



**UNIVERSITY OF
TORONTO PRESS**

WARMING UP TO INSCRUTABILITY

Author(s): Brian Sheppard

Source: *The University of Toronto Law Journal*, 2018, Vol. 68, Supplement 1: Artificial Intelligence, Technology, and the Law (2018), pp. 36-62

Published by: University of Toronto Press

Stable URL: <https://www.jstor.org/stable/10.2307/90019652>

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/10.2307/90019652?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

University of Toronto Press is collaborating with JSTOR to digitize, preserve and extend access to *The University of Toronto Law Journal*

Brian Sheppard*

WARMING UP TO INSCRUTABILITY:
HOW TECHNOLOGY COULD CHALLENGE
OUR CONCEPT OF LAW†

In this article, I discuss how technological development could change the way that we think about the essential features of legality. In particular, I focus on the strengths and weaknesses of machine learning in the context of legislation and adjudication. I argue that the content of those essential features could depend upon our willingness to make tradeoffs between intelligibility and results. These tradeoffs might lead us to reject a concept that requires critical officials (HLA Hart), reason-based tests of legitimacy (Joseph Raz), or deep justifications for coercion (Ronald Dworkin). I conclude that our concept of law will likely be shaped by our willingness to accept a growing disconnect between the way that we decide and the way that the system does.

Keywords: HLA Hart, Joseph Raz, law and technology, legal philosophy, machine learning, micro-directives, Ronald Dworkin, rules versus standards

I Introduction

I have been asked to give an answer to the question ‘will technology challenge the conceptual foundation of the law?’ For a long time, the standard answer would likely have been ‘no,’ on the ground that true concepts are universal – they are true in all places and in all times – so the development of technology or anything else could never unseat them.¹ Today, however, many legal philosophers have abandoned the notion of a non-contingent concept of law, emphasizing that law is, like technology, a

* Professor, Seton Hall University School of Law, Newark, United States

† I would like to thank Simon Stern, Mireille Hildebrandt, Frank Pasquale, Daniel Martin Katz, Jon Romberg, Michael Risinger, Edward Hartnett, Harry Surden, Paul Gowder, Anthony Casey, Anthony Niblett, and two anonymous reviewers for helpful conversations, suggestions, or criticisms. I am especially grateful to the *University of Toronto Law Journal* for the opportunity to write on this topic.

1 See e.g. Immanuel Kant, *A Critique of Pure Reason*, trans Werner S Pluhar (Indianapolis: Hackett Publishing, 1996); Warren Goldfarb, ‘On Quine’ (paper delivered at Centennial Celebration and Market Dedication honouring Willard Van Orman Quine, Oberlin College, 28 June 2008), online: Douglas B Quine <<http://www.wvquine.org/GoldfarbOberlin.pdf>> (describing philosophical enterprise after Immanuel Kant until the twentieth century).

human artefact.² As Brian Leiter recently explained, '[h]uman artifacts [like law] answer to human interests, thus their nature and character is hostage to changing needs and wants. Even so, we can try to take a conceptual snapshot of these artifacts that answers to our current concerns.'³ The strengths of future technology might increase our expectations regarding the goods that a legal system can deliver, but they might also increase our tolerance for systemic weaknesses that arise from that technology. These changes in attitude could be so profound that the theoretical normative systems we would now find objectionable and, importantly, non-legal become so desirable that we revise our concept of law to accommodate them.

Conceptual change can be important. Concepts are useful in normative debate. They provide a means to set forth the parameters of empirical inquiry, and they can highlight irksome inconsistencies or hidden parallels in our thinking. Moreover, their usefulness is arguably enhanced by their once-denied contingency: since they are influenced by the times in which they live, they can better respond to the anxieties and hopes of those times. Once we admit contingency, we open the door to a project like this one: we can fruitfully examine the relationship between contemporary concepts and emerging trends to make educated predictions about what our future concepts might look like.

In this article, I will first identify technology that has a significant chance of changing our interests and expectations with respect to the manner in which we are regulated by law-like systems. In doing so, I consider the trajectory of technological development, seeking to identify existing technology that is promising enough to come to bear on the way we generate authoritative norms and resolve disputes. Simply put, this will be an examination of the features of emerging technology that legal philosophers ought to care about. Thereafter, I will investigate whether this development will pressure us to revise our concept of law. Because there is no current consensus on a single concept of law, I will consider three leading approaches, analyzing how our changed attitudes toward the plausibility and desirability of technologically driven systems of governance might lead us to reject them. I close by briefly discussing the sources of tension that might shape our future concept of law.

² For a recent discussion of this issue, see Brian Z Tamanaha, *A Realistic Theory of Law* (Cambridge, UK: Cambridge University Press, 2017) at 38–56.

³ Brian Leiter, 'Legal Positivism About the Artifact Law: A Retrospective Assessment' in L Burazin et al, eds, *Law as Artifact* (Oxford: Oxford University Press, forthcoming).

II *How technology could impact our law-related concerns*

There are myriad ways that technology could change the way that stakeholders in the legal system operate, but my primary concern here will be the way that it might change our perceptions regarding the delegation of official tasks to machines.⁴ The replacement of human officials with automated processes would undoubtedly have a profound impact on our interactions with the legal system as well as our expectations and fears with respect to those interactions. It is, therefore, a suitable starting point.

A ON THE APPEAL OF HIGH-VOLUME, AUTOMATED LEGISLATION AND ADJUDICATION

I will begin with an illustration to show how machine-powered legislation might seem like a desirable way to regulate conduct. One way to demonstrate the manner in which a modern society might warm to the idea of automated rule making over a prominent and commonplace activity is to consider the problem of traffic congestion. Most theorists agree that one of law's core functions is solving or alleviating coordination problems,⁵ and many scientists classify automotive traffic as one such problem.⁶ Lawmakers could conceivably come to view thoroughly automated rule making in a technologically advanced environment as a desirable way to produce both faster travel and lower carbon emissions.

It is well known that automated driving vehicles could provide a means to alleviate traffic congestion.⁷ For one, the vehicles have already gained a foothold in the automotive industry, with some experts offering 'conservative' estimates that 20 million will be on the road by 2020.⁸

4 Because it is so difficult to predict the arc of technology, it is important to state at the outset that the following descriptions of our potential future are not predictions; rather, they are plausible scenarios that serve to pump our intuitions regarding the way that our concept of law might shift.

5 See e.g. Kenneth E Ehrenberg, *The Functions of Law* (Oxford: Oxford University Press, 2016) at 182.

6 See e.g. Bhuvanachithra Chidambaram et al, "Commuters" Mode Choice as a Coordination Problem: A Framed Field Experiment on Traffic Policy in Hyderabad, India" (2014) 65 *Transportation Research* 9 at 9.

7 See e.g. Yuki Noguchi, 'Self-Driving Cars Could Ease Our Commutes, but That'll Take a While,' *National Public Radio* (10 February 2017), online: NPR <<https://www.npr.org/sections/alltechconsidered/2017/02/10/514091049/self-driving-cars-could-ease-our-commutes-but-thatll-take-a-while>>.

8 Olivier Garret, '10 Million Self-Driving Cars Will Hit the Road by 2020: Here's How To Profit,' *Forbes* (3 March 2017), online: Forbes Media <<https://www.forbes.com/sites/oliviergarret/2017/03/03/10-million-self-driving-cars-will-hit-the-road-by-2020-heres-how-to-profit/#22bbb9ef7e50>>.

Widespread vehicle automation could drastically limit congestion by lowering the probability of conduct that causes backups, such as risky driving and inefficient route taking.⁹ It is by no means obvious, however, that the mere introduction of these vehicles into the stream of commerce will solve, or even reduce, congestion or emissions.¹⁰ The most positive forecasts regarding congestion reduction assume the networking of a high proportion of vehicles on the road with each other and widespread vehicle sharing.¹¹ Neither of these conditions are easy to satisfy. Even assuming that competitor companies would be willing to share driver information with each other¹² and that perfect strangers would embrace carpooling en masse, the congestion problem could persist. Initial benefits under the system, such as the ability to perform other work while travelling, could increase demand for vehicle travel with inefficient, high-speed engines, thereby bringing about a higher number of travellers and nullifying salutary effects on congestion or emissions.¹³

Thus, human choices under individualized circumstances could have a strong bearing on the utility of autonomous vehicles. It would not be implausible for a legal authority to harness its unique coercive powers to constrain individual decision making and bring about the coordination necessary to solve the problem. Indeed, heavy regulation has been contemplated in the models for pollution reduction by leading experts,¹⁴ and governments have already enacted legislation coordinating the testing, rollout, and operation of autonomous vehicles. For example, Nevada, an early mover in autonomous vehicle legalization, recently passed legislation

9 Zia Wadud et al, 'Help or Hindrance? The Travel, Energy and Carbon Impacts of Highly Automated Vehicles' (2014) 86 *Transportation Research* 1 at 1–18 [Wadud et al, 'Help or Hindrance?'].

10 Ibid.

11 Jeffery B Greenblatt & Samveg Saxena, 'Autonomous Taxis Could Greatly Reduce Green House Gas Emissions of US Light-Duty Vehicles,' *Nature: Climate* (6 July 2015) (making assumptions of networked fleet and ride sharing in offering estimates of up to 90 per cent reduction); Will Knight, 'A Simple Wireless Technology Promises to Make Driving Much Safer,' *MIT Technology Review* (April/May 2015), online: MIT Technology Review <<http://www.technologyreview.com/s/534981/car-to-car-communication>>.

12 Perhaps this is not such a fanciful notion. Uber recently agreed to provide anonymized driver data both as a gesture of good will and because it believes doing so might spur infrastructure improvements in cities. Darrell Etherington, 'Uber Debuts Movement, A New Website Offering Access to Its Traffic Data,' *TechCrunch* (8 January 2017), online: Oath Tech Network <<https://techcrunch.com/2017/01/08/uber-debuts-movement-a-new-website-offering-access-to-its-traffic-data/>>.

13 Wadud et al, 'Help or Hindrance?' supra note 9.

14 Ibid.

pertaining to the use of software for platooning, driving distance, and permits for autonomous vehicle networking companies.¹⁵

While these laws or others like them might go some distance toward solving the problem of traffic congestion, it is easy to imagine a considerably more aggressive regulatory approach, one that stands a better chance of realizing the most optimistic models for congestion eradication. Highly individualized laws or regulations, each one customized to coordinate the usage, route, and speed of each separate traveller on a second-to-second basis, could maximize the good that could come from automated and networked vehicles. By contrast, familiar legal norms of general application, like speed limits, would not be detailed or context sensitive enough on their own to provide adequate guidance about routes, lanes, and precise speeds. And even if a state were willing to pursue this aggressive, individualized legislative approach, the sheer quantity of laws or regulations would make it too costly or impractical for human legislatures or regulatory bodies to complete the task. Such a system would require an unfathomable number of individualized directives per day, perhaps billions.¹⁶

This is where technology-assisted law comes in. There might someday be a machine that provides up-to-the-second individualized directives to each and every traveller and their vehicles, indicating the vehicles that people may take, the direction that they may go, and the speed at which they must travel. The same system could take control of the vehicles or even support the automated adjudication of violations, using data records from the vehicles involved and environmental details to determine whether the conduct fell below a rule-based threshold, such as deviating from the mandated path or exceeding the speed limit. Existing technology is already capable of making such directives within networks, for example, and it is certainly not unrealistic to assume that it could function in a complex travel environment.¹⁷

¹⁵ Nevada Assembly Bill no 69 (2017).

¹⁶ To be sure, governments are not unwilling to enact rather detailed laws on minutiae. The sheer quantity of regulations governing health care has become a political hot button, for example. Candice Malcolm, 'The Pitfalls of Single-Payer Health Care: Canada's Cautionary Tale,' *National Review* (13 April 2017), online: National Review <<http://www.nationalreview.com/article/446689/canadas-single-payer-health-care-system-cautionary-tale>>: '[I]n Canada health care is technically delivered privately, although given the Kafkaesque regulations and restrictions that govern it, the system is by no means market-based. In fact, Canada's government-controlled health-care system has become more restrictive than communist China's.' What I imagine here, however, is on a scale never before attempted, as it covers millions of people, each one subject to potentially thousands and thousands of separate, individualized directives per day.

¹⁷ I thank an anonymous reviewer for the example of Internet traffic. Internet protocols rely on dynamic routing, which rapidly changes data pathways based on the discovery

If a system of individualized rules still seems implausible, it cannot be from the unwillingness of citizens to submit the necessary data to such a system or to accept and even welcome its commands. We are already willing to wear or carry devices that provide great detail about our circumstances to databases. Our cellphones are already capable of providing real-time spatial location data.¹⁸ Similarly, Fitbit will soon add glucose monitoring to its products, which currently track steps, sleeplessness, heart rate, distance, and so on.¹⁹ And we have already have embraced highly contextualized and automated directives in the travel context, eagerly (and sometimes blindly) accepting directions from Waze or Google Maps.²⁰

Legislators, too, might grow to accept or even favour such a system. Specific directives, like rules, are attractive to their promulgators because they are more likely to constrain subjects than are other types of norms, like standards.²¹ Unfortunately, the creation of specific rule content with a low risk of error is costly because it typically requires an understanding of the probable individual circumstances of application, which could take expensive data collection or research.²² These costs would be prohibitively massive in the system of individualized rules that we are imagining unless there were a way to generate reliable rule content at a low cost.

The promise of automated individualized rule making is that it could drastically reduce the cost of rule promulgation. Once the data collection and processing infrastructure is set up, the cost differential between issuing a directive once per week versus once per second is near negligible. This insight has informed a novel prediction regarding the type of legal

of faster routes. DARYL, 'How IP Routing Works,' *Think Like a Computer* (24 August 2011), online: Great White Hosting <<http://www.think-like-a-computer.com/2011/08/24/ip-routing/>>: 'Dynamic routing allows routers to "talk" to each other to find where other networks are located. When the network topology changes so do the dynamic routes. When routers go down or faster routes become available dynamic routing also detects this and reconfigures the IP routing table accordingly.'

- 18 Yu-Che Chen & Michael J Ahn, *Routledge Handbook of Information Technology in Government* (London: Routledge, 2017) at 109.
- 19 Greg von Portz & Satish Misra, 'Medtronic & Fitbit Partner To Connect Activity Data with Continuous Glucose Monitoring,' *iMedical Apps* (24 January 2017), online: iMedicalApps <<https://www.imedicalapps.com/2017/01/medtronic-fitbit-partner-connect-activity-data-continuous-glucose-monitoring/>>.
- 20 See Shasta Darlington, 'Waze App Directions Take Woman to Wrong Brazil Address, Where She Is Killed,' *Cable News Network* (8 October 2015), online: Cable News Network <<http://www.cnn.com/2015/10/05/americas/brazil-wrong-directions-death/index.html>>.
- 21 Brian Sheppard, 'Judging under Pressure: A Behavioral Examination of the Relationship between Legal Decisionmaking and Time' (2012) 39 Fla St UL Rev 931 at 980–5 (empirically examining the constraining power of rules and standards).
- 22 See e.g. Louis Kaplow, 'Rule Versus Standards: An Economic Analysis' (1992) 42 Duke LJ 557 at 621–2.

norm content we can expect in the future. Anthony Casey and Anthony Niblett ‘predict a world of precisely tailored laws (“micro-directives”) that specify exactly what is permissible in every unique situation. These micro-directives will be largely automated. If the state of the world changes, or if the objective of the law is changed, the law will instantly update. The law will become “self-driving.”’²³ The creation of individualized rules through automation also paves the way for automated adjudication because it shrinks the role of judicial discretion. In that vein, Casey and Niblett describe a process whereby the general information and predictions that technology provides will become so reliable that they will pave the way to thoroughly machine-based law creation and adjudication.²⁴ They describe the adjudicative replacement in this way:

A computer-driven algorithm to predict the likelihood of a defendant skipping bail is already being used in some jurisdictions in the United States. But this algorithm has not completely replaced human judges, yet. The transformation will take time. The algorithm is currently used to provide human judges with a better forecast of the risk of flight. Soon, we imagine, the algorithm will provide recommendations as to how the judge should decide. These recommendations could be followed or ignored by the human judge. But as more information is generated, and the evolutionary algorithm updates and becomes a better forecaster, we imagine that judges will increasingly rely on advice of the algorithm. Over time, with increased acceptance, the algorithm will become the law. The algorithm will effectively replace the judge.²⁵

Mechanizing adjudication would feed back on legislation, making it easier to predict the state of affairs that rule promulgation will bring about. It might lower cost and increase effectiveness. Traffic congestion is only one problem, but it is an important one, and heavy, technology-assisted regulation might be the best way to solve it. As time goes on, the appeal of an aggressive regulatory regime powered by machines could significantly grow and so too could our willingness to consider concepts that utilize it to even more ambitious ends. We might soon be keen to accept a similar system to minimize the risk of personal injury or to maximizing economic efficiency, for example. To be clear, these developments are not inevitable, nor do I necessarily desire them, but they will plausibly appeal to the officials and subjects of some future state.²⁶ Even short of that, this will be

²³ Anthony Casey & Anthony Niblett, ‘Self-Driving Laws’ (2016) 66:4 UTLJ 429.

²⁴ *Ibid.*

²⁵ *Ibid.*

²⁶ It is also imaginable that the problem could be approached with minimal state involvement. Today, many leading technology companies make frequent, automated adjustments to their offerings in light of user data without government oversight.

the stuff of future intuition pumps in legal philosophy. It is important, even for conceptualists, to have an understanding of what such systems would likely require and how they would likely work.

B MACHINE LEARNING IN AN ENVIRONMENT OF AUTOMATED LEGISLATION AND ADJUDICATION

A computerized system for human conduct regulation must be based on programming that is highly scalable; it needs to maintain its performance levels under increased and expanded workloads and with data derived from an ever-changing environment. For the system to keep up or even improve under such conditions, it is likely that the machine powering it will have to improve itself rather than rely on initial or iterative human programming. Otherwise, the tremendous costs of high-volume legislation and, in turn, adjudication could return in the form of programming costs. The most likely way in the foreseeable future to handle the programming load without a surplus of human intervention is a technique called machine learning (ML). ML employs an initially programmed algorithm capable of subsequent ‘learning’ from the data it processes and, thereafter, creating its own algorithms or other programming rules.²⁷

We have all taken part in massive ML projects, whether we realize it or not. As a safeguard against bots, many websites use tests designed to rule out the possibility that the thing accessing the site is a machine-powered bot. For the last few years, Google has used a bot screener that teaches an algorithm to identify objects in photographs.²⁸ When visiting a website, a user might be shown a series of photographs and asked to identify which ones contain objects of a certain kind, such as a ‘river,’ ‘store front,’ or ‘street sign.’²⁹ Google’s ML algorithm analyzes this user data, searching for patterns that correspond to positive or negative user responses on the object query. Once found, the program incorporates those patterns into its object-labelling algorithm, improving performance over time.

Thus, ‘[ML is] a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data).’³⁰ It is math, and the relations that

²⁷ See Gillian Hadfield, *Rules for a Flat World* (Oxford: Oxford University Press, 2016) at 152.

²⁸ Hal Hodson, ‘Google’s New Bot Trap Trains Machines to See the World,’ *New Scientist* (10 December 2014), online: New Scientist <<https://www.newscientist.com/article/mg22429992-400-googles-new-bot-trap-trains-machines-to-see-the-world/>>.

²⁹ *Ibid.*

³⁰ Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (Cambridge, MA: MIT Press, 2012) at 1.

it identifies based on the data it processes can feedback into the program to give it greater power in performing the overall task that it has been given.³¹ But this overall task must be given; ML is a component of artificial intelligence, but it is not going to give machines the capacity to choose to work on entirely different projects or to otherwise deviate from their overall goals. What it can do, however, is identify patterns and, thereafter, pursue improved sub-operations in service of the overall mission that it has been given.³² In this regard, the system is a good fit for a project seeking improved performance regarding a specified goal but not for a project where the overall objectives are in flux.

In the next Part, I will discuss how the manner in which ML operates is friendly to the creation of a rule-based legal system that is quite different from those to which we are accustomed.

1 *ML versus handcrafted rule systems*

When people consider their own legal systems, they tend to focus on the ‘legal’ part rather than on the ‘system’ part, taking for granted that the laws to which they are subject exist in some sort of orderly relationship to each other. Theorists, however, have spilled a lot of ink regarding the systematic character of ideal legal systems, often considering them in terms of maps, networks, or webs of logically related legal concepts.³³ While they might debate whether it is possible to create a web without gaps or conflicts, few deny that legal interpreters in modern legal systems (particularly common law systems) often consider, among other things, semantic elements that emerge from the relationships between laws or from structural aspects of a legal system as a whole.³⁴ In the context of adjudication, these interpretive components can supplement the persuasiveness and validity of a particular legal outcome.

If ML were to promulgate a set of rules, those rules would each be in service of an overall, pre-programmed purpose. Although they would

³¹ John Paul Mueller & Luca Massaron, *Machine Learning for Dummies* (New York: For Dummies, 2016).

³² Even Q-learning or reinforcement learning – the aspect of machine learning (ML) that has achieved the most high-profile recent successes (e.g., the Alpha Go victory) and arguably its most autonomous – is beholden to overall goals. Cf Patrick R Nicolas, *Scala for Machine Learning* (Birmingham, UK: Packt Publishing, 2014) at 444.

³³ See e.g. Gerald Postema, ‘Law’s System: The Necessity of System in Common Law’ (2014) 2014 NZLR 69 at 78–80 (discussing the popularity of a view of law’s necessary systematic character from the advent of common law in jurisprudence).

³⁴ Of course, syntax and pragmatics also play important roles in legal interpretation, as do more elaborate jurisprudential approaches. For a lengthy discussion of the relationship between technological development and legal interpretation, see Brian Shepard, ‘Incomplete Innovation and the Premature Disruption of Legal Services’ (2016) 2015 Michigan L Rev 1797 at 1849–98.

each share the same goal, the rules would not likely form an underlying structure of meaning with each other. Thus, there would not likely be a conceptual network within the system of rules that provides a measure of synoptic clarity regarding the interpretation of rule content.³⁵ Despite their shared goal, the directives would not collectively give rise to tools for culling meaning beyond that the simple connection between an individual directive and the purpose that drives the overall system. In this regard, our imagined traffic system would not resemble familiar legal systems. We could infer that the directive to, say, turn left at the intersection is the best alternative under the machine's analysis for furthering the overall goal of reducing traffic congestion – the very same inference that we could draw regarding any directive within the system – but ML would not give rise to explanatory principles that apply to families of directives because ML is not designed to make interpretive connections between directives, let alone to use those connections to limit or augment future directives. This might not be such a bad thing. Were ML to limit the revision of its rules to maintain a semantic network, it would likely hamper the efficiency and effectiveness of the ML system as it evolves. This deserves some explanation.

It is possible under existing technology to model semantic properties of real laws without the use of ML, so it is fair to question the necessity of employing ML in the creation and execution of the legal system we envision. In today's world, computer scientists are already engaged in large-scale semantic web coding, including the development of semantic webs that attempt to model existing laws.³⁶ These rely on ontologies, or pre-programmed rule structures, that seek to correspond with our understanding of the law and its concepts.³⁷ Often a semantic web requires

35 'Rule-Based Machine Learning,' *Wikipedia*, online: Wikimedia <https://en.wikipedia.org/wiki/Rule-based_machine_learning>: '[R]ule-based machine learning applies some form of learning algorithm to automatically identify useful rules, rather than a human needing to apply prior domain knowledge to manually construct rules and curate a rule set.'

36 Pompeu Casanovas et al, 'Special Issue on the Semantic Web for the Legal Domain Guest Editors' Editorial: The Next Step' (2016) 7:2 *Semantic Web Journal* 1: 'Ontology-driven systems with reasoning capabilities in the legal field are now better understood. Legal concepts are not discrete but make up a dynamic continuum between common sense terms, specific technical uses, and professional knowledge in an evolving institutional reality. . . . In this sense, Semantic Web tools are not only being designed for information retrieval, classification, clustering, and knowledge management. They can also be understood as regulatory tools, i.e., as components of the contemporary legal architecture, to be used by multiple stakeholders – front-line practitioners, policymakers, legal drafters, companies, market agents, and citizens.'

37 See Thomas R Gruber, 'A Translation Approach to Portable Ontology Specifications' (1993) 5 *Knowledge Acquisition* 199 at 199; Craig Roberts, 'Context in Dynamic

not simply the programming of a coherent concept for each object in the system but also one that is adequately ‘rule-ifiable’ to permit a machine to use it in a reliable classification.

One significant problem is that this approach would not likely scale to a project of the magnitude that we imagine.³⁸ Google’s object-identification project is instructive. We might think that the best approach to the automated identification of objects would be to create a decision-tree structure that follows our conceptual categories for objects. A pipeline of if-then tests that eventually leads an object to fall into a category of objects sounds intuitive – for instance, if it is round, orange, is being thrown at a rim, and so on, it is a basketball. As professors of introductory philosophy courses know, however, a satisfactory rule-based concept for something as simple as a basketball can be tough to come by. Google came to this realization after reaching a scalability ceiling using this approach.³⁹ The same impediments would exist for the legal project, we imagine.⁴⁰

ML, by contrast, does not require an initial concept of an object at all, so it limits upfront costs and unshackles performance in the process. Directives can be generated so long as they meet statistical tests showing that they further the overall objective in the best way; they need not cohere or otherwise maintain consistency with a semantic web. We can see, then, that a similarly top-down approach to human conduct regulation – even one that attempts to create a decision-tree-like structure for automated adjudication from existing laws – is probably not a favourable alternative.⁴¹ Thus, it is reasonable to guess that an automated system for legislation and adjudication that fully delivers on the promise of automation – namely, through the delivery of highly contextualized micro-directives in

Interpretation’ in Lawrence R Horn & Gregory Ward, eds, *The Handbook of Pragmatics* (Hoboken: Wiley-Blackwell, 2002) 197.

³⁸ Another problem is that technology is unlikely to develop competent understanding of the pragmatics of natural language for many decades. See Erik Cambria & Bebo White, ‘Jumping NLP Curves: A Review of Natural Language Processing Research’ (2014) 9 *Institute of Electrical and Electronics Engineers Computational Intelligence Magazine* 48 at 48 [Cambria & White, ‘Jumping NLP Curves’].

³⁹ Natalie Gagliardi, ‘Should Google Be Your AI and Machine Learning Platform,’ *ZDNET* (1 December 2016).

⁴⁰ See Andreas Hotho, ‘Data Mining on Folksonomies’ in Giuliano Armano et al, eds, *Intelligent Information Access* (New York: Springer, 2010) 57 at 73.

⁴¹ Indeed, we are only at the very beginning of understanding just how complex legal systems are, and there are ways that we might be able to harness both ML and rules-based approaches on the same projects. See JB Ruhl & Daniel Martin Katz, ‘Measuring, Monitoring, and Managing Legal Complexity’ (2015) 101 *Iowa L Rev* 191 at 240–4. But there can be little doubt that ML will be an attractive tool in technology-assisted legislative and adjudicative efforts on the horizon, thereby bringing its strengths and weaknesses into our future philosophical discourse.

the service of a broader goal – will neither follow established rules of inference nor form a semantic web. To be sure, future efforts at automation might successfully model the web of concepts in an existing legal system and use that as the basis for rule making and applying, but it would not likely produce satisfactory results at scale until long after a goal-based, unsupervised ML approach could.⁴²

2 *the strangeness and unintelligibility of ML-created directives*

Imagine again that a futuristic society has a ML-powered legislature addressing the problem of traffic congestion. Further imagine that each citizen must wear a device that tells her or him of any legislative directives. A citizen in this fictional state is rushing to the wedding of her best friend when suddenly she receives notice of the authoritative directive that she is to turn around immediately, return home, and remain there for the next six hours. She might not understand why she has received the directive; it just happens to be the case that the machine reached a statistical threshold whereby it could make a sufficiently reliable prediction that its overall goal of reducing congestion would be best served by issuing that directive. She might come to accept such a general and likely unsatisfactory justification for her constraint, but she might also feel like she is paying an unfairly high price or worry that there was a simple machine error.⁴³ In that case, she might desire a specific and compelling explanation regarding why the machine dealt with her particular situation in that manner. Unfortunately for her, she might want the impossible.

As algorithms have become an established part of high-stakes projects, there has arisen concern that they are not adequately transparent to allow for accountability when they are used as the basis for harmful or coercive decisions. Indeed, the European Union recently passed the General Data Protection Regulation, which provides the right ‘not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her’ unless certain safeguards are met.⁴⁴

⁴² Interestingly, computer scientists have been working on modelling modest semantic networks of legal concepts for decades. See e.g. *Law Moose*, online: Pritchard Law Webs <<http://www.lawmoose.com/>> (semantic search of Minnesota law).

⁴³ For an interesting recent discussion on the degree to which automated systems can provide satisfactory explanations in the context of criminal law in the United States, see Kiel Brennan-Marquez, ‘Plausible Cause’ (2017) 70 Vand L Rev 1249; Andrew D Selbst, ‘Response: A Mild Defense of Our Robot Overlords’ (2017) 70 Vand L Rev en Banc 87.

⁴⁴ EU Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, [2016] OJ L119/1.

However, safeguards could easily fall short. Tarleton Gillespie, the principal researcher at Microsoft Research New England, recently claimed: ‘There may be something in the end impenetrable about algorithms.’⁴⁵ Others are more hopeful, but there is an emerging consensus among computer scientists that it will take aggressive research to cut through algorithmic opacity, particularly in ML, where opacity is at its highest.⁴⁶

There is an important bottleneck, however. There are only so many decision factors that we can comprehend within a reasonable amount of time, and when analyzing a decision, we generally need the justification for that decision to be in intelligible language, sufficiently comprehensive, and reasonably short. Thus, there are semantic and pragmatic dimensions to our understanding of what makes a legal decision justified. Because these problems are by-products of our own epistemic limits, they are going to be very difficult to solve. Take a single dimension of law-related work: legal research. Leading companies like Westlaw and Lexis have begun to use ML to improve their natural language searches, using feedback from users to change the way that results are collected, selected, and ranked by the algorithm.⁴⁷ Even this process, with limited variables, basic inquiries, and similarly organized and structured texts like legal opinions, can produce inscrutable results to experts.⁴⁸ Bob Kingan, a data scientist at Bloomberg Law, explains: ‘Many machine learning techniques result in models of the data that consist of, say, hundreds of thousands to millions of numerical weights used to determine how input data is transformed to output. One can apply tests to such an algorithm and review examples from “gold standard” training data to get a feel for how it behaves, but it may be impossible to interpret the algorithm itself in human terms.’⁴⁹

An important question is whether the intelligibility problem is likely to be solved in the short term. Early signs are not encouraging. Even the

45 Tarleton Gillespie, ‘The Relevance of Algorithms’ in Tarleton Gillespie, Pablo Boczkowski, & Kirsten Foot, eds, *Media Technologies: Essays on Communication, Materiality and Society* (Cambridge, MA: MIT Press, 2012) 167.

46 See Aaron M Bornstein, ‘Is Artificial Intelligence Permanently Inscrutable?’ *Nautilus* (1 September 2016).

47 See Brian Sheppard, ‘Does Machine-Learning-Powered Software Make Good Research Decisions? Lawyers Can’t Know for Sure,’ *American Bar Association Journal* (22 November 2016), online: American Bar Association <http://www.abajournal.com/legalrebels/article/does_machine-learning-powered_software_make_good_research_decisions_lawyers> [Sheppard, ‘Does Machine-Learning’].

48 For an interesting article that goes into depth on the inscrutability of the computer-directed movements of autonomous cars, see Harry Surden & Mary-Anne Williams, ‘Technological Opacity, Predictability, and Self-Driving Cars’ (2016) 38 *Cardozo L Rev* 121.

49 Quoted in Sheppard, ‘Does Machine-Learning,’ supra note 47.

entities that are leading the transparency charge have humble expectations when it comes to ML. The IBM mission statement on transparency gives solving intrinsic opacity only fleeting consideration, stating only that it will provide information on why it is using artificial intelligence in cognitive solutions as well as ‘major sources’ of data that ‘inform’ the insights of those solutions, but stating that it will ‘protect’ client data and insights.⁵⁰ Likewise, the academic/practitioner watchdog group Fairness, Accountability, and Transparency in Machine Learning articulates guiding questions for practitioners regarding ‘explainability,’ stating that ‘[i]f you are using a machine-learning model: *consider whether* a directly interpretable or explainable model can be used.’⁵¹

The best hope is that there might be a separate scalable program that can render the operations of a constantly changing algorithm intelligible, but the results of cutting-edge research only show how far we are from breaking our epistemic limits. At the bleeding edge of progress is LIME, a program that claims to bring ‘a novel explanation technique that explains the predictions of any classifier in an interpretable and faithful manner, by learning an interpretable model locally around the prediction.’⁵² What this means is that the program is capable of boiling down and visualizing the factors that were most likely to have been determinative to the ML program when it classified something in a particular way.⁵³ LIME highlights factors against, or in favour of, a prior classification made by the learning program that were close to the line delineating one decision from a competitor decision. The particular way in which this occurs is too complex for this article, but LIME selectively turns on or off factors in the neighbourhood of those that are likely to have determined the classification. This allows LIME to determine the factors that have the most predictive weight and visualize them. Provided the factor survives an automated statistical analysis to rule out insignificance, LIME then displays the winner as the ‘explanation.’⁵⁴

50 IBM, ‘Transparency and Trust in the Cognitive Era,’ *IBM THINK* (blog) (17 January 2017), online: IBM <<https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/>>.

51 Quoted in Nicholas Diakopoulos et al, ‘Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,’ *FAT/ML*, online: *FAT/ML* <<https://www.fatml.org/resources/principles-for-accountable-algorithms>> [emphasis added].

52 Marco Tulio Ribeiro et al, “Why Should I Trust You?” Explaining the Predictions of Any Classifier’ (paper delivered at the Proceedings of the Twenty-Second Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining, 2016).

53 See *ibid.*

54 See *ibid.*

While LIME is doing important early work on intelligibility, it should be clear that its technique of filtering down factors and characterizing the locally weightiest one as an ‘explanation’ has some significant weaknesses. For one, knowing the factor that has been given the most weight close to a decision line is not necessarily going to explain the function of the model. It is possible that the weightiest factor under the analysis still only accounts for a tiny fraction of the overall weight of factors considered in the decision. In an analysis involving thousands (or even millions) of variables, such as would be the case in our imagined traffic system, the likelihood that the weightiest factors would suffer from this deficiency would be quite high. Moreover, in close cases, there might be thousands of factors, any one of which could qualify as determinative. Lastly, it is possible that the most troubling factors of significant weight did not happen to be in the presumed neighbourhood of the decision and, therefore, escaped the attention of the program.

Even if LIME or programs like it could reveal the most important factors, we might not comprehend the causal relationship between those factors and an automated decision, either because of their strangeness or because they are so numerous that we simply would not have the time or cognitive capacity to process them all. In the best case, we can see visualizations of factors and conclude that the program is sound because it mirrors our phenomenology of decision making in that context. However, this might not be true in non-banal cases. When faced with inscrutable explanations, there is a risk that we might conclude that the manner in which the model works is flawed or not to be trusted, though it might be a weakness in our own approach that is the problem. The risk of this sort of error increases with the sophistication of the ML program. A perplexing micro-directive might make sense only if we had a far greater understanding of the causal relations of our world, one that we might simply be too dumb to fathom. Even if we were shown these leading factors, we might dismiss them as being misplaced or silly, but they might actually reveal true relations that matter. This brings to mind a trope in science fiction: the wisdom of hyper-intelligent beings might sound like nonsense.⁵⁵

To produce a sufficient textual explanation of an adjudicative decision is even more demanding than explaining machine classifications, like LIME’s analysis of image recognition programs, because we generally expect those decisions to have articulable, reasoned justifications that are compact enough to be comprehended in a reasonably short amount of time. Listing the hundreds of thousands of factors that led to the

⁵⁵ This is not so different from the seemingly nonsensical game of Brockian ultra cricket in Douglas Adams’s most famous book, *The Hitchhiker’s Guide to the Galaxy* (New York: Del Rey, 1997) at 152.

directive that caused you to miss the wedding, any one of which might have led to a different micro-directive, will not do. We would hope that the machine would interpret the particular decision and produce writing that adequately answers to our semantic and pragmatic demands, something that is many decades away.⁵⁶ Besides, an ML-powered system is agnostic to the importance of a causal relationship beyond its predictive reliability; it answers only to prediction results. As Jenna Burrell explains, ‘[w]hen a computer learns and consequently builds its own representation of a classification decision, it does so without regard for human comprehension. Machine optimizations based on training data do not naturally accord with human semantic explanations.’⁵⁷

Worse yet, it may be the case that the more intelligible a system is, the less effective it will be at prediction, suggesting that human understanding of prediction might hold back performance. In an influential 2009 study, Jonathan Chang and colleagues tested human subjects’ ability to spot a non-conforming word in a family of concepts and a non-conforming label in a family of concepts and then compared their performance to the performance of ML programs on the same task. They discovered that the ML program with ‘[t]he highest probability of success did not have the best interpretability; in fact, the trend was the opposite.’⁵⁸

III *The legality of a plausible automated system of conduct regulation*

I have already described the first half of my hypothesis – namely, that the upside of ML-powered automation of highly contextualized, effective, and cheap directives will make it tempting for us to accommodate thoroughly automated systems based on ML in our concept of law. It remains, however, to explain how this accommodation, or even this temptation, could lead to a revision of the concept of law. I will consider the approaches of HLA Hart, Joseph Raz, and Ronald Dworkin.

A HART: AUTOMATION AND THE PROBLEM OF CRITICAL OFFICIALS

Hart is undoubtedly the most important legal conceptualist of the last century. His most fundamental contribution to jurisprudence was providing a better, richer alternative to deflationary concepts of law and legal systems, like those of John Austin, which conceptualize law as little

⁵⁶ See e.g. Cambria & White, ‘Jumping NLP Curves,’ *supra* note 38.

⁵⁷ Jenna Burrell, ‘How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms’ (2016) 3:1 *Big Data & Society* 12.

⁵⁸ Jonathan Chang et al, ‘Reading Tea Leaves: How Humans Interpret Topic Models’ (2009) 22 *Advances in Neural Information Processing Systems* 288 at 293.

more than the command of the sovereign backed by a sanction that leads to a habit of obedience among the general population.⁵⁹ Hart's alternative to this 'command theory' positivism emphasized the role of social conventions among officials that gave the laws that come from them a special normative character.⁶⁰ As a result of the pivotal role that officials play, it is fair to ask whether automation of that role would strip a system of its legality. To answer the question, we must first examine Hart's articulation of the role of officials in creating and maintaining a legal system.

Hart believed that command theory failed to explain how legal systems differ from systematic coercion, like a gunman in a bank demanding a purse writ large.⁶¹ Nor could it explain how legal systems persist when lawmakers change, such as when there is a succession of an absolute monarch.⁶² The persistence of a legal system requires that there be some sort of meta-rule governing the transfer of power that is distinct from the commands of a sovereign – namely, a socially accepted standard that keeps the legal system in place.⁶³ The same principle also distinguishes legal systems from the gunman writ large; a gunman might oblige you to pay the purse through the threat of death, but he does not obligate you to do so because there is no socially accepted standard that gives him that power.⁶⁴ Following this line of thought, Hart concluded that a legal system is a union of primary and secondary rules.⁶⁵ Primary rules are what people typically think of as laws (norms that directly regulate conduct by imposing obligations to perform or abstain from certain things, such as criminal prohibitions on murder or tax requirements), whereas secondary rules are norms that govern primary or secondary rules (rules of succession or rules of adjudication).⁶⁶ According to Hart, primary rules impose duties, and secondary rules confer power.⁶⁷ Command theory was defective because it had the former but not the latter.

The most important secondary rule is the rule of recognition. It is the rule that provides the test for identifying the primary rules (for example,

⁵⁹ John Austin, *The Province of Jurisprudence Determined*, 2d ed (London: John Murray, 1861) at 19–23, 173.

⁶⁰ HLA Hart, *The Concept of Law*, 3d ed (Oxford: Oxford University Press, 2012) at 116 [Hart, *Concept of Law*].

⁶¹ HLA Hart, 'Positivism and the Separation of Law and Morals' (1958) 71 Harv L Rev 593 at 603.

⁶² Hart, *Concept of Law*, supra note 60 at 54.

⁶³ Ibid at 56–65.

⁶⁴ Ibid at 85.

⁶⁵ Ibid at 90–110

⁶⁶ Ibid.

⁶⁷ Ibid.

rules prohibiting theft or imposing taxes) of a legal system, so it is the lynchpin that holds the system together.⁶⁸ It is important to point out that the rule of recognition is unlike the laws that we are accustomed to following: it is a social rule, meaning it comes into being based on the widespread adoption of a certain attitude, an internal point of view, regarding primary rules:

[I]f a social rule is to exist some at least must look upon the behaviour in question as a general standard to be followed by the group as a whole. What is necessary is that there should be a critical reflective attitude to certain patterns of behaviour as a common standard, and that this should display itself in criticism (including self-criticism), demands for conformity, and in acknowledgments that such criticism and demands are justified, all of which find their characteristic expression in the normative terminology of 'ought,' 'must,' and 'should,' 'right' and 'wrong.'⁶⁹

Thus, a legal system differs from mere coercion because it involves a convention of acceptance that changes the way that we reason about things moving forward. For a legal system to exist, then, Hart claims it must satisfy two conditions: (a) the primary rules must be generally obeyed and (b) its rules of recognition must be effectively accepted as common public standards of official behaviour by its officials.⁷⁰ Thus, the officials serve as the normative backbone of the legal system, connecting themselves to the social conventions by having critical reflective attitude to patterns of behaviour.

An essential characteristic of an official who is capable of setting up a rule of recognition is that she can accept that there is some sort of standard, outside of the primary rules, indicating that 'the violation of a rule is not merely a basis for the prediction that a hostile reaction will follow but a *reason* for hostility.'⁷¹ There must be some cognition that the critique of rule violation must be based, at least in part, on the notion that a violation of the rules agreed upon, regardless of their content, is a valid, correct, or right basis for hostility. Can a machine be an official in Hart's sense? It would depend on whether a machine can adopt a critical reflective attitude. While it is difficult to predict whether a machine will ever be able to adopt a conscious attitude of any kind, one is quite unlikely to arise from more powerful versions of existing technology.⁷² This might

⁶⁸ Ibid at 110.

⁶⁹ Ibid at 53.

⁷⁰ Ibid at 116.

⁷¹ Ibid at 90 [emphasis in original].

⁷² See e.g. Pedro Domingos, *The Master Algorithm: How the Great Quest for the Ultimate Learning Machine Will Remake the World* (Philadelphia: Basic Books, 2015) ch 1: 'The Terminator scenario, where a super-AI becomes sentient and subdues mankind with a

be enough to show that Hartian legality would not be satisfied by machines that develop along the current trajectory of technology.

Even if we are willing to accept a non-sentient machine that aims to perform the function of officials under Hart's understanding of legality – say, a computer that has been programmed to consider the validity or correctness of accepting the law as a reason for state coercion based on an emergent social convention – the result would likely be the same. Recall that ML is inapt as a means for choosing an overall goal; it can recursively generate sub-norms in service of a pre-programmed overall goal, but it does so without any critical mode with respect to that goal. It will always speedily chug along in service of the goal, improving as it goes, but it will not be able to step outside of the system of primary rule generation to revise its overall goal. The secondary rules have become inert, and all that would really change would be the constantly updated individualized primary rules. Thus, there is a tension between Hart's concept of law and the trajectory that technologically assisted law might take.

An example will help. It is easy to imagine a rather effective and nimble system of conduct regulation, for example, in which a machine is given a purpose, and in furtherance of that purpose, it creates billions of sub-norms which constitute the entirety of the rules of conduct in the system. We can further imagine that the machine was given its initial purpose by humans long ago and that there is not now, nor has there ever been, a body of officials to track social conventions for the system's secondary rules. In other words, it would be like the micro-directive traffic system but applicable to all legally governed areas of life. This automated system would effectively be a system of primary rules because the secondary rules, insofar as they existed at all, would be unalterable from within the system. The dead hand of the computer scientists who encoded the machine's purpose would rule over future generations.

Hart considered systems of primary rules in *The Concept of Law*, but he characterized them as 'primitive' due to their reliance on emergent, collective decisions of the masses to steer the system.⁷³ For this reason, he thought it fair to think of such systems as 'custom based.'⁷⁴ For him, systems that were not steered by living officials were effectively tethered to the typically slow changes of customary behaviour across the state. He did not contemplate that new primary rules of obligation could frequently arise through an automated system in the absence of officials. Rather, he

robot army, has no chance of coming to pass with the kinds of learning algorithms we'll meet in this book.'

⁷³ Hart, *Concept of Law*, supra note 60 at 91.

⁷⁴ Ibid.

assumed that any system of primary rules would suffer from a static character, inefficiency, and slowness in responding to social developments.⁷⁵

Technology might someday prove Hart wrong, displaying adaptability and utility through automation. We might even imagine an automated system that vainly tries to mimic Hartian secondary rules. ML could be harnessed to recognize the patterns of convergent behaviour necessary to form the social convention that serves as a rule of recognition. Indeed, Hart made some effort to explain how an outside observer might identify that others had adopted an internal point of view with respect to the laws of their system, stating: 'After a time the external observer may, on the basis of the regularities observed, correlate deviation with hostile reaction, and be able to predict with a fair measure of success, and to assess the chances that a deviation from the group's normal behavior will meet with hostile reaction or punishment.'⁷⁶ Importantly, however, even a system that makes an attempt to create secondary rules through simulated 'officials' who do not adopt a critical reflective attitude regarding that job. The statistical correlator, Hart argued, 'will be like the view of one who, having observed the working of a traffic signal in a busy street for some time, limits himself to saying that when the light turns red there is a high probability that the traffic will stop . . . [and] [i]n so doing, he will miss out a whole dimension of the social life of those whom he is watching.'⁷⁷

For Hart, officials recognize that any critique of a rule violation in the legal system must be based, at least in part, on the notion that a violation of the rules is valid regardless of their content. As a result of the power and dangerousness of that content-independent notion of authority, it is important that officials be at least capable of taking a self-critical attitude toward the endeavour. It would be a stretch to say that the machine, which has no alternative but to act as lawmaker in the service of the overall goal, is adopting an internal point of view toward a social rule. To acknowledge a reason means, at the very least, that the acknowledger has some capacity to disregard the reason, compare its weight against countervailing reasons, or follow the reason. And the capacity to decide among those options is important because it helps to keep the rule in lockstep with the changing values of a society while providing a reasoned source of normativity. Indeed, in order for a directive to be obligatory, it must regulate behaviour that is shown by convention to be essential or especially valuable to social life. It is silly to discuss a reason in favour of an action if the promulgator has no choice but to issue a directive

⁷⁵ Ibid at 91–9.

⁷⁶ Ibid at 88.

⁷⁷ Ibid at 90.

compelling that action. Therefore, this system, too, would fall short of Hartian legality despite its ability to mimic the outward performance of officials.

The tension between Hart and the trajectory of technology is now in view. It provides a clue that the following questions might become important to future philosophers: first, if a system of primary rules were to transcend the defects of custom through automation, should we then modify our concept of law to include it and, second, once non-customary primary rules can spring into existence in a system that, so far at least, has paid the goods, why ought we require revisable and normatively robust secondary rules to establish a legal system? The best answer could be that our participation in the formation of a social standard that can entirely change how we determine legal validity is precious and essential to the very motivation to be governed. However, future thinkers might arrive at a different conclusion. There can be little question that a system of critically reflective officials (and subjects) is a source of change, with the role of political and moral discourse taking centre stage. A system that automates legal officials, using a more developed version of the most promising technology we now have, would be comparatively passive. Our concept of law might someday be revised to accommodate this passivity.

B RAZ: UNINTELLIGIBILITY, REASONS, AND THE PROBLEM OF LEGITIMATE AUTHORITY

Raz has made great progress in explaining how it can be rational to subject oneself to law. In this regard, he provides a vivid account of how accepting law as an authority affects the reasons that a subject considers when making decisions, and this will be the focus of this Part. Raz asserts that a defining characteristic of law is that it claims legitimate authority over its subjects.⁷⁸ While some legal systems fail to be legitimate, they must at least claim to be legitimate authorities in order to qualify as legal systems.⁷⁹ The law's claim to legitimacy is a claim to moral authority: it will lead those subject to it to better comply with the reasons that apply to them.⁸⁰ A system that cannot sincerely claim to be seeking to perform this service is not legal.⁸¹ This test is premised on the notion that authoritative directives should be based on the relevant reasons under the

⁷⁸ See Joseph Raz, 'Authority and Justification' in Joseph Raz, ed, *Authority* (New York: New York University Press, 1990) 115 at 131–2.

⁷⁹ See Joseph Raz, *The Authority of Law: Essays on Law and Morality* (Oxford: Clarendon Press, 1979) at v–vi.

⁸⁰ See *ibid* at 14.

⁸¹ See Brian Leiter, *Naturalizing Jurisprudence* (Oxford: Oxford University Press, 2007) at 172: 'Raz's thesis is only that all laws (sincerely) claim moral authority.'

circumstances that already independently apply to the subjects of those directives ('the dependence thesis').⁸²

When we are deciding, Raz assumes, there is a right way of doing it – a way that comports with moral and other reasons that apply. People may perfectly balance the reasons for and against something, but people have some ability to identify better and worse ways of assigning weight or correct and incorrect final decisions. The normal way to assess whether a system is living up to its claim of authority is to ask whether a person subject to the directives of that system is more likely to comply with right reason by accepting those directives as authoritative than she would have by weighing the reasons under consideration on her own.⁸³ This is called the 'normal justification thesis' (NJT). When one accepts a directive as authoritative, 'the fact that an authority requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should replace some of them.'⁸⁴ At the risk of oversimplification, when a legal directive tells us to do X, then we must exclude from our thinking all of the reasons to not do X that are within the scope of the reasons upon which the authority's directive was based (and which we might have misbalanced without the directive); this stacks the deck in favour of doing X, which makes it much more likely that we will do it. If the authority really is better at figuring out right reason in that instance, and X is the right thing to do, then the law has performed a valuable service by increasing the likelihood that we will do X. On this account, it can make good sense to subject oneself to the authority of another; it can make you a better decider.

How do we know the limits of the range of reasons that are off limits for us to consider? Recall that the dependence thesis requires that authorities base their directives on reasons that apply to the subjects of those directives in the relevant circumstances. This means that the authority, in issuing the directives, must be balancing within some range of reasons what the right thing is for the subjects of its directive to do. This sets its scope: the scope of the reasons that are excluded is the range of the reasons that the authority considered when issuing the directive.⁸⁵ Anytime there are reasons outside of the scope of the directive that outweigh the reasons that remain in favour of the directive (after exclusion), then the person must violate the directive. So when a person is assessing whether a subject is justified in accepting a system's

⁸² Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986) at 53.

⁸³ *Ibid.*

⁸⁴ Joseph Raz, *Ethics in the Public Domain* (Oxford: Clarendon Press, 1995) at 214.

⁸⁵ See Heidi Hurd, 'Challenging Authority' (1991) 100 *Yale LJ* 1611 at 1638 (linking scope to the dependence thesis).

claim to be a legitimate authority, she should normally consider only the scope of the relevant reasons that formed the directives to which he or she is subject.

This is where technology-assisted law poses a big problem. Harnessing ML in a highly complex environment will almost certainly result in strange directives without intelligible justifications. A subject who is struggling to follow the NJT to assess the legitimacy of the system's claim to authority will not have a good sense of the set of reasons that applies to the directive and, therefore, will have a difficult time knowing whether a conflicting reason is excluded or falls outside of its scope.⁸⁶ The volume and individualization of the micro-directives would be problematic as well. People who are subject to dozens of inscrutable directives per day would find it impracticable, if not impossible, to apply the NJT to even a small fraction of them. This trouble is exacerbated by the fact that we cannot likely enlist the help of others because the directives are tailored to our individual circumstances. Still, these challenges might not be so insurmountable that we could say that the system is incapable of satisfying the NJT, which would be necessary to claim that a system of this sort is categorically not a legal system.

However, even if we ignore the NJT, we still have the dependence thesis to consider. ML programs generally base their directives on a statistical analysis of factors that apply to the individuals that are subject to the directives they issue. They do this, however, without any thought regarding the right balancing of moral and other reasons that go in favour of, or against, a particular choice. Instead, they simply evaluate their predicted impact as to the overall goal. This leaves us with the question of whether there is something intrinsically important about the process of balancing reasons that is distinct from the consequences of that balancing. Raz believes that reasoning is intrinsically valuable because it is our *telos* as human beings; it is the process by which we constitute that which we are supposed to be: 'In postulating that authorities are legitimate only if their directives enable their subjects to better conform to reason, we see authority for what it is: not a denial of people's capacity for rational action, but simply one device, one method, through the use of which people can achieve the goal (*telos*) of their capacity for rational action.'⁸⁷

⁸⁶ One counter-argument might be that subjects should simply assume that the machine is looking at all data (and, therefore, all reasons) that exist within its vast system of detection. Following this reasoning, the subject could simply assume that all decisions come from some sort of balancing of all machine-detectable reasons. But this subject would not likely be able to comprehend the presumably vast numbers of variables, let alone the frequently changing weight that the machine assigns to those variables.

⁸⁷ Joseph Raz, 'The Problem of Authority: Revisiting the Service Conception' (2006) 90 *Minn L Rev* 1003.

Rational action is not passive; it is not simply that the chosen action turned out to be the right one but also that it was done for the right reasons. In other words, the reasons matter too, not just the action that comes out of considering reasons. As Michael Sevel explains, '[legitimate authorities] tell us what to do, and what it is we're doing when we obey them. They not only provide reasons for action, but provide the very understanding and conceptualization of those actions, quite apart from a reason to perform them, to the obedient subjects who do in fact perform them.'⁸⁸ Thus, a clear directive with an inscrutable basis in reasons robs us of the most important aspect of accepting an authority – that we can better learn how to balance reasons and, thereby, better embody what we are supposed to be as humans. Nevertheless, we must not forget that these laws might be quite effective at bringing about favourable states of affairs under morally sound overall goals.

Still, at the risk of restating a bromide, the advance of technology into law-making or adjudication will challenge the degree to which law reflects humanness. But, if one understands Raz, one can pinpoint the source of the problem. ML-generated law will likely make it harder for us to identify the reasons that underlie the directives that promise to help us, damaging the capacity of law to perform that service. There is a certain irony, then, in the way that technology might steer our decision making through law: an automated legal system could be too good to be legal. Conceivably, a machine could be so good at identifying, weighting, and balancing reasons for and against an action that it would struggle to improve our rational, human capacities to do the same. Michael Jordan might simply be too good at basketball to be a good coach. Sadly, it does not appear as though there is a win-win situation; we will not likely be able to increase machine intelligibility without limiting its performance. The question, then, is whether we would be willing to put a leash on the system for the sake of keeping it within our own epistemic limits. If we are willing to sacrifice intelligible, knowing self-improvement through law for the faith that our actions are best improved through the help of an inscrutable law machine, we might feel compelled to abandon a Razian concept of law.

C DWORKIN: GOAL-BASED PROGRAMMING AND THE PROBLEM OF JUSTIFYING COERCION

Dworkin made a splash in the early phase of his academic career by espousing the 'right answer thesis' in which he claimed that in resolving

⁸⁸ Michael Sevel, 'The Constitution of Authority' (2014) 5:2 *Jurisprudence* 430.

all cases, even hard cases, legally authoritative standards recommend a single right answer.⁸⁹ This might initially seem to be a legal philosophy that is compatible with the theoretical automation project that I have been describing because it appears to be mechanical and determinate. However, this misunderstands the thesis. Dworkin, himself, foresaw the error. In *Law's Empire* (1986), he wrote the following:

Have I said what law is? The best reply is put to a point. I have not devised an algorithm for the courtroom. No electronic magician could design from my arguments a computer program that would supply a verdict everyone would accept once the facts of the case and the text of all past statutes and judicial decisions were put at the computer's disposal.⁹⁰

Dworkin's position on programmability comes from his belief that the correct method of adjudication is highly interpretive and evaluative. An adjudicator can find the answer to virtually any case but only by determining what fits settled law and determining which of those interpretations best justifies the coercion imposed under the morality embodied in the law.⁹¹ In doing so, the judge is supposed to preserve law's integrity by interpreting it in its best light, both legislated and judge made, as if it were created by a community speaking with a single voice and which is seeking to provide a justification for state coercion.⁹² Dworkin believed that every case has a right answer, but he nevertheless believed that every case is hard, requiring some exercise in political morality.

So while Dworkin sets forth a rather elaborate protocol for adjudication in all cases, he understands that a legal system, as a whole, is constantly adjusting and then attempting to cohere its parts. While I disagree with Dworkin that no computer could ever be programmed to perform his integrity-based approach, the question here is whether the current pace and trajectory of technology is such that a machine will meet his test for legality before we are likely to revise our concept of law. ML can provide a systematic way to revise a set of rules in service of an overall goal. And that goal might be morally legitimate. In this respect, it is dynamic and structured while also displaying neutrality and consistency as it creates and applies the norms in its system. Were ML to be harnessed to regulate conduct in a law-like system, it would do so with a sort of evolving integrity, albeit mechanical in nature. Still, it would leave those overall goals intact, while it constantly replaces and improves the sub-norms in service of that

⁸⁹ Ronald Dworkin, *Taking Rights Seriously* (London: Duckworth, 1978), ch 4.

⁹⁰ Ronald Dworkin, *Law's Empire* (Cambridge, MA: Harvard University Press, 1986) at 412.

⁹¹ *Ibid* at 93, 285.

⁹² *Ibid* at 218.

goal. The static nature of the policy animating the system would rob it of its Dworkinian legality. For Dworkin, integrity demands that all aspects of the system be subject to revision because disagreement goes all the way down to the very bottom of legality: ‘We do not follow a shared linguistic criteria for deciding what facts make a situation just or unjust. Our most intense disputes about justice – about income taxes, for example, or affirmative action programs – are about the right tests for justice.’⁹³

The justness of an overall policy goal, such as reducing traffic congestion, cannot be assumed on this view. Why is this important? If policy goals were agreed upon in advance under fair procedures, why should it matter if they are fixed? For Dworkin, the answer is that policies do not give rise to individual rights, which are trumps against collective goals. They arise from the principles that become evident through interpretation in the resolution of cases – a process that seeks to coheretize past political decisions in a way that provides a justification for state coercion from the perspective of the interpretive community.⁹⁴ These principles can be used to force deviation from policies. They can arise through the adjudication of cases that were not foreseen when the legislature invoked the policy rationale for its directive.⁹⁵

This is an area of tension between Dworkinian legality and the promise of technology. ML does not admit of principles – it operates by finding statistical relations, which are agnostic to whether they mirror the principles of a legal semantics. Nor might people want it to. Just as ML is hindered by established rules of statistical inference, so too would it be hindered by a system forcing it to limit itself to bonafide community principles and to cohere its directives. It might make it impossible to meet the demands of micro-directive regulation of conduct given its massive scale. Assuming such a technological system actually comes into being, the perceived value of rights in Dworkin’s sense might diminish. Why rock the boat if the system is delivering on its given, agreed-upon policy goals?⁹⁶

⁹³ Ibid at 73.

⁹⁴ Ibid at 381.

⁹⁵ Ibid at 311–12.

⁹⁶ People might further be willing to accept such a system if it could be programmed to provide convincing, ‘as-if’ principles-based explanations for why a particular micro-directive was issued. In other words, even if a punishment under the system were based on billions of factors of differing weight pursuant to an algorithm, it might someday be possible to generate a convincing simulation of a judicial opinion in favor of that punishment. Dworkin, however, is very clear that ‘as-if’ explanations do not function as adequate semantically expressed justifications and effectively abrogate legal rights. This, he assumes, is unacceptable to people. Ibid at 160–2.

IV Conclusion

I have attempted to provide a reasonable prediction regarding how the trajectory of technology might change the way that we think about the essential features of legality. My goal here has not been to sing the praises of an automated system of conduct regulation, especially not the rather extreme and theoretical systems that we have been imagining. Rather, I have sought to show how the perceived strengths of emerging technology in this domain might someday tempt people to revise their concept of law so as to accommodate its weaknesses. I believe that the promise of granular control through ML will be tempered by increasing concerns about its inscrutability. We will likely face important and unprecedented trade-offs between intelligibility and results. The way in which we strike a balance between those things is likely to have a strong bearing on whether and how we revise our concept of law. The shape of legality could depend on our willingness to accept an intimate, all-by-myself relationship with the legal system. It could depend on our willingness to accept broad collective goals even though the commands that come to us in service of those goals become narrower and less intuitive. And, most importantly, it could depend on our willingness to accept a growing disconnect between the way that we decide and the way that the system does.