage we cannot discern, allocations of resource that make no obvious sense, arranged in ways (and for reasons) we'll never understand, to advance ends we can only dimly perceive. Even our finely honed cynicism, tuned against centuries' experience of human venality, may not be the surest guide to this set of circumstances.

The black-box quality we see in so many algorithmic systems—the deep obscurity of the methods they use to decide whatever matters that are placed before them—aggravates our ability to make wise choices about them in one final way.

As we saw from the examples of Autopilot and Branch, Snaptrends and PredPol, what often matters most in weighing the degree to which we surrender control to an automated decision-making process isn't so much what a system can actually do, but what we believe it can do. In the absence of better information—guided mostly by the folk beliefs about the capabilities of autonomous systems that completely saturate popular culture, leavened significantly by commercial hype—our estimates of machinic competence can grow to the point that they become dangerous. As Tesla enthusiast Joshua Brown discovered, with fatal consequence, this confusion of desire,

belief and actual capability operates at the individual level. But it also functions at the level of entire societies.

A case in point is, again, automated driving. A few years back, a friend with experience in the trucking industry pointed out some of the many complications that would surely beset any attempt at automating away the human driver. He noted that the challenge wasn't merely guiding a cargo vehicle from one point to another, which is comparatively simple, but somehow accounting for everything else that needs to happen in and around that vehicle in order to accomplish the real goal: moving *freight* from one point to another.

As he explained it, "autonomous trucking" really means automating a whole bundle of processes and procedures dedicated to cargo handling, including specialized protocols for the management of live loads or hazardous materials; it means automating the balancing of loads in a moving vehicle, and (at least until route-optimizing algorithms dispense with the necessity of doing so) the swapping of loads between them; and it almost certainly means at least some redesign of loading bays and docks around the world, to accommodate whatever ancillary automation is required by all of this. And reasonably enough, given the magnitude of the effort involved in all this, he concluded from this that automated trucking is some ways off yet.[76]

But all of that doesn't mean that every aspect of the challenge he sketched out won't be essayed, and attacked, and worn down by attrition, however complicated it might have seemed from the outset. Once the conceptually central element of the problem—vehicle control, guidance and navigation—is accomplished, every other subtask wrapped up in logistics suddenly seems like an eminently reasonable and achievable goal, *even if each of them is in itself far more complex than the task of moving the vehicle across a continent.*

Belief, in other words, exerts a peculiar kind of gravitation, pulling history toward it—especially when the belief concerns something as widely desired, for as many reasons, as autonomous trucking. When desire is that overdetermined, the problem sufficiently modular or reducible, and the (hardware and

software) components that might be assembled in a solution already in existence in some context, however remote, the resolution of a challenge like this can come to seem very close at hand indeed. It would be absurd to think that that isn't already affecting investments, hiring, training and other allocations of resource, and that it isn't already reflected, however subtly, in the posture and disposition of all the institutions touched on by trucking.

What comes to be the object of belief, in short, resculpts the space of possibilities we're presented with. The conviction that autonomous operation isn't merely possible in principle, but actually imminently practicable, operates at multiple levels, and creates multiple kinds of consequences. I think it's by now reasonably well understood that the truly vexatious complications of automation are almost never technical but legal, regulatory, institutional, and those invariably take longer to settle out than any mere matter of invention and development.

In the meantime, just as was the case with Tesla's Autopilot, a chasm will open up between belief and realization, and we should understand that this is a "meantime" that might span anywhere from months to decades. And what we will contend with in the interim is an impoverished universe of possibilities.

Consider the set of arguments put forth by Florida state senator Jeff Brandes. In his successful 2014 attempt to eliminate subsidies for mass transit in his district, Brandes argued that it was futile to invest in mass transit when an age of autonomous vehicles was dawning upon us: "It's like they're designing the Pony Express in the world of the telegraph."[77] Never mind that Pony Express riders historically delivered mail, packages and other things the telegraph could not have; the argument from technological inevitability is a vivid and compelling one, especially for Americans nurtured practically from birth on the belief in a gleaming technological future. If autonomous cars really are just a year or two away, why invest in modes of public transit that would surely be rendered obsolete before they even entered service?

This sentiment carried the day, and the light-rail line Brandes

opposed was never built. But Pinellas County, where Brandes prevailed, is a place that desperately needs mass transit. As David Morris reports in *Fortune*, the city and its surrounding region "are consistently near the bottom in a number of transportation and livability indexes. They suffer high average commute times, astronomical pedestrian fatality rates, and massive per-capita spending on the private automobiles that, given today's inadequate public transit system, even the very poorest need to get by." And this will remain true for all the time between the present and any appearance of an automated mobility system capable of serving their needs.

Again, by being politically useful, the mere perception that automation is imminent has produced a new set of facts on the ground. Here the imaginary folds back against the actual, constraining the choices we have in the here and now, forcing us to redesign our lives around something that may never come into being. The lesson for all of us is clear: beliefs about the shape of the future can be invoked, leveraged, even weaponized, to drive change in the present. Even in advance of its realization, automation based on machine learning and the algorithmic analysis of data serves some interests and not others, advances some agendas and not others.