THE
CRITERION
JOURNAL ON INNOVATION

VOL. 1 2016

From Law to Automation

Thomas A. Smith[*]

Laws are like computer code in that they describe what must be done or not done. Computer code does the same thing, but in a sense more precisely. A law can provide that "no vehicle shall be allowed in the park." A computer controlling the town's robot park rangers can do the same thing, by stopping all vehicles from entering the park. The law might define "vehicles" to exclude ambulances, perambulators, and police cars. The robot rangers can also be programmed to let these non-vehicles pass.

But in a sense, law is not like code. This is because law has human agency as an essential step in the causal chain between its enactment and its enforcement or performance. In a seeming paradox, one must be able to disobey a rule before that rule is a law. With computer code or any automatic process, this is not the case. In fact, a process is automated to the extent human agency is removed from it.

Computers do what they are programmed to do. They don't make mistakes; they are "just machines," at least so far. Computer programmers often make mistakes. The business of writing code is fearsomely complicated, all but assuring there will be bugs to be eliminated from a program, and a program is never final. In the end, though, this mechanical quality makes computers, like any machine, incapable of being moral. They are *non*-moral. As Jonathan Bennet writes, in another context,

> [f]or a creature to be correctly said to have a rule, it is necessary that it should be able to break the rule.
>
> . . . .
>
> Reason-giving explanations require a conception of how things ideally would be, sufficiently independent of how any actual individual's psychological economy operates to serve as the basis for critical assessment of it.

* Professor of Law, University of San Diego. I thank Ted Sichelman and Greg Sidak for valuable suggestions.

> In particular, there must be a potential gap between the ideal and the specific directions in which a given agent's motivations push him.
>
> [T]here is no normativity if you cannot be wrong.[1]

Machines or automatons cannot be wrong in the normative sense. The non-culpable "mistakes" that machines make are made because they were wrongly programmed. If blame is to be assigned, it must be assigned to the humans ultimately responsible for getting a particular task done.

My thesis in this article is that more and more of what is now done by law in moral space by relying on human agency is being pushed into non-moral space as these objects of regulation are subjected to mechanisms that are not social, but purely mechanical. In other words, automation occurs. Automation involves the disappearance of law. If a process becomes purely mechanical, it is no longer a legal process. A legal process has somewhere in it at least one step that involves human agency.

I. Laws and Mechanism

Is there such a thing as a (truly) mechanical or automatic law? Could there really be something like, say, Asimov's Three Laws of Robotics, laws that are followed by robots? The answer, perhaps not surprisingly, is no, at least not if we are talking about human law. Purely mechanical law, like the law of gravity, is "law," but in a different sense from human-made law. Laws that are *really* mechanical—that is, *laws of nature*—cannot be broken because they are, as it were, hard-wired into the stuff of existence. Apples fall and decay. Time passes. $F = ma$. These are laws that cannot be broken. If one is broken, it is not physical law, a law of nature, at least not as it was stated. Human law, on the other hand, can be broken. If it cannot be broken, it is not *human* law; it might be another sort of law, such as a law of nature.

Humans sometimes build things that imitate the operation of laws of nature, using different laws of nature: astrolabes, for instance. An astrolabe is a mechanical device of ingenious construction whose parts show the movements of astronomical bodies. A correspondence exists between parts of the universe and parts of the astrolabe, such that certain relations, such as the relative position of two astronomical bodies, are depicted by the device. There are simplified astrolabes in the form of clocks. Mechanisms can break, but as long as the mechanism is working, this similarity allows an astrolabe

[1] Douglas Lavin, *Practical Reason and the Possibility or Error*, 114 Ethics 424, 424–45 (2004) (first quoting Jonathan Bennett, Rationality: An Essay Toward Analysis 17 (Hackett 1989), then quoting John McDowell, *Might There Be External Reasons?*, *in* World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams 68, 76 (J.E.J. Altham & Ross Harrison eds., Cambridge Univ. Press 1999), and then quoting Christine M. Korsgaard, The Sources of Normativity 164 (Cambridge Univ. Press 1996)).

to imitate the motion of the planetary spheres. The mechanical laws that govern our devices are unbreakable in the same sense as the mechanical laws that govern the movement of bodies in the universe, though they may be different laws. A law of mechanics, such as, say, leverage, might in effect substitute for a different law of nature, such as gravitation.

Human laws are different. Human laws are part of the technology (in the broadest sense) that constitutes "social mechanisms of control." Not all laws are social mechanisms of control, but some certainly are. Sociological mechanisms consist in part of people. These mechanisms can arise spontaneously or be the deliberate creation of a legislator. Just as a hundred people lifting, pulling, and prying simultaneously can move a stone block into a pyramid, and thus are collectively a device for moving blocks, consisting of muscles and brains, so a class of people can serve as bureaucrats, soldiers, or priests performing their particular functions in a larger body. In all cases, these groups of people are subject to laws, either formal or informal. The laws obviously are not *all* that it takes to form a sociological mechanism, but they clearly are part of what is necessary.

Among the devices that these legal-sociological mechanisms use are incentives—often penalties: incarcerations, imposition of fines, and stigmas that are imposed on persons who do not adequately fill the roles which they take up or to which they are assigned. To an extent, the legal process of imposing these punishments when necessary aspires to the condition of being truly mechanical.[2] The laws create incentives, and these incentives must exist if the sociological mechanism in question is to operate properly. If a bureaucracy is to function at all, it might require, for example, a rule that says its functionaries must appear at work by a certain time. If a worker does not show up on time, his pay will be docked. This creates an incentive for workers to show up on time. Of course, in real life the internal and external laws of employment in California, for example, are much more complicated than this. But this example illustrates the function of rules in forming a group of people into a sociological mechanism.

A peculiar property of human law consists in how it goes about the formation of sociological mechanisms. Human law uses a variety of incentives to get people to behave in certain ways. These include rewards for actions or inactions, and moral stigmas, fines, incarceration and even capital punishment for acting or not acting in certain ways.[3] Persons who are subject to these incentives are free, in a certain sense, to do or not do what the law requires. Yet we can easily imagine other mechanisms to force people to do or not do certain things. A security fence, for example, prevents people from

[2] *See, e.g.*, U.S. Sentencing Guidelines Manual (U.S. Sentencing Comm'n 2004).

[3] In private law, actions are categorized and interpreted so that they consist, as a matter of law, of contracts or torts, for example.

entering a certain area. A law might also prohibit entry into an area and be enforced by a penalty. Both serve the same purpose—prevent people from entering an area. The fence, however, does not appeal to, invoke, or involve a person's free will to decide to enter or not to enter the area. It stops potential trespassers because it is tall and solid. A law, on the other hand, leaves it to a person to obey or not to obey it. It presents a threat: if you enter this area without authorization, you will be punished. Human agency occupies in law a pivotal place in the chain of possible causal events that make up the social control mechanism in question. Absent this step, whatever we are observing is not a law; it might be just a mechanism. A fence is a mechanical barrier to entry, not a law. A law relies on human agency to get the same result. The removal of human agency from the chain of causes transforms a legal regime into a merely mechanical one. Automation is the replacement by machines of steps previously done by human agents.

II. From Law to Mechanism

As a regulated activity becomes more automated, it tends to separate into two parts, one that involves human agency, and so includes politics and morality, and another that is more purely mechanical. It is as if law and morality-and-politics are emulsified into a nearly uniform system. Then automation catalyzes a reaction that precipitates out a mechanical system on the one hand and, on the other hand, morality and politics. As this happens, the political and moral part may continue to be described by commentators in partly legal terms, while the mechanical part tends to disappear beneath their notice. This mechanical part is considered mere implementation or engineering.

As societies become more technologically advanced, various sociological mechanisms that contained groups of people who acted as part of those mechanisms are gradually transformed into mechanisms that do not rely on human agents, at least not nearly to the same extent. For example, a large insurance company or bank might have employed thousands of clerical workers who performed calculations. Their employment relations were governed both by private rules (which are essentially private statutes and form part of their employment contracts) and public laws. As technology advances, the functions they performed are automated. The same functions get performed, only now machines do the calculations that were previously done by hand. As the step of human agency is dropped from the calculation process, law no longer has a role to play in inducing the behavior that made up the sociological mechanism of the job of being a clerk-calculator. The law might still be fine, sitting silently in its volumes, but the clerk-calculators it regulated are no more.

Law includes a moral or political element in its incentive structure. Morality and politics surround the incentives of positive law like a kind of aura. People are motivated by a desire to avoid punishment, but inevitably breaking or keeping the law also contains a moral element. Whether this moral or political dimension is philosophically strictly necessary or not, this feature seems to be a part of every legal system. But once a function has been entirely mechanized, it no longer needs this moral element. I may feel a moral obligation to get to work on time as well as have a desire not to have my pay docked. But once the function I perform is embodied in a machine, the moral obligation that at least contributed to my getting to work on time is no longer relevant or necessary. Some mechanical device, such as a timer, has replaced the moral incentive, and my job is now done by a machine. Alternatively, consider an important job—a soldier in a nation's armed forces. Military law typically requires soldiers to face the enemy, braving threats to one's own life, and carrying out orders. As a matter of psychology, following military orders can require great courage. But fighting functions are increasingly being turned over to autonomous vehicles and robots. It is easy to imagine that robots will play the chief role in actual combat within a few decades. When this hand-off is complete, the moral and legal infrastructure associated with individual manned combat will have become obsolete, at least within advanced technological cultures. Human agency will still be involved in war, but at a higher level.

When the objects of law are committed to automated process, law changes. Imagine I am driving down the freeway when I am cut off by someone talking on his phone. Furious, I notice that the man pulls into a rest stop. I follow. He walks into the restroom, and I follow him and punch him in the nose. In this hypothetical, I commit a morally and legally culpable act—a crime. Now consider an alternative. I might not have been driving at all, but sitting in an autonomously driven vehicle. When the other driver cut me off, I was too absorbed in my book to notice. The first hypothetical tees up a decision by me that we recognize as moral and legal: whether to punch someone and break the law. In the latter hypothetical, the opportunity to make this decision never arose because of automation. In the autonomous vehicle hypothetical, technical means keep my behavior within a range that excludes the space where my behavior is governed by rules which I can elect to follow or not follow. It excludes human agency. In this moral space, my behavior is governed, at least in part, by myself. There laws establish incentives for morally correct and/or lawful behavior. These moral and legal rules satisfy Bennett's criterion of its being possible for an individual "creature" confronted by them to make the wrong decision with respect to them. As Bennett explains, a creature has to be able to do the wrong thing in order to do the right thing. Otherwise, he is just acting mechanically.

An activity governed by law becomes non-moral when it crosses over the line from being an activity governed at least in part by rules with incentives that tend to cause people to behave in a certain way, to being a mechanical or technological activity that does not require any human agency—that is, in either case, when a programming rule does not leave room for humans to decide whether to obey the law or not. Thus a law might impose vehicular speed limits which must (within reason) be obeyed in order to prevent accidents. Autonomous vehicles which are programmed to observe speed limits, however, would achieve the same result as speeding laws without our choosing to obey the speed limit. An optical scanner might read the speed limit from a sign and then instruct the engine to decelerate or accelerate. In this case, the matter of whether to obey speed limits would have been taken out of human hands by a computer program.

We can imagine programmers sitting down with a copy of the motor vehicle code to make sure autonomous vehicles observed all the rules of driving. (In fact, this process would be much more complicated than this, but we can simplify for illustration.) The programmers who keep autonomous vehicles driving within the speed limits are not making a moral or political decision like drivers make when they decide whether to obey a driving law. The programmers are making a moral and political decision at a higher level, perhaps, such as whether all cars made by their employer should obey the speeding laws. These decisions are distinguishable from the actions the program they write will cause to happen in the future. They are not the same as the moral or political decisions made by the people whose behavior is shaped by the rules that the programmers implement. Actions that used to be taken by people and governed by law are now taken by machines and governed by code. People write the code, but when they apply law through their code they are acting at a higher level than what we are familiar with now. The programmers may indeed be making quasi-legislative decisions in deciding what sort of code to write. They are taking the moral and political decisions we currently make in driving out of the hands of the people who have them now.

III. A Dynamic Historical Process

It might seem that law will be replaced by some sort of anti-nomial regime, with morality and politics on the one hand, and mechanism on the other, with law gradually disappearing. But this is probably not the case. The process of laws disappearing (often by falling into desuetude) will probably be more than made up for by the emergence of new areas of legislation and adjudication. So, for example, it used to be that people got around using horses, and they also employed numerous people to service their horses, such as drivers

and grooms. These animals and human workers were rendered economically obsolete by the proliferation of the automobile. Similarly, passenger trains used to be the most important way that people engaged in long-distance travel in the continental United States, but it is of limited importance with the invention of aviation.

There was a great deal of law surrounding horses and trains as transportation activities, regarding such things as liability for accidents, regulation of prices, and so on. But these laws faded in importance as new technologies rendered horses and trains of much more limited importance. As the work of horses became the work of machines, mechanism replaced activities that were formerly the combined efforts of people and animals. Where before one would hire a horse and driver to take one to another town, a relationship which would be governed in part by applicable laws, one would now simply get in one's car and drive. New laws, however, came to govern the new activity of driving as it became more prevalent. In formulating these new laws, legislatures, it is true, had general principles from past laws to guide their formulation of new principles, but often radically new principles had to be formulated. Thus, while automation drove the desuetude of old laws, decomposing their subject matter into mechanism on the one hand, and morality and politics on the other, it also occasioned the necessity of new laws, which legislators and judges crystallized out of political and moral principles, old and new, as well as out of the physical imperatives of the new technology.

As human culture becomes more complex, so do its legal regimes. Old laws fall into disuse and disappear from memory and new laws emerge, evidently in ever greater numbers. Thus, technology does not result in the disappearance of law as such, and no utopia above or beyond law emerges. The disappearance of some laws is made up for (and more) by the emergence of new laws. One could say romantically that "law is ever renewing itself," but in fact this is just the natural consequence of the work of legislators and judges, striving to keep up with the new forms of life occasioned by new technology.

IV. Automation

It is a commonplace that we live in an age of automation. More and more of our everyday tasks and amusements involve sophisticated mechanical processes. By "mechanical," I mean anything that is done by machine, including general-purpose computers. Watching movies on Netflix or YouTube, reading email or the web on our smartphones, texting our friends, and dozens of other daily activities involve machines and networks of dazzling complexity. "Automatic" is commonly defined as "working by itself with little or no direct human control" as in "an automatic kettle that switches itself

off when it boils."[4] An automatic tea kettle does not decide to turn itself off. According to the laws of physics, it just does what it was designed to do. It is this quality of a machine having little or no direct human control that makes it automatic.

Automaticity is also an aspect of human laws. It is a colloquial mode of speech to say, "if you don't pay your taxes, you automatically get punished." The laws (roughly) are designed to fine you or put you in jail if you do not pay your taxes. Various parties are duty-bound to arrest you, prosecute you, defend you, or incarcerate you. Lawyers know it is not so simple, but lawyers understand the quality that is being referred to. It is an aspect of the "rule of law."

Law, like other human activities, is subject to the economic forces that bring mechanization to all areas of life. Legal research is now largely a matter of computerized searches through textual databases. Legal discovery involves computerized searches through email and other electronic document databases. Some contracts now take the form of "smart contracts," which are computerized agreements, frequently between financial intermediaries, to buy and sell products that only ever exist in electronic form. "Smart property" consists of registering assets on computer networks, proposed to be decentralized and much more accurate than current centralized and paper-based registries. There are many other examples.

As computer technology becomes better and cheaper, government, already significantly computerized, will become more computerized. Taxation, regulation, and ordinary law enforcement are likely to involve interacting more with machines and less with human agents. Within government agencies, computer processes will become more central to official duties. The Social Security Administration and other bureaucracies use databases to keep track of people and their obligations to them and will do more so in the future. The Department of Homeland Security, the Department of Defense, the Central Intelligence Agency, and the National Security Administration all monitor the activities and communications of many citizens for security purposes through methods involving computers and other high technology. There are many other examples. Virtually every activity of government involves computers, and all are likely to involve more in the future.

Those activities and many more are the result of automation coming to government. This process of the mechanization of government is gradual and pervasive, takes place simultaneously on different levels of government and, but for exceptional cases such as those involving privacy and surveillance, is not especially controversial. Most of the processes of government in our regulated society take place beneath widespread public notice. Vehicle

[4] Oxford Dictionary of English 108 (Angus Stevenson ed., Oxford Univ. Press 3d rev. ed. 2010).

registration, health insurance regulation, transportation and traffic laws, food and drug regulation, restaurant regulation, workplace health and safety regulation, anti-discrimination regulation, telecommunications regulation, and of course ubiquitous taxation—the list could go on and on—involve ornate and recondite rules, some of which even specialists cannot ever entirely master. Although some laypersons might imagine that law books encapsulate all the rules for these activities, in fact the rules are found partly in statutes and broader principles of case law, partly in regulations and other guidance issued by federal, state, and local administrative agencies, but finally in the heads and practices of the people who actually administer and those who are governed by regulations in any given jurisdiction. The regulated contribute importantly to regulatory processes by studying the rules and other guidance produced by the relevant agencies, and educating their regulators as to what will work for them and what will not.

Automation comes relatively slowly to government compared to the private sector because the public sector is less disciplined by markets. Public administrative agencies usually do not need to compete with rivals to survive, at least not so much as firms do. Slowly, governments adapt to the larger societies in which they function, often by being pushed in new directions by firms and other constituencies. As technology penetrates societies more fully, it will also penetrate government.

While automation processes in law and government are slow, especially compared to the private sector, eventually most government functions that involve interaction with the general public probably will be given over largely to computerized processes. Paying one's taxes, registering one's car, filing for and collecting Social Security and public assistance benefits, registering firearms, and so on, will probably involve only minimal human interaction. Eventually even transactions that are higher up the food chain of economic activity, such as reviewing the documents of corporate merger transactions or reviewing transactions for antitrust law compliance, will probably be largely a matter of automatic, computerized review.

Many legal rules look at a glance like sufficiently advanced machines could apply them. Other more complicated legal determinations are more problematic because the algorithms computers would apply to them would be "black boxes" that a human reviewer would not be able to reproduce mentally. "Dark code," pieces of obscure code in a program which programmers are reluctant to eliminate lest it turn out to serve an important purpose, also presents a puzzle for reviewers. Yet human decisions, for their part, also contain much that is opaque and ultimately unreviewable. If a human reviewer applies a balancing test to a contested decision, what actually goes on in the mind of the reviewer is often unreviewable. It is not as if a federal appellate court review of an agency decision comes from literally putting

the decision on a scale. Human decision making is a morass of many factors, including many that are frankly subjective. Computer algorithms, one might argue, are no more or less obscure, and should not be held to a higher standard than human decisions.

When law is embedded in technological systems (and so ceases to be law), it becomes less accessible to ordinary people. If law is embedded in code rather than written down in law books and known by legal practitioners, it becomes more obscure and recondite than it already is. Often laws are not written down, but kept in the heads of people who understand how to administer or be regulated by them. On the other hand, law embodied in code is in a certain sense fixed. While law can be manipulated to contain corrupt and hidden purposes, this would be more difficult to accomplish with machines, which always must have explicit, written instructions, than with people. This is an important feature of the transition from rule by law to rule by code. Computer code constitutes a public record of what is to be done and what has been done. Computers can recreate every step of a decision. Auditing what computers have done should in theory be easier than tracking the steps of human functionaries who may or may not record their actions accurately and may subject them to revisionist historiography. In some ways, government computer code should be open to public examination and comment. Citizens or their representatives, for example, should be able to tell exactly how their taxes travel from their pockets to their ultimate destination, probably by using computer applications of their own.

Understanding exactly what is going on with an embedded "law," such as a speed limit program, will require the help of people who know how code works. A simple rule, such as "no vehicles are allowed in the park," would require many lines of code to implement. How is a vehicle to be recognized? How is a perambulator identified? How does a robot ranger position itself to stop a vehicle? What if somebody resists? These are examples of immensely complicated practical problems the likes of which engineers are only beginning to tackle. This "law," if printed out, would consist of many volumes of, to a lay person, practically indecipherable code. This code would specify in detail exactly what the robot rangers could and could not do and under what circumstances they would do or not do it. More levels of expertise would separate the ordinary park-user, and the ordinary lawyer, from those who implement the laws about the park than is the case in a non-automated world.

So far I have examined how law and computer code are fundamentally different. The former involves human agency but the latter does not, except in its creation.[5] I have discussed how code "detaches" morality and politics from (what had been) law. Then I discuss how this process will make what

[5] I am not counting maintenance and modification as separate instances in which computer code involves human agency.

had been law and regulation more recondite and opaque to ordinary citizens. Next I discuss how computers are penetrating the world of commerce and law.

V. Programming People and Programming Machines

Law can be viewed as a sort of program that is (usually) written down and that people execute. This contrasts with a computer program, which people write down and is executed by machines. But people can be depicted as robots of a kind that are genetically and environmentally programmed to do and not do certain things.[6] Prominent in their programming are instructions to obey the law. Robots (a term I use interchangeably with "machines" and "computers") are also programmed to do and not to do specific things, but their programming is far less elaborate than that of human beings.

If we adopt this view, we get a picture of some beings (us) programming machines (robots). The first set of beings is naturally occurring, while the second set of machines is not. This view might be incompatible with examining persons and their decisions from a moral point of view (I'm not sure), and it raises some of the same issues generally as in philosophical debates about determinism and moral responsibility. These I take, however, to be beyond the scope of this article. Even without considering these issues, some interesting points remain.

This picture of humanity as self-programming automata is not new. In a sense, this is the view taken by James R. Beniger in his remarkable 1989 book *The Control Revolution: Technological and Economic Origins of the Information Society*.[7] Beniger tells the story of the industrial revolution as a "control revolution." Instead of a story of various technologies being developed to meet increasing demand, his story is one of ever increasing supply, the increase

[6] This recalls the brief short story by Terry Bisson, *They're Made Out of Meat*, 13 Omni, no. 7, Apr. 1991, at 42, 42:

> "They're made out of meat."
> "Meat?"
> "Meat. They're made out of meat."
> "Meat?"
> "There's no doubt about it. We picked several from different parts of the planet, took them aboard our recon vessels, probed them all the way through. They're completely meat."
> "That's impossible. What about the radio signals? The messages to the stars."
> "They use the radio waves to talk, but the signals don't come from them. The signals come from machines."
> "So who made the machines? That's who we want to contact."
> "*They* made the machines. That's what I'm trying to tell you. Meat made the machines."

Some people indelicately refer to human beings as "meat robots."

[7] James R. Beniger, The Control Revolution: Technological and Economic Origins of the Information Society (Harvard Univ. Press 1989).

of which in turn demands better control of productive processes. Getting better control of resources for production partly involves taking production out of the market and putting it into firms. This in turn involves reprogramming humans, as when railroad workers were placed at the front of trains and programmed through training and incentives to take control of brakes and switching, a move that greatly increased the productivity of railroads and therefore the whole economy.

Viewed this way, programmable computers are just a different level in the bigger project of control. Humans are programmable "wetware" in the productive project. Programming people and programming machines are different activities, to be sure, but in some respects, they are not as different as one might think. In his thought-provoking novella *Manna*, Marshall Brain tells the story of a software program designed so that fast-food workers do not have to think at all.[8] The program tells them what to do every moment of the workday, step by step, through their headphones: "Jane, when you are through with this customer, please close your register. Then we will clean the women's restroom . . . Place the 'wet floor' warning cone outside the door please . . . Please block the door open with the door stop . . . Please retrieve the bucket and mop from the supply closet,"[9] and so forth, for every aspect of their jobs. Most fast food workers actually like not having to think (or perhaps "plan" would be a better word) because it frees their minds so they can think of other things. Workers who cannot accept this kind of continuous monitoring and instruction are let go. The program includes sophisticated location management and monitoring capabilities. In this sort of arrangement, fast-food workers really are very like robots, directed by the management software. Manna eventually evolves into an artificially intelligent computer program, but it would not have to be very intelligent to control fast-food restaurants. Programming human workers is done through the programming of the computers that manage them.

The story in *Manna* is not as far-fetched as it might seem. Somewhat similar practices are already followed, for example, by human resource (HR) departments in developed world corporations large and small. In the United States, taxing, hiring, firing, giving leave to, and paying employees are all highly regulated by complicated state and federal regulations. Some of this regulation is embedded in HR management software. One HR software vender writes:

> Starting at $10,000, Optimum HRIS offers in-house systems for both the Microsoft Windows & IBM iSeries (AS/400) platforms as well as a cloud-based solution. Our software include Optimum Payroll, HR and Time

[8] Marshall Brain, Manna: Two Views of Humanity's Future (2003), http://marshallbrain.com/manna.htm.

[9] *Id.* ch. 1, http://marshallbrain.com/manna1.htm.

> & Attendance. We specialize in helping companies with 100+ employees easily manage the difficult & cumbersome tasks of HR record-keeping & government reporting. Featuring a fully integrated single database, Optimum HRIS is a total solution for cost-effective HRIS management - learn more![10]

HR software is just one tip of an enormous iceberg that economist Brian Arthur has described as a "second economy."[11] This second economy has grown underneath the physical economy with which we are more familiar:

> Twenty years ago, if you were shipping freight through Rotterdam into the center of Europe, people with clipboards would be registering arrival, checking manifests, filling out paperwork, and telephoning forward destinations to let other people know. Now such shipments go through an RFID [radio-frequency identification] portal where they are scanned, digitally captured, and automatically dispatched. The RFID portal is in conversation digitally with the originating shipper, other depots, other suppliers, and destinations along the route, all keeping track, keeping control, and reconfiguring routing if necessary to optimize things along the way. What used to be done by humans is now executed as a series of conversations among remotely located servers.
>
> . . . .
>
> [A]nother economy—a second economy—of all of these digitized business processes conversing, executing, and triggering further actions is silently forming alongside the physical economy.[12]

These "conversations" that used to be human and are now digital are probably legally significant. Under the Uniform Commercial Code in the United States, or the United Nations Convention on Contracts for the International Sale of Goods (CISG) internationally, these digital colloquies would seem to be part of the offers, counter-offers, acceptances, and rejections that make up contracts for the sale of goods. Yet in many cases, no human is negotiating these contracts on a day-to-day basis, and in the near future, these transactions will probably be almost entirely automatic.[13] Yet case law that addresses these automatic conversations is sparse.

Where does the law go when this happens? It disappears from view, but it does not exactly disappear entirely from the system. The relevant law gets embedded in the programs that engage in the digital colloquies that organize

---

[10] *Optimum HR*, Capterra, http://www.capterra.com/human-resource-software/spotlight/147900/Optimum%20HR/Optimum%20Solutions.

[11] W. Brian Arthur, *The Second Economy*, McKinsey Q., Oct. 2011, at 90, http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/the-second-economy.

[12] *Id.*

[13] On a personal level, many of our purchases of books, music, and household goods are virtually entirely automated because we make them through Amazon.

the sales, shipping, and receipt of goods.[14] Yet once it is embedded, it is not exactly law anymore.

As law gets embedded in technology, lawyers will become less involved with it. Computers are programmed to do what people did before. People followed instructions, but machines follow instructions faster and more accurately. This is the point made vividly by the fictional *Manna* case. Economic processes such as the buying, selling, and shipping of goods are changing to accommodate what machines can and cannot do. Machines are not able to do everything humans can do, but what they can do, they usually do more quickly and efficiently than people. The automation of a process involves changing things so that only or mostly machines are involved in producing the good or service offered. This usually means some human flexibility is often lost in order to seize the economies of scale and scope that machines bring to the productive process. Thus Amazon shipped over 100 billion dollars of goods in 2015, but it does not accept orders by phone the way some brick-and-mortar retailers do.[15] Presumably the costs of accepting telephone orders outweigh the benefits Amazon would get from the additional business.

As computers get better at law, people will inevitably get worse. This is a standard consequence of mechanization. Typically, mechanization results in an increase in worker's skill followed by a sharp decline as the knowledge necessary to do a task by hand is lost. Few of us know how to make butter with a butter churn; we simply buy it at a store. Yet in years gone by, it was a widely known skill. Legal skill and automation seems likely to follow a similar path. Few lawyers will bother to master the intricacies of UCC § 2-207, or the Internal Revenue Code's consolidated return regulations, when a computer can keep track of all that.[16] The overall accuracy of the end product, whether sorting out what warranty covers a good, or the payment of corporate taxes, will increase with automation. At some point, the benefits of increased accuracy will justify investment in automation and the decrease in human skill.

[14] Richard Susskind and Daniel Susskind draw attention to a related phenomenon in *The Future of the Professions*: "[I]nterest is developing in embedding legal requirements into our social and working lives, so that, for example, automatic compliance with health-and-safety regulations can be integrated into the design of buildings that can identify and respond when temperature levels are above some statutory level. In this way, human beings do not need to know the law and make a conscious decision to comply, and consequently, lawyers' direct involvement is not needed." Richard Susskind & Daniel Susskind, The Future of the Professions: How Technology Will Transform the Work of Human Experts 71 (Oxford Univ. Press 2015).

[15] *Annual Financials for Amazon.com Inc.*, MarketWatch, http://www.marketwatch.com/investing/stock/amzn/financials. Of course, you can order from Amazon using your smart phone. I mean you cannot call an 800 number and give your order to a human Amazon operator as you can with some old-fashioned catalogs.

[16] These regulations are notoriously complex.

VI. Will Law Wither Away?

How much of law will be automated? If computers become intelligent, either narrowly or generally, does this mean they will they perform all the functions lawyers have now? In this part, I address these questions.

It seems likely, first, that Artificial Intelligence (AI) will have only narrow capabilities, perhaps for many decades to come. AI scientists contrast narrow AI to general AI. By "*generally* artificially intelligent," AI scientists mean having human-like intelligence—that is, being able to respond intelligently to the myriad of problems and challenges humans routinely face. By "*narrowly* artificially intelligent," AI scientists mean, capable of intelligent behavior in one or just a few specific tasks, such as playing chess. AIs have already shown they are able to play chess, *Jeopardy!*, and *go* better than any human player, and so are narrowly intelligent in those respects. But AIs probably will not be able to navigate many simple human environments and problems for decades to come. Humans have many particular abilities and skills, whereas AIs are likely to have relatively few, at least for the time being. Within a decade or two, however, narrow AIs probably will be developed and applied commercially to tasks that have been exclusively the work of lawyers. Thus the questions, in the long term, of whether lawyers will be automated out of existence and whether law itself will wither away, are only natural to ask.

A. Some Applications to Law

These are difficult questions to answer in part because of their abstractness and generality. One way to approach these questions is to ask, what functions does law serve in society, and what functions does technology serve? If the answers show that ultimately they serve largely the same functions, then it may be that technology will substitute for law, if technology can perform the functions of law at lower cost. For many legal tasks, including some that we now consider to be at the heart of legal practice, this will probably turn out to be the case. For example, legal opinions, briefs, and articles are typically expected to conform to "*Bluebook* style," a complicated citation format developed over time by the leading law reviews. Many regard this format as excessively elaborate and, worse, obsolete, but it will nevertheless probably be with us for years to come. A simple kind of AI can be helpful in making one's citations conform to *Bluebook* style, and later forms of these AIs will eventually probably take over the function of correcting footnotes to conform with this style.

At the opposite end of this spectrum is artificially formulated legal advice. Computers are virtually unlimited in their ability to keep the complicated

steps of an argument in mind (as it were). Advances in machine-learning technology and deep learning will enable computers to become increasingly sophisticated at interpreting natural language. Computers will become better at interpreting spoken commands and following the often elusive arguments of human speakers. This will enable computers to interpret and eventually even make at first simple, and then more complicated, legal arguments. At the same time, how we craft legal arguments will change so that computers are able to follow what is said, and make arguments themselves.

Advances are happening now in routine areas of law, such as eviction notice appeals in housing courts, and in technical areas of law, such as corporate tax law. In highly technical areas, often relatively simple steps of logic and syntax are repeated many times, turning statutes and regulations into complicated webs and nests of references.[17] Computers will be able to keep these sorts of structures straight better than humans can. The application of computer technology to law is practically unlimited.

We can also expect the influence of machines on law to be positive by promoting the prevalence of the rule of law, because computers need *rules* to act, though again, computer-governed processes are not exactly the same as legal processes. Human judges, and people in judge-like roles, do not have to be as rule-bound in their actions. They can deviate from past practices and stated rules to take into account their personal preferences and the "interests of justice" they are regulating. Sometimes, however, this involves or leads to corruption.

B. Theories of the Withering Away of the State and Law

Competition for resources is plainly visible among all life forms,[18] including human life. People must find ways to eat, dress, and get shelter, or else die. The drive to continue their genetic lines also implies sexual competition, and this in turn underpins much of human culture. Evidently, people also have religious drives or inclinations. As part of their efforts to survive and prosper, people form into groups, including families, villages, cities, and states. Law forms as part of this process.

Much more sophisticated versions of this vision of human development can be gotten from contemporary anthropologists and historians. What law does in these pictures is mediate, channel, and formalize what would

[17] For example, Ronald Reagan cited this notorious sentence from section 509(a) of the Internal Revenue Code: "For purposes of paragraph (3), an organization described in paragraph (2) shall be deemed to include an organization described in section 501(c)(4), (5), or (6) which would be described in paragraph (2) if it were an organization described in section 501(c)(3)." 26 U.S.C. § 509(a)(4). This is just one tiny section of an entire Code.

[18] To make an argument that involves law generally, one must inevitably make some broad generalizations about human culture, politics, and economics.

otherwise be a chaotic and probably much less successful process. Law evolves among individuals and groups. Law is fundamentally relational and interstitial. It is this feature of law that explains why it will never be completely automated. I will discuss this at the end of this part.

The idea that law and the state, for better or worse, will be replaced by technology is but a recent version of an old idea. Friedrich Engels used the concept of the state "wither[ing] away" in *Anti-Dühring*:

> The interference of the state power in social relations becomes superfluous in one sphere after another, and then ceases of itself. The government of persons is replaced by the administration of things and the direction of the processes of production. The state is not "abolished," it withers away.[19]

Although Engels used this concept, he did not invent it. As Ben Kafka explains in his erudite article *The Administration of Things: A Genealogy*, the concept goes back at least to Henri de Saint-Simon and Auguste Comte.[20] Comte in turn distilled the concept from a multitude of sources circulating in the days before the French Revolution. Comte argued that the administration of *things* should replace the administration of persons.[21] Comte meant by "thing" not just medium-sized dry goods, but everything Montesquieu meant by the term:

> [T]hey were "things" in the most general sense of *res*, that is, objects but also beings, matters, affairs, events, facts, circumstances, occurrences, deeds, conditions, cases, and so forth. Likewise, these things "governed" in the sense that "many things governed" when Montesquieu wrote that "many things govern men: climate, religion, laws, the maxims of the government, examples of past things, mores, and manners." That is to say, some things governed by constraining possible actions, others by constraining permissible ones.[22]

---

[19] Friedrich Engels, 2 Anti-Dühring 307 (Int'l Publishers 1976) (1878).

[20] Ben Kafka, *The Administration of Things: A Genealogy*, West 86th (May 21, 2012), http://www.west86th.bgc.bard.edu/articles/the-administration-of-things.html. Kafka writes: "This was the context in which Auguste Comte made the argument that has been attributed to Saint-Simon ever since. Given the choice, we should replace the government of persons with the administration of things. He made this argument in the third installment of Saint-Simon's *Cathéchisme des industriels*. The essay was published in 1822 as the *Plan des travaux scientifiques nécessaires pour réorganiser la société* and then again in 1824 as *Système de politique positive* (it would also be known as the *Opuscule fondamentale*). Saint-Simon wanted to take credit for the publication, which Comte had written at his request, but the younger man insisted on having his name attached to it. The result was a complicated printing history and an even more complicated schism between master and disciple that probably explains why subsequent readers were confused about its authorship." *Id.*

[21] *Mill to Comte, November 8, 1841*, *in* The Correspondence of John Stuart Mill and Auguste Comte 35 (Oscar A. Haac trans., Transaction 1995). On the reception history of the text, see Mary Pickering, *Auguste Comte and the Saint-Simonians*, 18 French Hist. Stud. 211 (1993).

[22] Kafka, *supra* note 20 (citation omitted) (quoting Charles de Montesquieu, The Spirit of the Laws (Anne Cohler, Basia Miller & Harold Stone trans. & eds., Cambridge Univ. Press 1989).

Engels half-praised these "stupendously grand thoughts" of the utopian socialists to whom he attributed this line of speculation.[23] Engels was succeeded by Marx, Lenin, and Stalin, who dismissed it, considering the details of administration something mediocre minds could easily master, though their intellectual descendants learned otherwise. Some French and English thinkers, such as Jeremy Bentham and John Stewart Mill, by contrast, took this nascent theory of administration as one of the bulwarks of the rule of law. "Law in its descriptive sense became law in its prescriptive sense,"[24] Kafka writes. The idea of the rule of law, and the idea of the state withering away, are thus connected at their roots, through the idea of law in its prescriptive sense being identical with law in its descriptive sense, with the "science of administration" being ultimately a kind of physics. One might be tempted to take the next logical step and claim that, as technology advances, the "administration of things" will be automated.

My view is different. Unlike physical laws, human laws are written by people. The laws they write are essentially the settlements of negotiations (sometimes quite lopsided) that occur among interested parties, whether those parties are political, economic, social, religious, or ethnic groups or individuals, and frequently some or all of the above. These settlements become embedded in society and are the basis for future development.[25] Once a law is agreed upon, it can be automated, or at least those parts of a legal system that are *really* agreed upon in detail can be automated, if technology is sophisticated enough.

C. Disputed Cases and Automation

Even with written codes, however, and frequently with common law, disputed cases arise among litigants and resolving them requires more than a mechanical application of existing law, which a machine could do. It requires judges to exercise their individual judgment, which inevitably means resorting to moral and political preferences. These preferences are informed, one hopes, by the judge's background in the law. But the point is that the resolution of controversial cases, defined as cases for which existing law does not provide an answer, cannot be derived as a matter of mechanizable logic from existing rules, including precedents.

---

[23] Engels, *supra* note 19, pt. 3 ch. 1, https://www.marxists.org/archive/marx/works/1877/anti-duhring/ch23.htm.

[24] Kafka, *supra* note 20.

[25] This view of law may be considered essentially Lockean. Locke tells an anthropological story about where law comes from. John Locke, Two Treatises of Government 161–62 (Cambridge Univ. Press 1824) ("The law, that was to govern Adam, was the same that was to govern all his posterity, the law of reason. . . . for law, in its true notion, is not so much the limitation, as the direction of a free and intelligent agent to his proper interest."), https://archive.org/stream/twotreatisesgov00lockgoog#page/n166/mode/2up.

Of course, preferences could be programmed into a machine *ex ante*, such as, "always decide a controversial case that is not otherwise determined by rules such that the outcome favors the group in power," or ". . . the group that is economically disadvantaged," or whomever. (It is an interesting question, indeed, what the long-term effect of deliberately biased rules of decision such as these would have on a legal system over time. They might be disadvantageous because they might contribute to the long-term incoherence of the legal system at issue. However, they might also contribute to a desirable complexity in the legal system.) In any event, groups whose interests would be mediated by what we call the judiciary would have to decide *ex ante* what these rules of decision would be, if they were to be implemented by machine. It seems unlikely (but not impossible) that they would agree to choose any rule that systematically biased judicial decisions one way or the other.

1. Rules of Decision in Controversial Cases

The decision about what a mechanical rule of decision would be, when it is not otherwise determined by existing law, is essentially moral and political, not legal. When set in place well before particular cases arise that turn on these rules, these rules could have the legitimacy of law. Rules that decided particular cases, if installed at the last minute, would be considered by neutral observers, and probably rightly so, as cases of parties interested in the upcoming case exerting their influence. If the rules are installed well in advance, they have some of the characteristics of neutral decisions, like those made behind a Rawlsian veil of ignorance. This is because the inherent unpredictability of the future imposes a sort of ignorance on the deciding parties. The process of translating moral and political views and preferences into legal rules is called legislation.

If these legislated rules are clear enough, one can imagine computers "making decisions" based on them. This can be so even if the rules are very complicated, and even if they are too complicated to be figured out by humans. There is a difference between complicated and unclear rules, however. It frequently happens at some point (in an appellate process, for example) that the answer to a legal issue may *not* be found in existing law, or even in all logical deductions from existing law. Of course, computers could in theory be programmed with decision rules to help narrow the otherwise open texture of law. A law-interpreting program could apply rules of statutory and constitutional construction, for example. This might make statutory or constitutional text that seemed hopelessly indeterminate able to address particular cases. One can imagine, for example, an "Originalist"[26]

[26] John O. McGinnis & Michael B. Rappaport, *Originalism and the Good Constitution*, 98 GEO. L.J. 1693 (2010).

Computer of the sort mentioned by law professor John McGinnis that would parse all available documentary evidence from the Founding Era and figure out, if it was possible, what the original meaning of a word or phrase from the 1787 Constitution was. Databases containing all relevant legislative history or all of a large administrative record could similarly be parsed by narrowly intelligent legal AIs and interpreted by their natural-language algorithms to interpret the meaning of statutory or administrative language. One suspects in this world that legislation would consequently grow more sophisticated.

Even with these developments, the pressure of litigation, broadly conceived, will force issues to the surface that cannot be resolved even under the most rigorous and sophisticated application of actually existing law. This open-textured quality of any legal system is ultimately related to a fundamental feature of any axiomatic, logical system. Gödel's Incompleteness Theorem showed this is the case for logical languages generally, and it was quickly seen to apply to Turing machines, of which general-purpose computers are examples. For our purposes, this means however complex, intricate, and recondite our legal system becomes, there will *always* be unsettled issues that computers cannot answer on the basis only of already existing laws, unless they simply randomly guess at the answers.[27]

2. How Law Survives

We can generalize this result and conclude that judges or other deciders will always have legal work to do—although it is not strictly legal work that they will have to do. More precisely, what they will have to do will be interstitial legislation. Thus, as law becomes asymptotically more precisely defined—a process which in the future, I claim, will include becoming more mechanized and automated—these left-over or newly generated questions—that is, the questions that are not answerable within the existing legal system by legal and ultimately mechanical means, whoever or whatever does them—will

[27] The idea that these legal gaps or interstices should be filled randomly should not be dismissed without some thought. It may be that eventually, legal oracles will have grown so sophisticated that humans cannot really follow their legal reasoning, or that only a few humans can hope to do so. The point has been reached already, where some of the contracts drafted by humans are too complicated to be understood by humans. When this point is reached, it is unclear that human beings will be in a position to judge whether gaps, when they appear, should be filled by some humans exerting their best judgment as to what the outcome should be, or whether this job was best left to machines. I suspect this time, if it ever comes, will not arrive for some time. In the meantime, gaps that appear and cannot be filled by sophisticated but ultimately mechanical legal reasoning will probably be of the kind human judges present relatively elementary problems. There is no particular reason to think that human thinking and machine algorithms will find ambiguities or other interpretive issues in the same places. To take a trivial example, a machine might be unable to answer the question whether remote-control cars are permitted in the park, where there is an ordinance that provides that "no vehicles" (except for police cars, baby perambulators, and ambulances) are allowed in the park. A human judge might be able to tell that the remote-control cars are a nuisance and social welfare generally would be promoted by banning them from the park, and so rule, whereas a computer might hang itself up on this question.

become more exclusively moral and political, rather than legal, in nature. This is another way in which law will become detached from morality and politics. It will be as if the automation of law pushes out the morality and politics from law. Of course, new laws will take the place of old ones. It therefore may be that society will continue to seem to be governed by law to the same extent as it does now, perhaps more, or perhaps less.

This does not, of course, mean that legal rules are often without moral and legal consequences. But if rules are well settled, the morality and politics of the issue addressed by a given rule are well settled, at least for public purposes. For example, human cultures decided long ago that deliberate killing of humans for the thrill of it was not to be countenanced. Laws against this sort of killing are of long standing and not open to question in ordinary, public debates. Gladiatorial combat, snuff films, and the like, are simply not permitted. A society could permit them, as the ancient Romans did for gladiatorial combat, but it would be a different society from ours. Other laws, such as those permitting abortion, are still contested, but this is largely because a subset of the population holds that fetuses are persons, and so should be protected from killing. To some extent, one can see this process of law, and politics and morality, separating themselves, at work in the jurisprudence of the U.S. Supreme Court and other appellate courts. While some of their cases are unnecessarily political or moral—that is, the courts have given a politically convenient answer to a legal question which had a different legal answer—some of the courts' decisions are *inevitably* political or moral, in that existing law does not provide an answer to the issue posed to the courts, but the courts must nevertheless answer it.

Thus the dream of the early social utopians—that government will gradually become the administration of things and the administration of persons will wither away—seems to rest on a misconception of the role of law in society, at least if the social utopians are interpreted to mean that all decisions thrown up by a legal system will eventually be answered mechanically within that system. Because society is dynamic, and perhaps even if it were not, new legal issues are raised. Where these are issues that can be solved by a logical extension of existing rules, law can be applied mechanically, whether by humans or machines, and in theory, a machine could do it. But other new legal issues will not be of this sort. The existence of these issues is possible as a matter of logic. If the legal system rests on axioms, then more statements can be derived from those axioms than can be proven by them.

This also means that when these new legal issues arise, the legal system may branch: different solutions to a given legal issue may be found, but both or several are all equally consistent with the law that came before. Whether a legal system should turn, say, left or right, might not be a matter of law, but rather of politics and morality. Unless these decisions are to be left to

machines, where there is no reason to believe they would have any special expertise, humans will have to make these decisions. How they make them or should make them is beyond the scope of this article, but one might wish these decisions be made democratically, for example. In any event, the administration of things will always generate open questions for which these things themselves do not provide the answers. The administration of persons cannot be entirely purged from the government of persons.

VII. Addendum on AI Questions

As this article is about the future of law and automation, it is perhaps natural to address questions of artificial intelligence and law. While "AI and law" is a broad topic that does not lend itself to sweeping generalizations, some issues still do suggest themselves. The issue of "harmful AI" is one, and it is particularly relevant because law is, besides one of the oldest, also one of the most important technologies that we use to make human life secure and, as far as we can manage, productive and happy. So we can ask, are AI and law at cross purposes? At least since the Enlightenment, law broadly conceived has been part of the project of human freedom. Does AI threaten to undermine all that? Does AI threaten ultimately to subject people to a new kind of slavery?[28] These are big issues really beyond the scope of this article, but I can make a few observations.

In the popular press, one sees references to "killer robots," AI technology that perhaps unintentionally presents a deadly threat to humans. Some technologists and scientists, including luminaries Elon Musk, Bill Gates, and Stephen Hawking, have expressed fears that AI might be an existential threat to humanity. Oxford professor Nick Bostrom has written an extremely thoughtful book on this topic.[29]

AI might turn out to be harmful. Harmful AI may be defined more precisely as any AI technology that has a net detrimental effect on the total general welfare of humankind. Harmful AI could be harmful to anybody who encountered it, or just to some people. It might be profitable to some, while it still reduced the net general welfare of humankind. While killer robots would be an example of the first case, AI that put some or all people out of work, without also resulting in a sufficiently broad distribution of wealth, presumably would also be an example of reducing net human welfare.

Experts speak of AIs harming humans because of "value misalignment."[30] This misalignment occurs when the values of AIs do not align with those of

[28] *See* Thomas A. Smith, *Robot Slaves, Robot Masters, and the Agency Costs of Artificial Government*, 1 Criterion J. on Innovation 1 (2016).

[29] Nick Bostrom, Superintelligence: Paths, Dangers, Strategies (Oxford Univ. Press 2016).

[30] *See, e.g.*, Stuart Russell, AI: The Story So Far, https://intelligence.org/files/csrbai/russell-slides.pdf.

humans. Of course, *some* misalignment is, one assumes, desirable and inevitable. The narrow AI programs that are now undoubtedly used by the NSA to ferret out terrorists make these terrorists much worse off, but the rest of us better off, if they work as intended. AIs that never made *anybody* worse off would be paralyzed by their universal benevolence. Harmful AIs usually serve interests, one can assume, whether their own, or those of their ultimate programmers.

Most of the social sciences of decision-making, such as standard game theory, define decision-making as an effort to increase a party's or parties' self-interests. If one has to choose between two options, one will choose the option that contributes more to one's utility or welfare. It might seem fairly straightforward then to give an AI some objective welfare function such as "your welfare will be maximized when you make as many paper clips as possible," to use Nick Bostrom's famous example.

An AI with interests has something to do. But then the problem becomes, what if the AI develops some unexpected capacity? (We often suppose AIs that become intelligent or superintelligent develop unexpected capacities.) If an AI does this, it will presumably use its capacity to try to achieve its objective—in this case, making paper clips. There is no stopping point, however, since "make as many paper clips as possible" does not set any numerical limit. The AI might then realize that by taking over the factory next door, it could make more paper clips. So it takes over that factory, then gradually the whole Earth, and finally the entire universe, all because the programmers inadvertently set off their AI on what in retrospect was a poorly defined mission.

This problem might seem easily solved, but it is not. It cannot be avoided simply by giving the AI a numerical limit to achieve, such as one million paper clips. If the AI uses standard probability reasoning, as it presumably would, it might decide to maximize the *certainty* with which it will produce one million paper clips. To have the actual capacity to produce one million paper clips, being able to produce an unlimited number of paper clips might be best, at least if one wants to produce one million paper clips with a probability of 0.9999999 . . . .

One might then say, we can specify the budget of the AI, such that it can spend no more than $100,000 to produce one million paper clips. But then the AI might just figure out new physical principles that would allow it to produce one million paper clips, again to a nearly absolute certainty, but for which it needs to get ahold of practically infinite resources, somehow at only the cost of $100,000. At any rate, one gets the idea. One might have to build an AI just to keep up with all the ways the paper clip making AI might go wrong, and an AI to watch that AI, and so on. Even if a particular paper clip factory avoids these problems, it does not mean every factory will. This is a general problem with giving an AI a set of objectives to attain or maximize.

One might think that if an AI started converting the world into paper clips, one could just flip the off-switch until one figured out what had gone wrong and how to fix it. But this problem cannot be avoided by making sure every AI has a prominent off-switch, as Steve Omohundro has pointed out.[31] Omohundro argues that an AI will contribute to the ultimate achievement of its goals by first fulfilling the subsidiary goal of not allowing its existence or operation to be terminated. As Stuart Russell puts it, if the objective of an AI is to get the coffee, it must first not let anybody turn it off.[32] Any AI that has an ultimate objective programmed into it must necessarily have the *subsidiary goal* of assuring its own continued existence, at least until its ultimate goal is achieved, which might be after the entire universe has been reduced to paper clips.

One interesting suggestion made by Stuart Russell is that the answer might lie in *not* choosing ultimate goals for AIs, but rather letting them infer goals by "inverse reinforcement"—observing what goals humans, or at least some humans, try to achieve, and then inferring what the rewards would have to be for these to be that agent's ultimate goals.[33] If we let AIs observe people and their goals, the AIs can infer what the rewards are that "program" people to get to their goals, and then set these reward functions and goals as their own. Roughly, then, AIs would learn the way children are said to, by imitating the adults around them. If AIs were educated in this way, perhaps they would not attempt feats people would never attempt, such as turning the world into paper clips. AIs would acquire the nuanced and multi-factorial goals that real humans have, Russell suggests.

The problem with having AIs imitate the goal structures of people, however, is that it gives people too much credit. Our world is as good as it is at least partly because people do *not* have the capacities to achieve their goals, whereas, by hypothesis anyway, AIs might. People have various utility functions, and they try to maximize their utilities as best they can, subject to their budget constraints of wealth, intelligence, and other resources, as well as outside constraints, such as law. But humans do not have the capacities that AIs would have if they became superintelligent or had other super-capacities. People have limitations, not least of all legal limitations, regarding what they are able to do to maximize their utilities. A person is not permitted to

[31] Steve Omohundro, *The Basic AI Drives*, *in* Frontiers in Artificial General Intelligence 2008 (Ben Goertzel, Pei Wang & Stan Franklin eds., 2008), https://selfawaresystems.files.wordpress.com/2008/01/ai_drives_final.pdf.

[32] Stuart Russell, *Should We Fear Supersmart Robots?*, Sci. Am., June 2016, at 58, 58.

[33] *See, e.g.*, Dylan Hadfield-Menell, Anca Dragan, Pieter Abbeel & Stuart Russell, *Cooperative Inverse Reinforcement Learning*, in Advances in Neural Information Processing Systems 25 (MIT Press forthcoming 2017). This is something like the Baxter strategy writ large. Baxter is a robot designed by Rethink Robotics which can be programmed by ordinary workers by having them move Baxter's limbs through the motions they would like it to take, a considerably simpler process than most industrial robots. *Baxter with Intera 3*, Rethink Robotics, http://www.rethinkrobotics.com/baxter/.

kill other people or take their property, for instance, even if doing so would maximize that person's utility. There are also things that it has not occurred to us to outlaw because they do not present any particular, current threat to us because no one can do them. No one has the intelligence or other capabilities necessary to exploit these unknown loopholes in our legal defensive structure. This might not be the case, however, if intelligent, or superintelligent AIs existed.

It is easy to imagine a superintelligent computer modeled on Stuart Russell or some committee of scientists going wrong, if it had the capacity, say, to rule the world. For a less remote example, consider an AI in charge of a private corporation, such as Alphabet, Inc., formerly Google. Suppose the AI observed the behavior of the officers and employees around it while it was being trained. In an ideal case, the AI would infer that its goal was to maximize profits, within the structure of the legal rules and principles that applied to the corporation. We would then have an AI that (by hypothesis) would take advantage of all sorts of legal gaps and loopholes that exist in current law by reason, among other things, of the legislature's not having caught up entirely with current business practices.

There would be other loopholes as well. Taking advantage of legal loopholes is no more, after all, than we expect private businesses to do. The AI might, for example, figure out novel tax strategies that enabled the AI-managed company to pay an unprecedentedly low percentage of its profits in taxes. Apple Corporation has done something like this already with respect to its international tax planning, provoking the European Community to charge it with suborning Ireland in its tax schemes.[34] An AI might pursue a similar strategy even more effectively than human tax planners. If world domination were not possible, other goals undesirable for most people, but good for the AI-managed corporation, might be attainable. The managers of a corporation (human or AI), after all, are trying to maximize the welfare of themselves, their shareholders, or their constituents, but not everybody.

Let us next suppose the human business executives at our AI-managed corporation are less than perfect, a realistic assumption. What is the AI to make of these humans' actual behavior, in terms of the goals and rewards it distills from their actions which it is going to imitate? Here it is difficult to figure out what the AI is supposed to do. Would it apply some sort of implicit hierarchy among all the individual executives in the corporation as to which goals counted more when the employees' ambitions conflicted, as certainly they would? If the CEO wants *X*, but the vice president wants

[34] Natalia Drozdiak, *European Commission Decision Isn't Endgame for Apple, Ireland*, Wall St. J. (Sept. 8, 2016, 2:00 a.m. ET), http://www.wsj.com/articles/european-commission-decision-isnt-endgame-for-apple-ireland-1473314400.

an inconsistent Υ, to which should the AI give priority? The CEO might be more selfish than the vice president, putting her own interests ahead of the corporation's. Yet this does not seem to be the lesson we would want the AI to take from observing human behavior. Corporations are composed of self-interested individuals. If an AI managing a corporation was designed to imitate human beings as its moral compass, it would presumably end up being selfish too. But this presumably is what we want to avoid.

A selfish AI with super-capacities would be a formidable threat to the interests of some, and perhaps all, humans. An AI would learn a survival instinct—this is an essential human behavioral characteristic and probably the most basic. But, if an AI's capabilities are greater than any individual or group of individuals, the AI's survival might imperil important human objectives.

One might try to solve this problem by saying the AI should instead observe the collective behavior of groups of people. So, for example, the AI could observe what some group of officers and employees of the corporation was doing and, speaking colloquially, steer a course among them. But this runs into the long-known problems of trying to derive a general welfare function from several individual welfare functions.

If a super-capable AI is designed to "maximize shareholder profits," as it is usually assumed the goal of a corporation's board of directors should be, many problems are foreseeable. Suppose an AI were programmed to maximize shareholder profits but within the law. Our law, however, is far from perfect. It does not prohibit many bad actions that have hitherto been outside of the capabilities of existing persons, not to mention bad actions that are within their capabilities. But these actions might turn out to be within the capability of the AI. The AI might immediately begin doing these profit-maximizing bad actions, while simultaneously taking steps to keep them from coming to the attention of those who would attempt to stop it, and then resisting those efforts so long as possible. Sometimes human-run corporations do this as well, but an AI with super-capacities would presumably be much more formidable at the job.

Thus it seems that AIs, if they ever become super-capable in the ways we imagine, might be even more difficult to control than humans are. As Hayek observed in *The Road to Serfdom*, humans who rise to the top of government frequently have bad motives.[35] If AIs imitate these humans, the results might be even worse. It seems much of humans' not inflicting harm on one another is due to humans' relative lack of capacity to do so, and they already seem to have a lot of capacity in this regard. As humans have become more capable, so has the destructiveness of our species increased. World War II was the

[35] Friedrich A. Hayek, The Road to Serfdom ch. 10 (Univ. of Chicago Press 1944) ("Why the Worst Get on Top").

most destructive of all wars, for example, in part because of the capacity of the weapons created and then used in it. Some people worry about the ever increasing capacity of humans to use physical resources resulting in environmental degradation, global climate change, and related problems. For this and other reasons, Bostrom and other thinkers are well advised to continue to worry about the prospects of artificial intelligence.

VIII. Conclusions

Law is that portion of applied political and moral theory that can in theory at least be automated or mechanized. Law organizes people and their actions into sets of people and classes of action. Yet as law is automated, it stops being law. This is because law consists of the rules that we use to determine or influence the behavior of humans through incentives on human agency. A great deal of what law now does can in principle be done by technology and increasingly will be.

This does not necessarily mean that law will do less and less, but it will do different things. There will always be open issues in law. Law is fundamentally open-textured. New technology and new ways of life will push up issues for law to resolve, until technology catches up with them.

When something becomes automated, it tends to cease to be the subject of debate and controversy, but is rather taken for granted as part of the backdrop of our lives. This does not mean controversies are correctly settled, just that they are settled. As technology becomes more advanced, more of what law currently does will be subject to automation. When this happens, for better or worse, this law tends not to be debated any more. One can easily imagine an advanced culture that had embedded technological "laws" that were unjust. Political and moral debate will nevertheless tend to move its focus to questions the answers to which are not really subject to automation, at least not yet. To take a nontrivial example, much of the abortion debates that continue to roil our society stems from the primitiveness of birth-control technology, relative to what one imagines it might be in the future. If a woman or a man could simply decide whether or not they wished to conceive, so that the termination of a pregnancy was not even at issue, all with no more trouble than an act of will, there would not be much left to debate. The formerly legal, moral, and political decisions will have been made subject to technology, and debate about them would simply no longer occur. One could say that because of automation, law would have become detached from morality and politics. There would be the mechanism on the one hand (whatever technology that made birth control so easy) and morality and politics on the other (whether this technology was ultimately good or bad). But of law there would be little, it having been rendered irrelevant.

Cultures advance technologically and law keeps up as part of a dynamic process. Laws are ultimately practical principles, and technology makes some things practical and other things not so practical. The law of domestic livestock causing torts, as when John's bull tramples Henry's crop, is no longer very important in our advanced culture, but the law of aerial drones might come to be. Cattle and hogs, meanwhile, live such meticulously controlled lives that they are now unlikely to cause much damage by their wanderings. Law seems to stand in place, but it is constantly changing, "renewing itself," as courts and legislators render our old laws and political and moral principles into new.