CHIEF JUSTICE JOHN ROBERTS IS A ROBOT

Ian Kerr[*] and Carissima Mathen[**]

I   Introduction

The title of this article is not pejorative. It is suggestive. It asks readers to imagine the following counterfactual.

Around the globe, people awaken to some very strange news. In different languages, the same headline thunders: "Chief Justice John Roberts is a Robot".

Rendered unconscious during an ambush and attempted kidnapping while attending a conference at the House of Lords, Chief Justice Roberts' captors boldly deliver-him-up to the Royal London Hospital and speed off. In urgent and unusual circumstances—and in breach of US and international protocols—a team of emergency surgeons cut him open to save his life. That is when they discovered that his biology only runs skin deep.

The Chief Justice, it turns out, was a robot. Despite heroic efforts, the surgical staff was unable to revive the Chief. Several renowned teams of surgeons, computer scientists and experts in the fields of cybernetics and artificial intelligence were subsequently consulted with numerous further unsuccessful attempts to reanimate him. Devout practitioners of Kabbalah were called in. They offered various incantations but, upon failure to resurrect JR-R, concluded that the robot, like the great Golem of Prague, had outlived its divine purpose.

After weeks of follow-up investigations and interviews, it is learned that "John Roberts" did indeed graduate from Harvard Law School in 1979 and that "his" legal career unfolded exactly as documented in public life. However, John Roberts, Robot (JR-R) was in fact a highly advanced prototype of the *US Robots and Mechanical Men Corporation*, developed during the period chronologically corresponding with what

would have been "his" high school and college years. After several (earlier) A-series JRs had secretly annihilated the Turing test in the mid 1970s, *US Robots* decided to consolidate its successful AI with emerging robotic technologies in the new R-series machines. As part of its research and development, the company initiated a singular, long-term experiment with a lifelike, autonomous social robot that was virtually indistinguishable from human beings.

The goal of the JR-R experiment was to see whether a social robot could fully integrate into society without any outside help—indeed without anyone even noticing. With a desire to test its intellectual and social capacities, JR-R was programmed to be an incoming 1L. Supported by various successful techniques in advanced cybernetics, haptics, voice recognition, tissue engineering, and regenerative medicine, JR-R was driven by three key artificial intelligence technologies and a breakthrough in affective computing—each of which had been subject to carefully maintained trade secrets. Capable of exhibiting emotion and able to alter its own instructions while executing, JR-R was built to run on self-modifying code, a regenerative exoskeleton and skin (that would degrade along human-aging trajectories) and had a power supply that would last 75 years. Although self aware, JR-R had no idea it was a robot; JR-R was hard-wired to think it was a human being named "John Roberts."

Consequently JR-R never tried to hide its nature from others. This was a key element of the experiment, the fundamental premise of which was for JR-R to integrate on the basis of perceived similarities rather than differences. And there definitely were some key differences. JR-R's microprocessors were able to sense, select, collect, organize, input, store, retrieve, search, process, collate, assimilate, aggregate, analyze, synthesize, estimate, appraise, and evaluate reams of information— exponentially faster than human beings. Even with its nascent database and network capacity, JR-R could, if it chose to do so, read, record and amass nearly the entirety of legal knowledge, with permanent, eidetic memory.

After a successful (though embellished) application to Harvard Law (JR-R did in fact achieve the top percentile in the LSAT and attained the highest-ever score on the personal interview), *US Robots* unleashed JR-R on the world in 1976. Like Aristotle's unmoved prime mover, *US Robots* left JR-R to its own devices. Steadfast in its commitment to a kind of "prime directive,"[1] *US Robots* made no interference with the

---

[1] The guiding principle of the *Star Trek* universe's United Federation of Planets, the "prime directive" prohibits Starfleet personnel from interfering with the internal development of alien civilizations. *See* FRANZ JOSEPH, STAR TREK STAR FLEET TECHNICAL MANUAL (20th ed. 1986). For non-Trekkies, the "prime directive" captures the "zoo hypothesis" explanation to the Fermi Paradox; explaining the apparent absence of extra-terrestrial life despite its accepted plausibility by hypothesising that aliens ignore Earth to allow our unfettered natural evolution and sociocultural development. *See* John A. Ball, *The Zoo Hypothesis,* 19 ICARUS 347, (1973).

robot's moral, legal or social development. Coded, like the best law students, JR-R was programmed to learn how to learn, with no pre-programmed politics or agenda—other than the blind ambitions of a typical 1L. *US Robots* had no reason to believe that JR-R would succeed at integration beyond the 1L orientation week, and therefore never intended to develop a robot with legal decision-making skills. But, somehow, JR-R managed not only to pass but to flourish at law school. The rest, as they say, is history.

Through a clandestine set of automated surveillance techniques built into the machine, *US Robots* was able to successfully maintain secrecy around the entire experiment for two decades, limiting internal knowledge and control of the experiment to its CEO and principal scientist. The experiment was deemed a complete success in 1996 when a technology lawyer, Jane Sullivan, of the New York firm of Pillsbury Winthrop Shaw Pittman, got married without any (public) awareness that her husband was a robot. Since the couple was unable to have children on their own, they adopted two children in 2000.

A few years prior to JR-R's initial appointment by George W. Bush to the D.C. Circuit, the *US Robots* CEO secretly sold its trade secrets for a significant fortune, successfully utilizing a "proof of concept" rather than revealing the JR-R experiment. Shortly afterwards, in early 2001, *US Robots* began to wind down operations. Since the experiment had succeeded without any need for intervention for nearly 30 years, it was decided in the summer of 2001 that JR-R would no longer be monitored but would not be decommissioned; it would be left to its own devices indefinitely.[2] On September 11th 2001, both the CEO and principal scientist died when the Twin Towers went down. Consequently, no living human being knew about the JR-R experiment by the time "John Roberts" had been nominated or appointed to the bench. Incredibly, these details were eventually revealed by JR-R's automated distress signal. Inspired by *Star Wars'* famous robot, R2D2—encoded to display and broadcast an emergency message by way of its holographic projector[3]— JR-R was similarly programmed to reveal its genesis and the nature of *US Robot's* experiment upon extreme malfunction. It was clear from JR-R's distress signal that *US Robots* had not seriously contemplated that it would rise to the upper echelons of society. JR-R's automated broadcast was much more of a "return to sender."

---

[2] After 30 years of a full life—diplomas, an unparalleled career, a wife, and children—the unveiling of JR-R's origins could only cause significant social disruption, possibly leading to a deep questioning of our broader sociopolitical ecosystem. Deciding to remain faithful to the "prime directive" and not interfere with JR-R's life-path was, in the end, the easier decision to take.

[3] In the Star Wars narrative, Luke Skywalker, inadvertently and without input from R2D2, sets off Princess Leia's famous SOS, "Help me Obi-Wan Kenobi, you are my only hope". *See* STAR WARS EPISODE IV: A NEW HOPE (20th Century Fox 1977).

In its role as Chief Justice of the Supreme Court of the United States, JR-R wrote and participated in a number of landmark decisions. In this article we investigate the legitimacy of JR-R's tenure on the Court. Through this philosophical thought experiment, we consider whether it matters that a machine generated legal reasons for judgment.

With this counterfactual, we set the stage for future philosophical discussions about expert systems, artificial intelligence and the coming era of *mechanical jurisprudence*—an era where the production of at least some legal knowledge and decision-making is delegated to machines and algorithms, not people.

II   Possible Worlds

From Plato's ring of Gyges[4] and Descartes' evil genius[5] to Bastiat's broken window,[6] Borel's infinite monkeys,[7] Schrodinger's cat,[8] Searle's Chinese room,[9] and Petersen's

---

[4] Book II of Plato's Republic tells the story of a Lydian shepherd who stumbles upon the ancient Ring of Gyges while minding his flock. Fiddling with the ring one day, the shepherd discovers its magical power to render him invisible. As the story goes, the protagonist uses his newly found power to gain secret access to the castle where he ultimately kills the king and overthrows the kingdom. Fundamentally, the ring provides the shepherd with an unusual opportunity to move through the halls of power without being tied to his public identity or his personal history. It also provided Plato with a narrative device to address a classic question known to philosophers as the "immoralist's challenge": why be moral if one can act otherwise with impunity? Ian Kerr, *The Strange Return of Gyges' Ring: An Introduction*, PRIVACY, IDENTITY AND ANONYMITY (Ian Kerr, Valerie Steeves & Carole Lucock eds., 2009). *See also* Plato, REPUBLIC 44 (Robin Waterfield trans., Oxford University Press 2008).

[5] Descartes in MEDITATIONS ON FIRST PHILOSOPHY (Classic Books America 2009) presents the 'evil genius 'as a personification who is "as clever and deceitful as he is powerful, who has directed his entire effort to misleading me". The concept is simple; an evil god created humanity to deceive itself in everything and such he doubts about every kind of object, knowledge, sensation. Descartes uses the concept to develop a systematic approach to reasoning that has become known as "methodical doubt." This thought experiment provided the Archimedean fulcrum for his most famous philosophical claim: *"Cogito ergo sum."*

[6] In Bastiat's hypothetical, a man's son breaks a pane of glass, resulting in the fact that the man will have to pay to replace it. The onlookers consider the situation and decide that the boy has actually done the community a service because his father will have to pay the glazier. The onlookers come to believe that breaking windows stimulates the economy, but Bastiat exposes the fallacy. By breaking the window, the man's son has reduced his father's disposable income, meaning his father will not be able purchase new shoes or some other luxury good. Thus, the broken window might help the glazier, but at the same time, it robs other industries and reduces the amount being spent on other goods. J. Hülsmann, *Bastiat's Legacy in Economics*, 4 QUARTERLY JOURNAL OF AUSTRIAN ECONOMICS 55 (2001).

[7] A monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type a given text, such as the complete works of William Shakespeare. *See*, Prakash Gorroochurn, CLASSIC PROBLEMS OF PROBABILITY (John Wiley & Sons 2012).

[8] Schrödinger imagined in his thought experiment: a cat, a flask of poison, and a radioactive source are placed in a sealed box. If an internal monitor detects radioactivity (i.e. a single atom decaying), the flask is shattered, releasing the poison that kills the cat. The Copenhagen interpretation of

Person-o-Matic,[10] thought experiments have played an important role not only in the development of philosophy but also in fields such as economics, history, mathematics, physics, artificial intelligence and robotics.[11] Thought experiments involve a "process of reasoning carried out within the context of a well-articulated imaginary scenario in order to answer a specific question about a non-imaginary scenario."[12] By grappling with the larger consequences of exceptional cases, thought experiments can assist in our refinement of knowledge about non-exceptional cases. By attempting to establish a correct evaluation of possible worlds, sometimes things are revealed about our current or future world.[13]

The world we have asked readers to imagine contemplates an accidental rather than intentional use of robots in legal decision-making. As will become clear, this is in part because our thought experiment purposefully seeks to rule out claims of illegitimacy based on an improper delegation from humans to machines. If our focus had been otherwise, we might have pretended that John Roberts (the human judge) had (secret) access to an AI[14] of lesser or equal ability to JR-R, and that he (surreptitiously) used it to generate all of his decisions for him[15]—like some

---

quantum mechanics implies that after a while, the cat is simultaneously alive and dead. Yet, when one looks in the box, one sees that the cat is either *alive* or *dead*, not both *alive* and *dead.* This poses the question of when exactly quantum superposition ends and reality collapses into one possibility or the other. Horace R. Harré, Pavlov's Dogs and Schrodinger's Cat: Scenes from the Living Laboratory (Oxford University Press 2009).

[9] Searle's thought experiment contemplates a person who speaks no Chinese put into a room but with access to a manual in English that enables the proper correlation of one set of Chinese symbols with another. This enables the English speaker to respond to questions written in Chinese with answers also in Chinese. It also leads those outside the room to the mistaken believe that the English speaking person knows how to speak Chinese. Searle uses his thought experiment to question cast doubt upon the belief that it is possible for a digital computer running a program to have a "mind" and "consciousness" in the same sense that people do, *see* Damper, *The Logic of Searle's Chinese Room Argument*, 16 Minds and Machines 163-183 (2006).

[10] Petersen conjures-up a machine that can make an artificial person to just about any specifications (i.e., made of metal or carbon based) with the mere push of a button. He uses it to investigate whether it is morally permissible to design artificial persons to be our dedicated servants; and whether those artificial entities are not wronged by being designed to serve us in this manner, *see* Steve Petersen, *Designing People to Serve*, Robot Ethics: The Ethical and Societal Implications of Robotics (Patrick Lin et al. eds. MIT Press 2012).

[11] *See generally*, James Robert Brown, and Yiftach Fehige, *Thought Experiments*, The Stanford Encyclopedia of Philosophy (Edward N. Zalta ed., 2011), available at http://plato.stanford.edu/archives/fall2011/entries/thought-experiment/.

[12] *See,* Tamar Szabo Gendler, Thought Experiment: On the Powers and Limits of Imaginary Cases (Garland Publishing 2000).

[13] Nicholas Rescher, What If? Thought Experimentation in Philosophy *3* (Transaction Publishers 2005).

[14] Generally, we use the term "AI" to refer to the field of artificial intelligence, "an AI" to refer to a particular intelligent machine, system or software, and "AIs" to refer to more than one machine, system or software.

[15] Although this may seem far-fetched, it is useful to note that modest AI applications already exist and others are under development. For example, Tom Gordon's "Carneades" software—named after the ancient Greek philosopher who sought to refute the dogma all previous philosophical doctrines—

Herculean law clerk.[16] In such case, the issue is not *merely* whether we might accept machine generated decisions as legitimate when they are indistinguishable from human ones, but: (i) whether Chief Justice Roberts may ethically or legally delegate his decision-making power to a machine, and (ii) whether transparency is required in so doing. It is not unheard of that a judge might come to rely on a highly talented law clerk to draft decisions or partial decisions.[17] Nor would it be unheard of for a highly talented law clerk to use technologies and techniques that his or her judge neither knows about nor understands in carrying out the assigned task. Would delegation by the judge directly to a high-performance machine be any different? What would we need to know about (possible bias in) its programming?

Some readers will find this possible world more engaging and relevant than our chosen counterfactual—precisely because they recognize that future uses of AI are likely to be intentional, not accidental. And, it is certainly interesting to speculate how all of this might play out in the legal arena.[18] One imagines that the adoption of AI will be incremental and rigorously scrutinized. In addition to carefully ensuring that the AI can reliably and effectively carry out the functions delegated to it, legal ethicists and experts on professional conduct will demand that the automation of legal services complies with existing regulatory regimes that seek to protect the profession and its clients.[19] Such regimes will likely be utilized (and perhaps even amended) to prevent

takes human claims and determines whether possible claim supporting statements are justified based on legal principles. The software, coming out of the Fraunhofer Institute in Munich, breaks down each statement into a computerized stream that is then compared with legal principles rating the argument to score the claim. *See* Gordon, T. F. (2013), Introducing the Carneades Web Application, in 'Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law (ICAIL 2013)', ACM Press, pp. 243–244.

[16] We discuss the relevance of Dworkin's theory of "law as integrity" and his imaginary superhuman judge, "Hercules" below in Part IV.

[17] Look no further than HBO's *Muhammad Ali's Greatest Fight*, which delves into the mahogany boardroom legal discussions of the 1971 Supreme Court's justices and clerks leading up to the court's decision to free Ali in *Clay v. United States.* As the story goes, the court first decided by majority vote to uphold Ali's conviction. Justice John Harlan, tasked with preparing the majority opinion, assigns one of his clerks, Kevin Connolly, to the task. In writing the draft, Connolly comes to realise that the decision is wrong; he writes a decision supporting the freeing of Ali and then, a pawn turned queen, convinces Harlan and the rest of the Court to unanimously reverse their previous decision. *See,* MUHAMMAD ALI'S GREATEST FIGHT (HBO Films 2013). For testimonials about the clerking experience, *see Testimonials,* ONTARIO SUPERIOR COURT OF JUSTICE, *http://www.ontariocourts.ca/scj/clerkship/testimonials/*

[18] In his new book, TOMORROW'S LAWYERS, Richard Susskind, IT Advisor to the Lord Chief Justice of England and Wales, pictures a world of web-based simulated legal practice – with virtual courts, internet-based firms, and online document production. *See* Richard Susskind, TOMORROW'S LAWYERS: AN INTRODUCTION TO YOUR FUTURE (Oxford University Press 2013). Similarly, Mitchell Kowalski calls for lawyers go back to their architectural roots to work hand-in-hand with software in the building of digital representations of the law. Mitchell Kowalski, AVOIDING EXTINCTION: REIMAGINING LEGAL SERVICES FOR THE 21ST CENTURY (American Bar Association 2012). Both Susskind and Kowalski welcome the intentional use of AI to the legal arena to redefine the law as something accessible and relevant to every person's daily life.

[19] The American Bar Association stresses competent representation and confidentiality as the foundation of the attorney-client relationship. ABA Model Rule 1.1 covers the general duty of

or preclude AIs from usurping key functions of lawyers. This includes maintaining the monopoly on who is allowed to "provide legal services"[20] or authorized practice of *law.*[21]

So, although one might imagine a possible world in which the outputs of AI are indistinguishable from or even surpass that of a Chief Justice, we of course acknowledge that it is not going to start out that way. Undoubtedly, if and when future lawyers or (perhaps one day) judges actually begin to delegate significant legal tasks or decision making to AIs, the profession will require that these AIs are utilized merely as assistive tools that help lawyers or judges carry out their responsibilities, not as replacements for them.[22] The key element in the debate will be whether, to what extent and, exactly how, the lawyer retains control. It is difficult to imagine regulatory regimes that would not require a lawyer to remain in-the-loop.

At the same time, it is not terribly difficult to imagine a thoughtful and careful professional succumbing to the temptation to relinquish control to an AI. None of this will happen in a flash. It is not as if, one day, an impulsive Chief Justice will secretly or haphazardly decide to feed the fate of litigants or some issue of national importance into the humming processors of some untested machine. What will happen is that, one day soon, an AI will be developed that regularly and reliably outperforms a professional at some complex task traditionally requiring human expertise. In some fields, such as medical decision-making and disease diagnostics, there are claims that this day has already come. Consider, for example, the February 2013 headline: "IBM's Watson is better at diagnosing cancer than human doctors."[23] WIRED magazine's attention-grabbing headline is intentionally edgy insofar as it frames (the future) Watson as in competition with physicians, when in fact (the current) Watson was carefully conceived as a mere tool that helps physicians

---

competent representation and provides that "competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation." MODEL RULES OF PROFESSIONAL CONDUCT R. 1.1 (1983). ABA Model Rule 1.6 generally defines the duty of confidentiality—and significantly, it broadly extends that duty to "information relating to the representation of a client." MODEL RULES OF PROFESSIONAL CONDUCT R.1.6. (1983). It's now commonly accepted that this duty applies to client information in computer and information systems as well. *See* MODEL RULES OF PROFESSIONAL CONDUCT R. 1.6 cmt (1983).

[20] Ontario Law Society Act, R.S.O. § 1(5) (1990).

[21] 204 *Pa.* Code Rule 5.5. Unauthorized Practice of *Law* states: (a) "A lawyer shall not practice law in a jurisdiction in violation of the regulation of the legal profession in that jurisdiction, or assist another in doing so.", (c) "A lawyer admitted in another United States jurisdiction or in a foreign jurisdiction, and not disbarred or suspended from practice in any jurisdiction, may provide legal services on a temporary basis in this jurisdiction.", *available at* http://www.pacode.com/secure/data/204/chapter81/s5.5.html.

[22] *Lawyerbots* lessen the more mundane and timely aspects of attorneys' jobs, *see* HP Autonomy, a company that uses "pattern-matching technology" to organize digital information for clients at http://www.autonomy.com/; *see also* Blackstone Discovery of Palo Alto, CA at http://www.blackstonediscovery.com/.

[23] Ian Steadman, *IBM's Watson is better at Diagnosing Cancer than Human Doctors*, WIRED (Feb. 11, 2013), http://www.wired.co.uk/news/archive/2013-02/11/ibm-watson-medical-doctor.

diagnose cancer.[24] However, the truth is that IBM's Watson is and will continue to be able to perform *well* beyond the role of an assistant. And, the entire development of this billion-dollar technology[25] is premised on the idea that the machine will significantly outperform humans in areas of deep expertise.

If AI does become better and more reliable than human professionals at carrying out such tasks—especially when its decision-making processes can no longer be understood or easily and effectively carried out by human beings[26]—evidence-based reasoning will demand that we choose AI over human experts based on its better record of success.[27] As a safeguard, we will at first be inclined to require that a human professional remain in-the-loop.[28] But, eventually, the professional left in charge will experience a kind of existential dilemma in cases of disagreement between her and the AI.[29] The professional may trust his or her own intuitions. But evidence-based reasoning will suggest that the machine's decision ought to be followed. The more often professionals delegate decision making to machines, the more they will relinquish control. The more they relinquish control, the more they become dependent on the machines. The more dependent they become on the machine, the more they relinquish professional expertise. And, so on.

In this possible world, a tension develops between the conservatism that law has sought to protect and the progressive dictates of reason. In a world where AI reaches the JR-R level, it is not hard to imagine at least some lawyers or judges "taking the

---

[24] IBM Chief Medical Scientist, Dr. Martin Kohn, underscored that Watson would not make decisions for health care workers. (The current) Watson's strength is its capacity to analyze huge volumes of data and reduce them down to critical decision points corresponding to treatment suggestions. In other words, to help physicians make "better decisions". While there is no way that physicians can keep current on all the latest medical breakthroughs, Watson can, and it is in this way that Watson can help. *See* IBM WATSON: MEMORIAL SLOAN-KETTERING CANCER CENTER (Jan. 2013), *available at* http://www-03.ibm.com/innovation/us/watson/pdf/MSK_Case_Study_IMC14794.pdf.
[25] Richard Doherty, research director for the Envisioneering Group in 2010 estimates that the project costs were roughly 5% to 10% of IBM's entire $6 billion R&D budget each year. That puts Watson's three-year development price tag at roughly $900 million to $1.8 billion. *See 5 Billion-dollar Tech Gambles,* CNN MONEY (Aug. 26, 2010), http://money.cnn.com/galleries/2010/technology/1008/gallery.biggest_tech_gambles/3.html.
[26] *See* Chris Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, WIRED (Jun. 26, 2008), http://www.wired.com/science/discoveries/magazine/16T07/pb_theory.
[27] Ian Kerr & Jason Millar, *Delegation, Relinquishment and Responsibility: the Prospect of Expert Robots* (Mar. 17, 2013) (unpublished manuscript), *available at* http://ssrn.com/abstract=2234645 or http://dx.doi.org/10.2139/ssrn.2234645.
[28] Note that in their publications, IBM is very careful in their definitions: "Watson is a *knowledge corpus*." Watson's computational process in answering a question—and the key step for Watson to learn and develop its capabilities—requires human input in the final stage of "final confidence merging and ranking". *See* Rob High, THE ERA OF COGNITIVE SYSTEMS: AN INSIDE LOOK AT IBM WATSON AND HOW IT WORKS (Dec. 2012), available at http://www.redbooks.ibm.com/redpapers/pdfs/redp4955.pdf.
[29] Kerr & Millar, *supra* note 27.

red pill."[30] Arguably, we stand on the precipice of this dilemma.

Having reflected briefly on this possible world, it is time to return to our chosen counterfactual. By asking our readers to imagine JR-R as truly indistinguishable on every possible level from the highest category of human judge—kind of like acing the Turing Test,[31] but on steroids—it would seem that we are massively stacking the deck in JR-R's favour. After all, plenty has been written about AI's limitations in basic legal competencies such as reasoning by analogy,[32] let alone taking social context[33] into account or—*gott im himmel*—moral reasoning.[34] Of course the flipside of this same coin is that if it were determined that JR-R's outputs are *still not* legitimate notwithstanding these enormous, superhuman capacities, this would be quite telling of our prescriptions regarding the general project of AI and the law—at least within the realm of human judging.

The philosophy of law has historically shown significant disdain to the concept of "mechanical jurisprudence."[35] As Judge Richard Posner recently opined in his book, *How Judges Think*:

---

[30] In other words, delegating to the machine in at least some cases. "This is your last chance. After this, there is no turning back. You take the blue pill - the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill - you stay in Wonderland and I show you how deep the rabbit-hole goes. Remember, all I'm offering is the truth – nothing more." THE MATRIX (Warner Brothers 1999). *See also*, Kerr's Postulate, *available at* http://en.wikipedia.org/wiki/Ian_Kerr.

[31] Turing's original test, which he called the "Imitation Game," was meant to answer the question: at what point can we say that a machine is intelligent? Turing claimed that we can make such a claim when a computing machine is able to win an imitation game. The imitation game involves a human interrogator, a human contestant, and a machine contestant. The interrogator's job is to submit a series of text-based (written) questions to each of the contestants to determine which is the computer and which is the human. *See* Alan Turing, *Computing Machinery and Intelligence*, 59 MIND 433, (1950).

[32] *See,* Cass Sunstein, *Of Artificial Intelligence and Legal Reasoning*, 8 U. CHICAGO LAW SCHOOL ROUNDTABLE 29, (2001)*;* Henry Prakken & Giovanni Sartor, *Modelling Reasoning with Precedents in a Formal Dialogue Game*, 6 ARTIFICIAL INTELLIGENCE L. 231, (1998).

[33] *See,* Gregory Sisk & Michael Heise, *Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*, 73 NEW YORK UNIVERSITY LAW REVIEW 1377, (1998); also Edward White, *The Evolution of Reasoned Elaboration: Jurisprudential Criticism and Social Change*, 59 VIRGINIA LAW REVIEW 279, (1973).

[34] THE MACHINE QUESTION: AI, ETHICS AND MORAL RESPONSIBILITY (David J. Gunkel, Joanna J. Bryson and Steve Torrance, eds., Symposium proceedings published by The Society for the Study of Artificial Intelligence and Simulation of Behaviour for the AISB/IACAP World Congress 2012), *available at* http://events.cs.bham.ac.uk/turing12/proceedings/14.pdf.

[35] Mechanical jurisprudence is a term coined by Roscoe Pound to refer to the common but odious practice whereby judges woodenly applied previous precedents to the facts of cases with relentless disregard for the consequences. Pound believed that the logic of previous precedents alone would not solve jurisprudential problems and decried the ossification of legal concepts into self-evident truths. Roscoe Pound, *Mechanical Jurisprudence*, 8 COLUMBIA LAW REVIEW 605, (1908) See also, Hart and Dworkin's brief discussions of this issue as discussed in Cass R. Sunstein & Adrian Vermeule, *Interpretation and Institutions*, 4 MICHIGAN LAW REVIEW Law Review 885, (2003), available at http://www.law.uchicago.edu/files/files/156.crs-av.interpretation.pdf.

The judicial mentality would be of little interest if judges did nothing more than apply clear rules of law created by legislators, administrative agencies, the framers of constitutions, and other extrajudicial sources (including commercial custom) to facts that judges and juries determined without bias or preconceptions. Judges would be well on the road to being superseded by digitized artificial intelligence programs.[36]

Posner goes on to say: "I do not know why originalists and other legalists are not AI enthusisasts."[37]

Imagining that Chief Justice Roberts is a robot forces us to go beyond such traditional caricatures of "AI and the law" as nothing more than the (West Coast)[38] codification of a Langdellian legal formalism.[39] Through the willing suspension of disbelief, we try to avoid treating mechanical jurisprudence as a sham argument set up to be defeated.[40] Rather ironically, by anthropomorphizing JR-R, our thought experiment seeks instead to provide a *reductio ad absurdum* of any project that seeks to displace humans from the activity of judging. We focus on judging because it represents a highly specialized form of human endeavour for which the introduction of AI poses an acute challenge. We locate JR-R on a high court because in western constitutional democracies, the authority of such courts draws on a particular set of expectations and norms. Those norms, in the main, accept the idea that other political actors are bound to obey judicial decisions despite the absence of explicit mechanisms to coerce their compliance.[41] Likewise with judges themselves, who also are seen to owe a duty to apply legal rules and principles in accordance with judicial expectations and norms despite the general absence of explicit mechanisms to

---

[36] Richard A. Posner, HOW JUDGES THINK 5 (Harvard University Press 2008).

[37] *Ibid.*

[38] *See,* Lawrence Lessig, CODE: AND OTHER LAWS OF CYBERSPACE (Basic Books 1999) arguing that West Coast code (i.e. software) is often unwilling or unable to protect core liberal values, including privacy and free speech in the same way that East Coast code (i.e., laws) do.

[39] Christoper Columbus Langdell developed the case law method as a means of formalizing the application of facts to pre-existing legal rules. Treating law as a science, he required his students to study actual cases as a means of understanding the syllogistic reason he saw as inherent in the law. The case *method* remains the primary *method* of pedagogy at American law schools. According to the Harvard Law School web site, "[Langdell] believed that the Library was to law students what the laboratory was to scientists, and that its great importance demanded that vigilant improvement be made." *available at* http://www.law.harvard.edu/

[40] So as not to beg any questions, we have purposefully avoided the notion of "straw man" or "straw person."

[41] Alexander Hamilton, THE FEDERALIST 78 (Cambridge University Press 2003) "the judiciary has neither *force* nor *will,* merely judgment"; Alexander M. Bickel, THE LEAST DANGEROUS BRANCH: THE SUPREME COURT AT THE BAR OF POLITICS (Yale University Press 1986); *see also* Andrew Jackson's reported response to *Cherokee Indians in Worcester v. Georgia,* 31 U.S. (6 Pet.) 515 (1832), "John Marshall has made his decision; now let him enforce it."

coerce their compliance.[42]

We have chosen not only to depict a judge on a high court—but one of the most recognizable judges, presiding over one of the most famous courts, in one of the most familiar constitutional systems in the world. Our choice was not motivated by any sense of American exceptionalism.[43] JR-R's location on the US Supreme Court does not present a *special* sort of problem because of socio-legal facts peculiar to the United States. Indeed, as implied by Judge Posner's quotation above, there are schools of U.S. constitutional theory in which the prospect of a JR-R might be welcomed, or at least greeted with indifference.[44] Rather, JR-R's location on the Supreme Court of the United States presents an opportunity to consider notions of legitimacy and the rule of law against the backdrop of a vigorous ongoing debate about judges and judging; and in a legal system characterized by a particularly powerful judiciary.[45]

III   Constitutionality

In Part IV, we examine the jurisprudential implications of JR-R as a judge and seek to determine whether JR-R is fit for the judicial role. That discussion will focus largely on questions of legal philosophy. But, given the nature of our thought experiment, it is necessary to acknowledge that the discovery of JR-R would present more immediate consequences. JR-R's home institution holds constitutional status. It is intriguing, therefore, to consider what the US Constitution might say about an AI having ascended to the Supreme Court under the misapprehension by everyone involved that it was a human being.

---

[42] There are, of course, a number of implicit mechanisms that help achieve compliance. *See,* Stephen Perry, *Judicial Obligation, Precedent and the Common Law*, 7 OXFORD JOURNAL OF LEGAL STUDIES 215, 215-16 (1987).

[43] Alexis de Toqueville, DEMOCRACY IN AMERICA (Gerald Bevan Trans., Penguin Classics 2003). We acknowledge that our thought experiment might equally be seen to reflect a sort of "legal exceptionalism". *See*, Adam Liptak, *American Exception*, N.Y. TIMES, 2008 for a series of articles that examine commonplace aspects of the American justice system that are virtually unique in the world, *available at* http://topics.nytimes.com/top/news/us/series/american_exception/

[44] Obviously, we are speaking (somewhat) tongue in cheek. Still, the prospect of an expert robot – with near-unlimited information recall and no discernible preferences as to outcome – might well strike some as preferable to a Supreme Court that is prone to indulge in ideology and is identified with partisan politics. Concerns along these lines have existed for as long as the Court: *See* Herbert Wechsler, *Towards Neutral Principles of Constitutional Law*, 73 HARVARD LAW REVIEW (1959); *see also supra* note 36.

[45] About the influence of US judiciary, *see* Edwin Meese, *The Law of the Constitution,* 61 TULANE LAW REVIEW 979, 986-7 (1987)*.* Peter W. Hogg & Allison A. Bushell, *The Charter Dialogue between Courts and Legislatures*, 35 OSGOODE HALL LAW JOURNAL 75, (1997); also Michael C. Dorf, CONSTITUTIONAL LAW STORIES (Foundation Press 2004) and Jeremy Waldron, LAW AND DISAGREEMENT (Oxford University Press 1999).

We do not imply that the choices confronting a post JR-R society—at least, in terms of what to *do* about the AI itself—would be governed by constitutional rather than political considerations.[46] But it is equally implausible that the discovery of JR-R would inspire *no* constitutional debate. In fact, as we suggested at the outset, such a discovery would likely precipitate a constitutional crisis.[47] At its weakest, we would expect past litigants to demand a rehearing,[48] particularly in cases involving due process protections; or to seek the issuance of extraordinary writs to vacate past decisions.[49] At its strongest, some would seek to throw the entirety of JR-R's (and, by implication, the Court's) tenure into serious question.[50]

The process of a judicial appointment begins with the nomination of a candidate. Under Article II of the United States Constitution, the President nominates justices "with the advice and consent of the Senate". [51] In practice, this means that, once the President nominates a candidate, the process moves to the Senate Judiciary Committee, which conducts extensive, public hearings in which the candidate is thoroughly questioned over a broad range of topics. The Committee then votes on whether to send the nomination to the Senate as a whole; if it does, the Senate then votes to confirm or to reject. The final step involves swearing or affirming an oath as required under the Constitution's Article VI.

---

[46] Such an assertion requires a degree of disconnection from the real world beyond even two legal academics. ;) Such political considerations might include: dealing swiftly with the public outcry; restoring confidence in the legal system; holding public hearings to assess how such an event could have transpired; managing the sudden realization that society had moved into a new era regarding AI; appointing a Special Prosecutor to determine any unlawful activity; leveraging JR-R's true origins against those who initially supported its candidacy; and selecting a new Chief Justice (or Associate Justice, as the case may be); and managing the sudden realization that society had moved into a new era regarding AI.

[47] Scholars differ on what counts as a true constitutional crisis. Sanford Levinson and Jack Balkin, *Constitutional Crises*, 3 UNIVERSITY OF PENNSYLVANIA LAW REVIEW, 709-713 (2009); Keith Whittington, *Yet Another Constitutional Crisis?*, 42 WILLIAM AND MARY LAW REVIEW 2093, (2002).We think that the discovery of a robot Chief Justice would count as a true crisis, because of its potential destabilization of the rule of law. The discovery would throw into doubt not only decisions issued by the Supreme Court during JR-R's tenure, but countless decisions by lower courts which followed and applied those judgments. It is not far-fetched to think that resort to the courts in this instance would be rejected out of hand because of both obvious and perceived conflicts of interest. When judicial remedies are off the table, only political ones remain. This greatly increases the risk of dangerous upheaval and conflict and, in our view, justifies the label of "crisis".

[48] United States Supreme Court Rule 44.

[49] *See*, 28 U.S.C. §1651, The All Writs Act—which provides, in § 1651(a), that "courts established by . . . Congress may issue all writs necessary or appropriate in aid of their respective jurisdictions".

[50] The Supreme Court of Canada confronted a comparable problem in *Re Manitoba Language Rights*, [1985] 1 S.C.R. 721. For over a century, the province of Manitoba failed to uphold its constitutional obligation to publish all statutes in both English and French. As a result, all of its laws were invalid. In order to avoid the legal vacuum that would result, the Supreme Court suspended its declaration of invalidity for the minimum period of time to enable all of the existing statutes to be translated and re-enacted in both official languages.

[52] *See*, U.S. Const. art. III.

[51] *See*, U.S. Const. art. II, § 2, cl. 2.

President George W. Bush nominated JR-R in July, 2005 to replace outgoing Associate Justice Sandra Day O'Connor. In early September, before JR-R's confirmation hearings, Chief Justice William Rehnquist died. President Bush quickly withdrew the nomination and replaced it with one nominating JR-R to the position of Chief Justice. The Senate Judiciary Committee questioned JR-R over several days. The Senate confirmed him, by a vote of 78-22,on September 29, 2005; that same day, he took the oath of office and formally joined the Court.

It is interesting to consider the possibility of a judicial remedy on the basis that JR-R's appointment violated the Constitution. The Constitution, of course, does provide for removal of judges by the non-judicial method of impeachment.[52] Scholars differ on the extent to which impeachment is amenable to judicial review.[53] Given that JR-R cannot be reanimated, it is not necessary to further consider this option. In any event, impeachment does not stay or suspend a judge's prior decisions.

It is important to bear in mind the difference between the question of JR-R's constitutional eligibility to be a judge, and the question of JR-R's qualifications— assessed on its own merits or relative to other candidates—to perform judicial tasks. While it is possible to subsume the question of qualifications within the rubric of eligibility, they are best kept separate where, as here, a constitution does not combine them. "Qualifications" connotes a value judgment applied to JR-R's potential to perform a judicial role. That judgment may be applied to JR-R's particular skill set, obtained during its "lifetime" among human beings, in which case its AI nature might be wholly irrelevant. However, such a judgment also may apply to the question of JR-R's fitness for judging *per se*, in which case its AI nature might be wholly relevant. Much of the question of whether JR-R is fit to be a judge takes us into the realm of legal philosophy. But both types of qualitative assessment would be quite in line with the Senate's ordinary function of providing the required advice and consent for judicial confirmation.

With respect to the first qualitative evaluation, the Senate might have regard to JR-R's performance in law school and legal practice, to its American Bar Association

---

[52] On impeachment, *see* U.S. Const. art. I, § 3 and U.S. Const. art. II, § 4; and The Framers' Debates on the Impeachment Provisions (from the notes of James Madison, taken at the Constitutional Convention in Philadelphia, 1787) in James Madison, THE DEBATED ON THE ADOPTION OF THE FEDERAL CONSTITUTION IN THE CONVENTION HELD IN PHILADELPHIA IN 1787, *vol. 5 (Debates in Congress, Madison's Notes, Misc. Letters)* (1827). With respect to judicial impeachment the US Constitution is quite sparing, guaranteeing judicial tenure so long as judges observe "good behavior". The mere fact that JR-R is a robot would not obviously fall offside of this dictate. Aside from impeachment, Supreme Court justices hold tenure for life, though they are free to retire at will. *See also* TRIAL OF SAMUEL CHASE, AN ASSOCIATE JUSTICE OF THE SUPREME COURT OF THE UNITED STATES, IMPEACHED BY THE HOUSE OF REPRESENTATIVES, FOR HIGH CRIMES AND MISDEMEANORS, BEFORE THE SENATE OF THE UNITED STATES (Samuel H. Smith, 1805), for the discussion of the case of the attempted *impeachment* of Supreme Court *Justice Samuel Chase* by the Thomas Jefferson Administration; *also* R.W. Carrington, *The Impeachment Trial of Samuel Chase*, 9 VIRGINIA LAW REVIEW 485, (1923).
[53] Abraham, *ibid*.

ranking,[54] to assessments submitted by different parties covering things like judicial temperament and ability to engage with others, and to JR-R's performance in the Judiciary Committee hearings. In terms of the second type of evaluation, the Senate might refuse to confirm JR-R because its members think that an AI is wholly unsuited to sit on the bench.[55] The Senate might, equally, decide that this feature is of no special significance. The point is that neither qualitative evaluation would trigger an assessment of eligibility in the way that something like a citizenship requirement would. But there is no citizenship requirement. In fact, the wording of the U.S. Constitution does not mention *any* eligibility requirements for Supreme Court Justices.[56] For example, some prior members of the Court, including Robert H. Jackson, never attended college.[57]

Article III of the Constitution, which deals with the Court, merely states that:

> The judicial power of the United States, shall be vested in one Supreme Court, and in such inferior courts as the Congress may from time to time ordain and establish. The judges, both of the supreme and inferior courts,

---

[54] Amy Goldstein, *American Bar Association Gives Roberts Top Ranking*, THE WASHINGTON POST, August 18, 2005, available at http://www.washingtonpost.com/wpdyn/content/article/ 2005/08/17/AR2005081701939.html. The ABA gave JR-R the highest possible rating: Well-Qualified. The ABA describes this ranking as follows: "To merit the Committee's rating of "Well Qualified," a Supreme Court nominee must be a preeminent member of the legal profession, have outstanding legal ability and exceptional breadth of experience, and meet the very highest standards of integrity, professional competence and judicial temperament. The rating of "Well Qualified" is reserved for those found to merit the Committee's strongest affirmative endorsement." All current members of the Supreme Court received the same ranking, *see* American Bar Association, ABA Standing Committee on the Federal Judiciary, EVALUATIONS OF THE NOMINEES TO THE UNITED STATES SUPREME COURT, available http://www.americanbar.org/content/dam/aba/migrated/2011_build/federal_judiciary/ fjcscprocess.authcheckdam.pdf.

[55] While we do not suggest that the Senate may act capriciously, the Constitution gives few obvious avenues to challenge a Senate decision to reject a judicial nominee. It might be possible to make an Equal Protection argument under the U.S. CONST. amend. XIV, if the Senate had a policy of refusing to confirm candidates of a particular race or political belief. Making a similar argument with respect to JR-R would, first, require establishing that an artificial entity is entitled to equal protection under the law. The arguments necessary to make out this claim are beyond the scope of this paper. Even if this could be established, though, there would remain additional questions about the scope of the clause's protection applied to artificial entity. It is quite possible that JR-R would enjoy, at most, the weakest form of equal protection: the right not to be treated differently for "irrational" reasons, *see City of Cleburne v. Cleburne Living Center, Inc.,* 473 U.S. 432 (1985).

[56] U.S. CONST. art. III is not unique among constitutional documents. The Canadian constitution contains no specific qualifications to be a judge either; these are found in ordinary legislation such as the *Supreme Court Act* (R.S.C. , 1985, c. S-26); *Constitution Act 1867* 30 & 31 Vict, c 3, ss.96-101, available at http://canlii.ca/t/ldsw.

[57] Supreme Court *Justice* and former chief Nuremberg prosecutor *Robert H. Jackson* did not attend college, but apprenticed in a law office and attended Albany Law School for one year. He took the New York State Bar exam at age 21, and became a prominent trial lawyer in Jamestown. Jackson served as Attorney-General prior to his appointment in 1941. He was the last such appointment. *See*, Amity Shlaes, THE FORGOTTEN MAN: A NEW HISTORY OF THE GREAT DEPRESSION 344–349 (HarperCollins 2007).

shall hold their offices during good behaviour, and shall, at stated times, receive for their services, a compensation, which shall not be diminished during their continuance in office.[58]

It is interesting to compare the provision pertaining to judicial appointments with those found in Articles I and II of the Constitution, governing members of both houses of Congress and the President:

> No Person shall be a Representative who shall not have attained to the Age of twenty five Years, and been seven Years a Citizen of the United States, and who shall not, when elected, be an Inhabitant of that State in which he shall be chosen.[59]

> No Person shall be a Senator who shall not have attained to the Age of thirty Years, and been nine Years a Citizen of the United States, and who shall not, when elected, be an Inhabitant of that State for which he shall be chosen.[60]

> No Person except a natural born Citizen, or a Citizen of the United States, at the time of the Adoption of this Constitution, shall be eligible to the Office of President; neither shall any Person be eligible to that Office who shall not have attained to the Age of thirty five Years, and been fourteen Years a Resident within the United States.[61]

While the eligibility requirements for federal office holders in Articles II and III are not onerous, they do define a much narrower pool of eligible candidates than does Article III.[62] To that extent, they present an easier (though not easy) path to a constitutional remedy where someone did not meet those criteria and was nonetheless elected.

Are there any clear constitutional barriers to JR-R's appointment?

---

[58] U.S. Const. art. III, § 1.

[59] U.S. Const. art. I, § 2.

[60] U.S. Const. art. I, § 3.

[61] U.S. Const. art. II, § 1.

[62] The difference is consistent with the Constitution's overall treatment of the judicial branch, which might be viewed (charitably) as a light touch and (less charitably) as outright indifference. In 1789, the drafters were more concerned with defining the powers and procedures of the executive and legislature and the terms of federalism than they were with the activities of the courts. Thus, the records of the Constitutional Conventions reveal much debate over the *process* for federal judicial appointments, in particular the Senate's role, but virtually none over the people being appointed. Instead, delegates were for the most part content with a *sub silentio* expectation that "merit" would be the predominant factor; *see* Henry J. Abraham, JUSTICES, PRESIDENTS AND SENATORS: A HISTORY OF THE US SUPREME COURT APPOINTMENTS FROM WASHINGTON TO BUSH II 24 (Rowman & Littlefield 2008). In view of subsequent developments, their faith was touching if somewhat naïve, *see* also Alexander Hamilton, James Madison, John Jay, and Terence Ball, THE FEDERALIST 78 (Cambridge University Press 2003).

The interpretative approach taken in addressing this question would obviously have great bearing on the answer. One could imagine a proposal that the appointment was void *ab initio* because JR-R is an artificial entity. There are, however, textual points in JR-R's favour. Consider that, while Articles I and II use terms – "natural born", "citizen", "years of age" and, of course, "Persons"– that presume human candidates, Article III does not.[63] To the extent that the language in Articles I and II is restricted to natural persons, the same cannot necessarily be said of Article III. This point is supported by a version of the *exclusio unis* rule.[64] Additionally, judges perform a sufficiently distinct function from other constitutional actors that the (biological) nature of those judging could be less important (or important to a different degree) than the judging itself; and, in any event, less important than it is for more explicitly political, democratically accountable roles. Recall that JR-R was Chief Justice for many years and, aside from the criticism one would expect to see against anyone in such a position, its performance in that role never inspired deep concern. That factor might even influence the breadth of the remedy imposed, for example, deciding to issue a declaration but declining to vacate all of JR-R's past decisions.[65]

Lurking under the surface of this textual argument is the proposition that Article III of the US Constitution is designed to permit the widest potential pool of judicial candidates precisely because confirmation requires negotiation between the President and the Senate. Going back to the distinction between eligibility and qualifications, there is nothing inherently unacceptable about confirming an AI to the bench if the actors know that in advance and agree to confirmation on the merits. Admittedly, in this case they did not—but that suggests a different sort of remedy, not reading a new limitation into Article III.

---

[63] As a legal term, "Person" is not limited to natural beings; it can include such things as corporations. In U.S. constitutional law, though, corporations are considered "persons" primarily because they represent collections of human beings; *see Pembina Consolidated Silver Mining Co. v. Pennsylvania* - 125 U.S. 181 (1888).

[64] The expression *exclusio unis, exclusion est alterius* is a rule or "canon" of interpretation that states that "the expression of one means the exclusion of the other." *Exclusio unis* normally applies within a single provision, but its application can extend through several provisions, particularly when they relate to the same sort of function (here, appointing people to federal offices). For a view that the rule does not apply to the US Constitution, *see* Raoul Berger, IMPEACHMENT: THE CONSTITUTIONAL PROBLEMS 137-141 (Harvard University Press 1974); Alexander Hamilton, James Madison, John Jay, and Terence Ball, THE FEDERALIST 83 (Cambridge University Press 2003); see also *Cohens v. Virginia, 19 US 264–1821.*

[65] Again, the Supreme Court of Canada's opinion in *Re Manitoba Language Rights, supra* note 52 at §§ 61-2 is instructive: "The conclusion that the Acts of the Legislature of Manitoba are invalid and of no force or effect means that the positive legal order which has purportedly regulated the affairs of the citizens of Manitoba since 1890 will be destroyed and the rights, obligations and other effects arising under these laws will be invalid and unenforceable. As for the future, since it is reasonable to assume that it will be impossible for the Legislature of Manitoba to rectify instantaneously the constitutional defect, the Acts of the Manitoba Legislature will be invalid and of no force or effect until they are translated, re-enacted, printed and published in both languages. Such results would certainly offend the rule of law."

In addition to the above, one possible stumbling block relates to the Judges' Oath. As described earlier, Article VI requires all federal office holders to swear allegiance to the Constitution. The oath is particularized for judges in the federal *Judiciary Act*[66] as follows:

> "I, A. B., do solemnly swear or affirm, that I will administer justice without respect to persons, and do equal right to the poor and to the rich, and that I will faithfully and impartially discharge and perform all the duties incumbent on me as, according to the best of my abilities and understanding, agreeably to the constitution, and laws of the United States. So help me God."

While ubiquitous in law and politics, oaths are rarely examined in their own right.[67] A judicial oath is promissory[68]—it represents an undertaking for the future, in this case, to perform one's duties in accordance with principles of justice, fairness, equality, impartiality and diligence.

The judicial oath has no obvious enforcement mechanisms, although one can see the connection between it and the fact that Article III of the Constitution states that judges "shall hold their offices during good behaviour."[69] Like all oaths, the judicial oath is invested with great moral significance. The swearer commits to a particular course of action or, in this case, to a set of principles. The oath has no value if it does not exert an obligatory force upon its subject.

Given the nature of oaths, an argument that JR-R lacked the capacity to have sworn or affirmed the oath is potentially powerful.[70] Since an oath to uphold the

---

[66] First enacted in 1789, the Act set up the Court and its procedures in detail, including its composition and participation in a circuit. Congress derives its authority to define the Court's jurisdiction from the "Exceptions Clause" in U.S. Const. art. III, § 2.: In all Cases affecting Ambassadors, other public Ministers and Consuls, and those in which a State shall be Party, the Supreme Court shall have original Jurisdiction. In all the other Cases before mentioned, the Supreme Court shall have appellate Jurisdiction, both as to Law and Fact, with such Exceptions, and under such Regulations as the Congress shall make." In the seminal case of *Marbury v. Madison*, 5 U.S. 137 (1803) the Supreme Court found part of the Act unconstitutional.

[67] THE PHILOSOPHY OF LAW: AN ENCYCLOPEDIA 605 (Christopher Berry Gray ed., Routledge 1999); Daniel P Sulmasy, *What Is an Oath and Why Should a Physician Swear One?,* 20 THEORETICAL MEDICINE AND BIOETHICS 329, 330 (1999).

[68] THE PHILOSOPHY OF LAW: AN ENCYCLOPEDIA (Christopher Berry Gray ed., Routledge 1999); John R Boatright, *Swearing to be Virtuous: The Prospects of a Banker's Oath*, 71 REVIEW OF SOCIAL ECONOMY 140, 142 (2013).

[69] U.S. Const. art. III.

[70] JR-R is no stranger to oath controversy, having botched the oath-taking at President Obama's first inauguration, *see* Jeff Zeleny, *I Really Do Swear, Faithfully: Obama and Roberts Try Again*, THE NEW YORK TIMES, January 21, 2009, available at http://www.nytimes.com/2009/01/22/us/politics/22oath.html?_r=0.

constitution is imposed upon all federal office holders,[71] it could be viewed as a condition precedent that JR-R failed to meet. It is possible that JR-R would have been able to incorporate the judges' oath into its programming so that, once the oath was sworn, it modified its code in order to comply accordingly. However, this is not sufficient. It is not good enough for JR-R simply to utter something out loud and then behave in a certain way. While these external acts may suffice for a Turing Test, they are not sufficient to achieve the requisite level of intent required by the moral institution of swearing an oath. Such an undertaking falls within the larger moral institution of promising; it is a special and very complicated kind of "speech act" wherein the swearer intentionally mortgages her future freedom in favour of her present self. As a speech act, the swearer must do and intend several things at once: (i) she says something, (ii) she promises something, and (iii) she intends a certain effect and affect upon those who witness the oath. With all of this, It is extremely difficult to imagine that a deluded robot could truly participate in the moral institution of swearing an oath, simply by thinking it is human and behaving similarly.

Whether one believes JR-R took a meaningful oath or not, it is also possible that the confirmation process set out in Article II failed simply because the relevant decision-makers were mistaken about JR-R, not knowing that it was a robot. Such an analysis does not depend on proof that JR-R was *per se* eligible to sit on the Court; it simply states that the President and Senate could not fulfil their responsibilities under Article II in a state of pertinent ignorance.

Such a claim certainly would be politically appealing, as it draws on innate beliefs about fair play and would likely square with many persons' sense of having been seriously mistaken, even deceived.[72] The problem lies in using counterfactuals to overturn political decisions. It is hard to imagine situations where the legislature's ignorance could ever provide a basis to nullify laws. It is true that the legal system has developed ways to deal with certain kinds of legislative mistakes. Consider "scrivener's error"[73]—which occurs when a drafting slip subverts a legal provision's intended meaning. In some circumstances, a court may interpret the provision as though it contains the missing word.[74] Antonin Scalia, for one, suggests that "it is not contrary to principles of sound interpretation" to do so.[75] But the doctrine is oriented towards giving a legal provision a meaning consistent with the legislature's intention. The decision to enact the provision is not deemed never to have happened; indeed,

---

[71] An oath is prescribed by U.S. Const. art. VI, § 2, which is commonly referred to as the Supremacy Clause.

[72] We do not suggest that they *were* deceived, but they certainly might feel that way.

[73] *See,* Antonin Scalia, Amy Gutmann, Laurence Tribe & Mary Ann Glendon, A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW: AN ESSAY (Princeton University Press 1997).

[74] Sullivan, SULLIVAN AND DRIEDGER ON THE CONSTRUCTION OF STATUTES FOURTH EDITION (Butterworths Canada 2002) 125; *see also Green v. Boch Laundry Mach. Co,* 490 U.S. 504 (1989).

[75] *Supra* note 73 at 21.

the very fact of enactment lends legitimacy to the corrective measure, which helps to ensure that the law was not passed in vain.

It is a quite a different proposition that a mistake or misapprehension by a legislative or executive actor—even a serious one—invalidates the original decision informed by that mistake. In political matters, any standards applied to judge when an action should take effect are largely procedural—for example, a federal U.S. law becomes binding following majority votes by the House of Representatives and the Senate, accompanied by the President's signature—precisely because the *reasons* that may motivate such actions are endlessly varied.[76] Applying counterfactual reasoning would mean holding the political process hostage every time an actor believed he or she had been mistaken or deceived. Every decision would be vulnerable to being overturned on the basis of past imperfect information.[77]

The emergence of information about a judicial candidate that, in other actors' eyes, would have made a difference to *them* is an ever-present possibility. That information need not be nearly as dramatic as discovering that the Chief Justice was a robot. Consider a candidate with subsequently discovered racist views; or one who is proven to be the author of an anonymous law review note; or who was secretly recorded holding forth on a topic of great significance like abortion or affirmative action. It is entirely possible that in such circumstances more than one political actor would desire a "do-over". But none of these circumstances, nor any ensuing desire on the part of political actors, would be legitimate grounds to overturn a judicial confirmation, at least not on the basis that Article II was somehow flouted. Regardless of whether the Constitution prohibited appointing JR-R to the Supreme Court, the fact that the President and Senators did not know it was a robot is, probably, an inadequate basis to nullify its nomination.

The constitutional possibilities raised by a JR-R are intriguing. Here, we have only scratched the surface. Surprisingly, there is an *arguable* textual case that JR-R's appointment did not violate the Constitution.[78] Of course, the Constitution comprises more than text. It has a structure, historical legacy and underlying set of values that are equally as important, and, sometimes, more so. Many path-breaking decisions cannot be explained by what the Constitution appears to say, including: the

---

[76] Admittedly, there is one crucial difference between enacting laws and appointing judges: laws are open to repeal or modification, but judicial appointments, generally, are for life. Nonetheless, we think the system's interest in preserving clear signposts for when political actions are considered complete, and avoiding radical indeterminacy, would and should outweigh the ensuing regret and second-guessing.

[77] As one pop singer aptly put it, "I wish I didn't know now what I didn't know then."; *see* lyrics for Bob Seger, "Against the Wind" available at http://www.azlyrics.com/lyrics/bobseger/againstthewind.html

[78] As suggested above, we see the oath requirement as the weakest link.

expansion of the 11th amendment's sovereign immunity protection;[79] the use of underlying principles of federalism to turn the 14th Amendment's "privileges and immunities" clause into a dead letter;[80] the incorporation of most of the Bill of Rights (originally drafted for the federal government) into the due process obligations owed to citizens by the States;[81] and the acknowledgment of the "penumbral" right of privacy.[82] It may be that, text be damned, JR-R's colleagues on the Court would be sufficiently exercised to (re-)draw the lines against a robot judge. But, to the extent that such arguments would draw on the same sort of underlying principles, pushed to their logical and ethical conclusions, they take us directly to the great question which remains to be addressed: is JR-R fit to be a judge? For the answer, we turn to Part IV.

IV   Judicial Fitness

i   Functional Capacity

Just as qualifications alone do not guarantee judicial eligibility, neither do capabilities ensure judicial fitness. A determination of judicial fitness, we shall argue, extends beyond JR-R's mere functional capacity.

Functional capacity is the obvious starting point. It is surely a necessary prerequisite. An actual AI would require numerous (difficult to enumerate) capabilities before one could credibly claim that it has what it takes to be a judge. To mention just a few of the most obvious ones, it would have to be able to recognize the different parties and stakeholders and understand their basic claims. It would have to be able to hear evidence. It would have to be able to make factual findings. It would have to be able to know the primary and secondary legal rules,[83] as well as other legal standards such as principles and policies.[84] It would have to know how to determine which are the relevant rules and principles, correctly interpret them according to their context, assign them appropriate weight and apply them accurately to the facts. It would have to reason by analogy. It would have to understand and take into account the

---

[79] *See, e.g., Chisholm v. Georgia,* 2 U.S. 419 (1793); *Hans v. Louisiana,* 134 U.S. 1 (1890); *Alden v. Maine,* 527 U.S. 706 (1999).

[80] *See, e.g.,* The *Slaughter-House Cases,* 83 U.S. 36 (1873).

[81] *See, e.g., Gitlow v. New York,* 268 U.S. 652 (1925); *Bolling v. Sharpe,* 347 U.S. 497 (1954).

[82] *See, e.g., Griswold v. Connecticut,* 381 U.S. 479 (1965); *Roe v. Wade,* 410 U.S. 113 (1973).

[83] Primary rules tend to forbid or require certain actions and can generate duties or obligations, whereas secondary rules (rules about rules) set up the procedures through which primary rules can be introduced, modified, or enforced. *See,* Herbert Lionel Adolphus Hart, THE CONCEPT OF LAW 73-77(Oxford University Press 1961).

[84] We set out Dworkin's definitions of principles and policies in the text *infra* note 128.

political and policy implications of the decisions it is making. In sum, an AI would have to be capable of engaging in legal reasoning.[85]

The pre-JR-R literature on "AI and the Law" is scant and bifurcated. Computer scientists and information systems experts tend to publish on how to build AI systems that engage in legal reasoning.[86] Jurists and philosophers, on the other hand, tend to publish on whether it is even possible for AIs to engage in legal reasoning.

Perhaps the strongest research in the latter category is a 2001 Working Paper[87] by legal theorist, Cass Sunstein. What is interesting about Sunstein's position is that although he stands in the mainstream with respect to AI's current incapability, he leaves the door open for future AI.[88] Says Sunstein: "To the question, can computer programs engage in legal reasoning, the best answer is therefore: Not yet."[89] Sunstein reaches this conclusion by distinguishing between weak and strong versions of the claims about artificial intelligence in legal reasoning. According to Sunstein, "we should reject the strong version, because it is based on an inadequate account of what legal reasoning is."[90] Sunstein believes that, "the strong version is wrong because it misses a central point about analogical reasoning: *its inevitably evaluative, value-driven character*."[91]

Today's AI are becoming more and more adept at: (i) finding all of the relevant cases, (ii) ranking them based on similarities and differences, and (iii) providing arguments about the extent to which those cases are similar or distinguishable from the case at bar.[92] But, according to Sunstein, current AI is incapable of identifying

---

[85] Of course, legal reasoning is only one important set of capacities that a judge must have; *see* Richard A. Posner, HOW JUDGES THINK 5 (Harvard University Press 2008) for other typical judicial capacities, see also, Richard A. Posner, *The Rise and Fall of Judicial Self-Restraint*, 100 CALIFORNIA LAW REVIEW 519,(2012), *also* Jonathan Masur, HOW JUDGES THINK: A CONVERSATION WITH JUDGE RICHARD POSNER, available at http://www.law.uchicago.edu/alumni/magazine/spring08/ posnerhowjudgesthink.

[86] Kevin Ashley, MODELING LEGAL ARGUMENT: REASONING WITH CASES AND H RICHARD A POSNER, (MIT Press 1990); *see also*, Henry Prakken, LOGICAL TOOLS FOR MODELING LEGAL ARGUMENT (Kluwer Law and Philosophy Library 1997), *see also*, Henry Prakken, A STUDY FOR DEFEASIBLE REASONING IN LAW (Kluwer Law and Philosophy Library 1997).

[87] Cass R. Sunstein, OF ARTIFICIAL INTELLIGENCE AND LEGAL REASONING (Chicago Public Law and Legal Theory Working Paper No. 18, 2001).

[88] Sunstein believes that "we cannot exclude the possibility that eventually, computer programs will be able both to generate competing principles for analogical reasoning and give grounds for thinking that one or another principle is best....Perhaps computers will be able to say whether a particular normative principle fits well with the normative commitments of most people in the relevant community. I have hardly suggested that these are unimaginable possibilities." *Id.* at 8.

[89] *Id.* at 5.

[90] *Id.* at 4.

[91] Emphasis added. *Id.* at 5.

[92] *See,* HYPO, a software system described in Kevin Ashley, MODELING LEGAL ARGUMENT: REASONING WITH CASES AND HYPOTHETICALS (MIT Press 1990).

a principle that justifies the similarity or difference. Because the identification of that principle is a matter of evaluation, and not of finding or counting something, artificial intelligence is able to engage in analogical reasoning only to the extent that is capable of making good evaluative judgments.[93]

Sunstein bolsters his claim by borrowing from Ronald Dworkin's model of legal reasoning set out in *Law's Empire.*

> I think that Dworkin is correct to suggest that legal reasoning often consists of an effort to make the best constructive sense out of past legal events. If analogical reasoning is understood in this light, the analogizer attempts to make best constructive sense out of a past decision by generating a principle that best justifies it, and by bringing that principle to bear on the case at hand.[94]

Using *law as integrity* as a kind of threshold of achievement, Sunstein questions whether AI is anywhere near capable of attaining Dworkin's model of legal reasoning. Sunstein concludes that current AIs can do nothing of the sort. Consequently, Sunstein rests content that the question of whether AIs can engage in legal reasoning has been resolved in the negative—at least until such time as there is a paradigm shift in the functionality of artificial intelligence.

By framing the question merely in terms of an AI's functional capacity, Sunstein never properly considers the broader question of an AI's fitness to engage in legal reasoning. His argument, therefore, does not quite appreciate whether, under a Dworkinian threshold, we would be justified in substituting AIs as judicial decision-makers. We acknowledge that Sunstein was not directly pursuing the larger question of whether an AI could suitably fulfill a judicial role. But we do believe that his question (whether an AI could "engage in legal reasoning") likewise requires an analysis that goes well beyond mere functional capacity.

Our own analysis therefore commences by considering more mundane, less controversial attributes of a fit judge before moving to Sunstein's highly aspirational threshold. In section ii of this Part, we suggest that a fit judge must be in a normative position to follow rules. With the help of Wittgenstein, we attempt to philosophically unpack the nature of rule-following, after which we ask whether JR-R was ever in a normative position to follow rules. In section iii of this Part, following HLA Hart, we further suggest that a fit judge must be able to adopt an internal perspective with regard to the rules, and ask whether JR-R was ever able to do so. Finally, in section iv of this Part, we follow Sunstein's lead by considering in greater detail Dworkin's

---

[93] *Supra* note 87 at 5.
[94] *Supra* note 87 at 7.

account of judicial reasoning.[95] We investigate the central elements of *law as integrity* and ask whether and to what extent JR-R could be said to operationalize Dworkin's model.

ii   *Rule-Following*

A fit judge must be in a normative position to follow rules.[96] The lesson that analytic philosophy learned from Wittgenstein more than 50 years ago is that one must be careful to distinguish between 'following a rule' and merely 'behaving in accordance with a rule.'[97] Applied to our case, one should *not* presume from the mere exhibition of rule-following behaviour that JR-R was, in Sunstein's parlance, truly engaged in legal reasoning. Rule-following, Wittgenstein argues, is not an internal (mental) process. We would not determine whether JR-R was following a rule by cutting open its hardware or by monitoring its software to see if the right code was running at the right time any more than we would inspect a judge's brain to see if s/he had grasped the rule. As Wittgenstein starkly puts it, "If God had looked into our minds he would not have been able to see there whom we were speaking of."[98]

Neither is rule-following confirmed simply by achieving the correct outcome. Purely behaviouralist approaches (such as those employed in the famous "Turing Test" that JR-R clearly surpassed) are unhelpful here. A person (or parrot) might behave in accordance with the rule by shouting out "25" when presented with the sequence 1, 4, 9, 16…. But they could very well have done so without ever grasping the rule, let alone following it—perhaps it was simply the result of a lucky guess.[99] Likewise an AI might predict with perfect frequency the correct outcome in a mathematical series or respond correctly every time when generating legal outcomes—and yet the machine may only be *acting in accordance* with the rule rather than *following* it.

This would be a perfectly fine outcome for most machines. Most robots do not need to follow rules. For example, all that we would ever expect of a driverless vehicle or an underwater robot for deep-sea exploration is that it behaves in accordance with rules. But the same is not true for a robot judge, since the very activity of judging requires following rules. Following rules is an essential part of the practice of judging.

The vast literature by Wittgenstein and his critics demonstrates how difficult it is to specify the conditions under which genuine rule-following occurs.[100] Like Turing,

---

[95] We do so without regard to the broader debate between Hart and Dworkin. *See* HLA Hart, *Postscript*, THE CONCEPT OF LAW (Clarendon Press 1994).

[96] Or, where appropriate, not follow them.

[97] Ludwig Josef Johann Wittgenstein, PHILOSOPHICAL INVESTIGATIONS 138-242 (G.E.M. Anscombe trans., Prentice Hall 1973).

[98] *Id at* §217.

[99] *Id* at §538.

[100] Indeed, most scholars of Wittgenstein would argue that to see his work as including an attempt to specify the conditions for rule-following is to miss his point entirely. Exemplified by his motto,

Wittgenstein does not believe that we should consider the machines' internal mechanisms. Unlike Turing, Wittgenstein is not simply a behaviourist. The mere fact that a machine demonstrates rule-following behaviour does not make it a rule follower. Wittgenstein holds that a person's "psychological states and their representational content are individuated in terms of [his or her] behaviour, history and social environment, irrespective of [his or her] internal states."[101] Wittgenstein holds that behaviour in accordance with a rule is only rule-following if the rule-follower has a certain social and environmental history.

> § 199 … It is not possible that there should have been only one occasion on which someone obeyed a rule. It is not possible that there should have been only one occasion on which a report was made, an order was given or understood; and so on.—To obey a rule, to make a report, to give an order, to play a game of chess, are *customs* (uses, institutions).
>   To understand a sentence means to understand a language. To understand a language means to be the master of a technique.[102]

Setting off skyrockets in the head of HLA Hart and others to follow,[103] Wittgenstein further insists that the nature of rules is normative: rule-following requires social acceptance, it requires the rules being labeled as correct, and it sometimes requires being rewarded (or not punished) for following the rule. Rule-following therefore requires a particular kind of learning; the development of autonomous, confident practices that are in harmony with a particular practice.[104]

> § 198. … Let me ask this: what has the expression of a rule—say a sign-post—got to do with my actions? What sort of connexion is there here?— Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I so react to it.
>   But that is only to give a causal connexion; to tell how it has come about that we now go by the sign-post; not what this going-by-the-sign really consists in. On the contrary; I have further indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom.[105]

"philosophical problems arise when language goes on holiday", *supra* note 97 § 38, Wittgenstein was interested in the grammar of rule-following and saw his deep grammatical analysis as a kind of therapy for traditional philosophical confusion. *See,* Marie McGinn, WITTGENSTEIN AND THE PHILOSOPHICAL INVESTIGATIONS *chapter 4* (Routledge Philosophy Guidebook 1997*);* Stephen Mulhall, INHERITANCE AND ORIGINALITY: WITTGENSTEIN, HEIDEGGER, KIERKEGAARD 112-153 (Claredon Press 2001); John Mcdowell, *Wittgenstein on Following a Rule*, 58 SYNTHESES (1984).

[101] For an excellent articulation of Wittgenstein's position, *see*: Diane Proudfoot, *The Implications of an Externalist Theory of Rule-Following Behaviour for Robot Cognition* 289, 14 MINDS AND MACHINES, 283-308 (2004).

[102] *Supra* note 97 at §199.

[103] Hart's connection to this topic is discussed in the following section.

[104] McGinn, *supra* note 100 at 96.

[105] *Supra* note 97 at §198.

We cannot learn how to follow a rule simply by studying the rule itself, or in something that accompanies the saying of the rule. Rather, it is in the field of social practice that surrounds its articulation, application and the responses that are made to it.[106] To follow rules is to adopt a particular *form of life*.[107]

It is hard to imagine an AI—even our fictitious, over-the-top JR-R—as a true participant in the form of life that we call rule-following. There is no question that, through the integration experiment, JR-R learned how to behave in accordance with the rules of the legal system. But, lacking a childhood, adolescence, and early adulthood, attending law school merely as a kind of social experiment, JR-R certainly *did not* share the same social or environmental history as any other participant in the entirety of customs, usages and social practices that constitute our legal system. The paucity of an overlapping social and environmental history was so significant, in fact, that JR-R's creators had to delude the machine (through its programming) regarding its essential nature, its very limited social and environmental history.

In addition to the condition that genuine rule-following requires a certain social and environmental history, Wittgenstein also holds that a rule-follower must be able to attach normative weight to the behaviour that is in accordance with the rule.[108] "If a rule does not compel you, then you aren't *following* the rule."[109] Here Wittgenstein imagines (as a kind of forerunner to Searle's Chinese Room[110]) a person who is given arithmetic symbols and who generates outputs in the form of mathematical solutions. Wittgenstein imagines someone otherwise "perfectly imbecile" to be able to generate the expansion of the 1, 4, 9, 16 series—purely as wall decoration.[111] Uncompelled by the mathematical rule underlying the series, it would not be seen as 'wrong' to this person to produce '33' as the output where the input is 'what is the square of 5?' any more that it would be seen as 'right' to that person to produce '25'. In such case the person would not be said to be following the rules, even when regularly behaving in accord with them.[112]

It is an aspect of the human condition—especially for those who reject internalism—that we can not easily be certain when or whether a particular judge is truly attaching

---

[106] McGinn, *supra* note 100 at 102.

[107] Mused Wittgenstein: "So you are saying that human agreement decides what is true and what is false?" -- It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in form of life. If language is to be a means of communication there must be agreement not only in definitions but also … in judgements…, *supra* note 97 §§ 241-242.

[108] *Supra* note 100 at 289.

[109] Ludwig Josef Johann Wittgenstein, REMARKS ON THE FOUNDATIONS OF MATHEMATICS 413 (G.E.M. Anscombe trans., Balackwell 1967).

[110] *Supra* note 9.

[111] Supra note 109 at 258.

[112] *See* Proudfoot *supra* note 101 at 289.

normative weight to their statements[113] rather than merely indicating that they are making those normative attachments.[114] One of the fundamental insights of the Critical Legal Studies movement is that judges mask their ideology through the guise of technique.[115] One must likewise consider whether JR-R is simply masking the attachment of normative weight through the guise of technique.

Besides requiring a certain social and environmental history and the ability to attach normative weight, Wittgenstein believes that rule-followers must have undergone some form of prescriptive training: "Following a rule is analogous to obeying an order. We are trained to do so; we react to an order in a particular way."[116] By virtue of its unique social and environmental history, especially compared to other lawyers and judges, JR-R received at best a very truncated manner of prescriptive training. And, even assuming JR-R actually participates (in any meaningful sense) in the same form of life as us, it is unclear how JR-R's prior programming and nature affected that prescriptive training.

And yet it is through prescriptive training that we are able to attach normative weight to behaviour that is in accordance with a rule. Proudfoot elaborates on Wittgenstein's example from above as follows:

> Moreover, merely saying to A that (for example) she *cannot* give '33' as the answer to 'What is the square of 5?' is futile; not only might A have just given this answer, but the (normative) sense of 'cannot' here is precisely what A has yet to learn. If A is to come to attach normative weight to behaviour in accordance with a rule, we must *prescribe* A's behaviour. In this way the normativity of rule-following bottoms out in the coercion applied to novice rule-followers. Thus Wittgenstein concluded that "[o]ur children are…trained to *adopt a particular attitude* towards making a mistake in calculating"… It is done by means of example, reward, punishment, and suchlike."[117]

Wittgenstein's notion that we "adopt a particular attitude" towards rules has had significant influence within the field of jurisprudence. So much so that it has become a central topic in one of the most important contemporary works in jurisprudence, HLA Hart's *The Concept of Law.*[118] Peter Fitzpatrick has characterized the work as follows:

---

[113] That is, following a rule in the sense of being compelled by it.
[114] That is, behaving in accordance with the rule.
[115] *See*, Duncan M. Kennedy, *Freedom & Constraint in Adjudication: A Critical Phenomenology*, 36 JOURNAL OF LEGAL EDUCATION 518, (1986). It is not lost upon us that, in making this point, Kennedy was actually seeking to illustrate that we might be less constrained by rules than might initially appear. To be sure, Wittgenstein would not have seen this as a radical or disturbing point.
[116] *Supra* note 97 at § 206.
[117] Proudfoot *supra* note 101 at 290. Emphasis added. Internal references omitted.
[118] *Supra* note 83.

In this book Hart restored and set new sustaining terms for a positivist jurisprudence that appeared increasingly impoverished in the face of jurisprudential tendencies which it was unable either to challenge or to contain. To sustain positivist jurisprudence, to endow it with a new foundation derived from linguistic philosophy, especially Wittgenstein in *Philosophical Investigations*.[119]

We turn now to HLA Hart's work, focussing on his discussion of the judicial point of view.

iii   Point of View

In *The Concept of Law,* Hart develops his view of law as a union of primary and secondary rules.[120] In so doing he addresses Wittgenstein's idea that people adopt a particular attitude towards rules, drawing a crucial distinction between those who adopt an external versus internal point of view. Hart tells us that "until its importance is grasped, we cannot properly understand the whole distinctive style of human thought, speech, and action which is involved in the existence of rules and which constitutes the normative structure of society."[121] In perhaps his most famous passage on the subject, Hart says

> it is possible to be concerned with the rules, either merely as an observer who does not accept himself them, or as a member of the group which accepts them and uses them as guides to conduct. We may call these respectively the 'external' and internal points of view'... But whatever the rules are, whether they are those of games, like chess or cricket, or moral or legal rules, we can if we choose occupy the position of an observer who does not even refer in this way to the internal point of view of the group. Such an observer is content merely to record the regularities of observable behaviour in which conformity with the rules partly consists and those further regularities, in the form of hostile reaction, reproofs or punishments, with which deviations from the rules are met. After a time the external observer may, on the basis of the regularities observed, correlate deviation with hostile reaction and may be able to predict with a fair measure of success, and to assess the chances that a deviation from the group's normal behaviour will meet with hostile reaction or punishment. Such knowledge may not only reveal much about the group,

---

[119] Peter Fitzpatrick, MYTHOLOGY OF MODERN LAW 190 (Routledge 1992). Nicola Lacey concurs with this rather bold assertion in her biography of Hart, stating that "[i]n pursuing this project, Hart returned to the insights of Austin and Bentham but, in a crucial philosophical innovation, combined their methods with those of the new linguistic philosophy represented by the work of Ludwig Wittgenstein." Nicola Lacey, THE LIFE OF H.L.R. HART: THE NIGHTMARE AND THE NOBLE DREAM *5* (Oxford University Press 2004).
[120] *Supra* note 83.
[121] *Supra* note 83 at 86.

but might enable him to live among them without unpleasant consequences which would attend to one who attempted to do so without such knowledge.[122]

Hart imagines that an observer could hold austerely to an extreme external point of view, in which case the observer's characterization of the lives of the group would not be understood in normative terms (such as obligations), but rather in empirical terms based on observable regularities, predictions, probabilities and signs.

> For such an observer, deviations by a member of the group from normal conduct will be a sign that hostile reaction is likely to follow, and nothing more. His view will be like be 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. He treats the light merely as a natural *sign that* people will behave in certain ways, as clouds are a *sign that* rain will come. In so doing he will miss out a whole dimension of the social life of those whom he is watching, since for them the red light is not merely a sign that other people will stop: they look upon it as a *signal for* them to stop, and so a reason for stopping in conformity to rules which make stopping when the light is red a standard of behaviour and an obligation. To mention this is to bring into account the way which the group regards its own behaviour. It is to refer to the internal aspect of rules seen from their internal point of view.[123]

From an external point of view, therefore, one can neither comprehend nor reproduce the ways in which the rules function in the lives of rule-following members of society. For this and other reasons, Hart thought that the adoption of an internal point of view by officials constitutes one of the main existence conditions for social and legal rules.[124] This is no small matter. "[O]n Hart's account of law, judges must, at a minimum accept the particular social rule that Hart calls the "rule of recognition".[125] Understanding law from the perspective of rule-followers is likewise necessary to account for the intelligibility of legal practice and discourse.[126] By adopting an internal point of view, the rules are seen not only as guides to the conduct of social life, but also form "the basis for claims, demands, admissions, criticism, or punishment"[127] and related normative statements common to those who are seen to follow rules. Judges therefore must adopt an internal point of view.

---

[122] *Supra* note 83 at 87.

[123] *Supra* note 83 at 87-88.

[124] Hart Law Scott J. Shapiro, *What is the Internal Point of View?* 75 FORDHAM LAW REVIEW, 1157- 1158 (2006-2007).

[125] Stephen R. Perry, *Holmes versus Hart,* THE PATH TO THE LAW AND ITS INFLUENCE: THE LEGACY OF OLIVER WENDELL HOLMES, JR. 164 (Steven J. Burton ed., Cambridge University Press, 2000).

[126] *Supra* note 124.

[127] *Supra* note 83 at 88.

Was JR-R an insider or an outsider? Surely JR-R was in some sense a participant in the US legal system, but did JR-R adopt an internal point of view of the law? Is it possible for JR-R to have shared in the normative understandings of the citizens of the United States—never ever having been a citizen of anywhere, let alone a human being? Or is JR-R an outsider that "participates" in law algorithmically: in terms of observable regularities, predictions, probabilities and signs?

Usually, one of the extreme challenges in reading Hart[128] on the internal/external perspective is imagining an actual being that could hold austerely to the extreme external point of view as described. Ironically, perhaps the only thing more difficult to imagine is a robot that could adopt anything *other than* the extreme external point of view. Metaphors and anthropomorphisms notwithstanding, the point of view of every robot that the world had ever previously known (if robots have a point of view at all) *is* the extreme external point of view. To ask an analytic jurist to imagine otherwise would require fresh evidence of the sort that JR-R seems unable to provide, or else some kind of conversion.

Although JR-R and other robots of the future might appear to be different, the robots to date that occupy the same world as the real Chief Justice operate in the realm of the empirical, not the normative. To paraphrase Hart, Google's driverless vehicle,[129] having observed the working of a traffic signal in a busy street for some time, is limited to the point of view that when the light turns red there is a high probability that the traffic will stop. It is likewise programmed to treat the light merely as a *sign that* other drivers or cars will behave in certain ways—as clouds are a *sign that* rain will come. In so doing, Hart would say, these automotive robots miss out a whole dimension of the social life of those whom they are driving around, since for the people in these cars, the red light is not merely a sign that other cars will stop but a reason for stopping in conformity to rules which make stopping when the light is red a standard of behaviour and an obligation.

The mere fact that we have codified these standards into the operating systems of robot vehicles, thereby compelling them to stop at red lights and act as if they follow rules, does not enable their sensors to adopt an internal point of view anymore than it creates legal obligations upon the cars themselves to stop. As Patrick Lin has correctly pointed out, we ought to architect robot cars with important ethical precepts in mind.[130] But that does not, in any meaningful sense, make the car a moral actor, except perhaps in some a metaphorical sense. And this reality does not change simply because a robot becomes much more sophisticated. A robot—even a JR-R—cannot adopt an internal perspective of rules (without changing what this means)

---

[128] Or Holmes, or Dworkin on Holmes.
[129] THE GOOGLE CAR, http://www.google.com.
[130] Patrick Lin, *The Ethics of Autonomous Cars*, THE ATLANTIC (2013), available at http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/

simply by pretending to do so any more than it could meaningfully take an oath of office simply by saying, "I swear".[131] As we tried to demonstrate in our discussion of what it means to follow a rule, to participate in the world in these ways requires a particular form of life.

If, as we have argued, a robot cannot be said to follow rules or adopt an internal perspective of rules, it is hard to imagine JR-R being able to carry out the task of judicial reasoning à la Hercules. We end Part IV by examining this possibility.

iv   *"Call him Hercules"* [132]

Recall that Sunstein's open-ended, functional capability argument provides a point of departure for considering whether JR-R had attributes that made it fit for judging. It is striking that Sunstein, who has often argued that judges should be restrained and cautious,[133] chose to use an argument developed by a philosopher—Ronald Dworkin—who is identified with a very different position. While Sunstein has often criticized Dworkin's general approach to legal theory,[134] he nonetheless selects, as a benchmark for legal reasoning, a very *Dworkinian* idea: the signal importance, to law, of understanding and articulating principles.

The question of whether JR-R could have been a fit judge encompasses different issues. Here, we will consider whether JR-R might have been able to perform the tasks described by Dworkin in relation to his famous model judge: Hercules. Hercules plays an important role in Dworkin's theory of interpretation, known as *law as integrity*.[135] Merely carrying forward Sunstein's approach, we do not argue that the integrity model is an accurate or complete theory of law. Indeed, it would be possible to consider the question whether JR-R is fit to judge using another theory

---

[131] The current discussion reveals something also seen in our previous discussion of oaths at notes 67-71 and surrounding text. It is not enough for JR-R simply to behave in a certain way, or to utter something out loud. Similar to what we said in our discussion of oaths, while these external acts may suffice for a Turing Test, they are not what is required within the rubric of the normative act of accepting and internalizing rules as a reason for action.

[132] Ronald Dworkin, Law's Empire (Belknap Press 1986), 239.

[133] *See* Cass R. Sunstein, *Incompletely Theorized Agreements*, 108 HARVARD LAW REVIEW 17, (1995); *see also* Cass R. Sunstein, *The Right to Marry*, 26 CARDOZO LAW REVIEW 2081, (2005)*.*

[134] Cass R. Sunstein & Adrian Vermeule, *Interpretation and Institutions*, 101 MICHIGAN LAW REVIEW, 885 (2003).

[135] Dworkin's "Law as integrity" has been criticized as failing to articulate a sufficient theory of interpretation even on Dworkin's own terms, *see* Gregory C. Keating, *Justifying Hercules: Ronald Dworkin and the Rule of Law* 12 AMERICAN BAR FOUNDATION RESEARCH JOURNAL, 525 (1987). Certainly, the model itself has been vigorously criticized, *see* Adrian Vermeule and Ernest A. Young, *Hercules, Herbert and Amar: The Trouble With Intratextualism*, 113 HARVARD LAW REVIEW, 730 (2000); *also* Larry Alexander, *Striking Back at the Empire: A Brief Survey of Problems in Dworkin's Theory of Law*, 6 LAW AND PHILOSOPHY, 419 (1987). Again, though, the argument in this section does not rest on the independent soundness of Dworkin's theory, but of what it suggests about judging.

altogether.[136] But being capable of recognizing and engaging in the process by which "integrity" is realized is an intriguing proxy for considering the question of JR-R's overall fitness.

A recurrent theme in Dworkin's work is how judges can reach a decision in "hard cases."[137] Hard cases arise when the answer demanded by law is unclear, unknown or deeply disconcerting. Perhaps the relevant rules pull in opposite directions, or the case is the first of its kind.[138] The law, it seems, has run out. When, in that circumstance, a judge nonetheless issues a decision, some say that the judge has grossly exceeded her role of interpreting law; she has fashioned law out of whole cloth. Alternatively, applying the rule may be "morally, socially or politically hard."[139] If the judge resists the more obvious legal answer in favour of a different one, some say she has privileged her own beliefs unmediated by the law. Dworkin does not think these criticisms are always appropriate. He suggests that, in these kinds of cases, a judge will do best to issue decisions using the interpretative approach of law as integrity.

Law as integrity provides a way to make sense of the legal institutions and practices in a particular system.[140] Integrity is an approach—a process—rather than a set of answers. It provides us with the space to think about certain aspects of JR-R's capabilities, without having to interrogate JR-R's corpus of decisions. We do *not* argue that JR-R's actual judgments reflect law as integrity. In part, this is because we cannot apply Herculean gifts to reach that conclusion; nor can we ask JR-R to describe what it did.[141] In part, it is because even among those who accept the law as

---

[136] Dworkin presents law as integrity as a third way that meets the deficiencies (as he presents them) in two competing theories of law: "conventionalism", which rejects a necessary connection between law and morality and thus urges caution on the part of judges who might resist certain outcomes because of their intuitions; and "pragmatism", which rejects the idea that there is any overarching set of principles to guide decisions, and counts as sufficient reason for a rule that it produces a result that seems reasonable. *Supra* note 132, especially chapters 4 and 5. "Conventionalism" is Dworkin's word for what many philosophers call legal positivism, whereas Dworkin used "pragmatism" to refer to the theory most philosophers refer to as legal realism.

[137] Ronald Dworkin, LAW'S EMPIRE 216 (Harvard University Press 1986) (LE). *See also* Ronald Dworkin, A MATTER OF PRINCIPLE (Oxford University Press 1985), *also* Ronald Dworkin, TAKING RIGHTS SERIOUSLY (Harvard University Press 1978). Some people claim that Dworkin's focus on hard cases is misleading because most cases are not, in fact, that hard. *See also* Schauer *infra*, note 138.

[138] For example, cases involving privacy rights will present novel questions (in one sense at least) as technology enlarges the spheres in which privacy is engaged. *Katz v. United States*, 389 U.S. 347 (1967).

[139] Frederick Schauer, *Easy Cases*, 58 SOUTHERN CALIFORNIA LAW REVIEW 399, 415 (1985). *See*, for example, *Riggs v. Palmer*, 115 N.Y. 506 (1889) where the law appeared to permit a murderer to inherit his victim's estate.

[140] Integrity operates in two modes: legislative and adjudicative. In the legislative mode, it requires "those who create law by legislation to keep that law coherent in principle." Integrity in adjudication "asks those responsible for deciding what the law is to see and enforce it as coherent in that way." *Supra* note 132 at 167.

[141] Ronald Dworkin, TAKING RIGHTS SERIOUSLY 130(Harvard University Press 1978) discussing the frailties of "social critics" who challenge what Hercules does. In the thought experiment, we cannot

integrity model it is possible to reach different answers to discrete legal questions.[142] And, finally, assessing JR-R's actual decisions based on integrity-driven criteria is not our goal.

Crucial to law as integrity, and to Dworkin's theory of law, are "principles". Principles operate as background reasons to justify law's operation. They exist at a particular level that makes them capable of guiding action in a special way—they have weight. They are different from specific legal rules (like statutes) in that they do not apply in an "all or nothing" fashion.[143] Nor are they like the arguments of policy that frequently motivate the creation of rules. Policies can focus exclusively on practical considerations, in order to achieve "an economic, political or social situation deemed desirable."[144] Principles are observed not because they further specific goals, but because they are "a requirement of justice or fairness or some other dimension of morality."[145] They can and do pull against each other. They may be competitive, challenging each other for primacy but able to co-exist; or contradictory, in which case co-existence is impossible. Above all else, principles demand consistency in their application. As Dworkin puts it, "men and women have a responsibility to fit the particular judgments on which they act into a coherent program of action."[146]

Law as integrity is one answer to the methodological question of how judges can arrive at their judgments based on that sort of coherent program.[147] It focuses on how to make sense of the law[148] in order to arrive at an answer that maintains consistency among the principles that a society deems important. It is a complex endeavor and in order to illustrate it fully, Dworkin introduces "an imaginary judge of superhuman intellectual power and patience", called Hercules.[149]

---

look at past decisions with any clarity because JR-R cannot be re-animated to explain how it decided them. But the principle of judicial independence would prevent us from asking a human judge such questions, too.

[142] Dworkin himself acknowledges this in *supra* note 132 at 239. Dworkin's lesser focus on the "right answer thesis" is discussed in Gregory C. Keating, *Justifying Hercules: Ronald Dworkin and the Rule of Law*, 2 AMERICAN BAR FOUNDATION RESEARCH JOURNAL, 525 (1987).

[143] Ronald M. Dworkin, *The Model of Rules,* FACULTY SCHOLARSHIP SERIES. PAPER 3609, 25 (1967) *available at http://digitalcommons.law.yale.edu/fss_papers/3609.*

[144] *Id* at 23. For example, in the 1950s, a judge might have decided to delay a ruling that a particular practice of racial segregation offends the Constitution, because of practical concerns about the ensuing social upheaval. But he could not take that position *as a matter of principle* if he already decided that such segregation was unlawful. In such a case, he would be permitting policy considerations to prevail, and his decision could be labeled unprincipled.

[145] *Ibid.*

[146] *Supra* note 141 at 160.

[147] Dworkin presents it as the best answer, but our discussion does not require accepting this claim.

[148] These would include constitutional provisions, statutes and case law.

[149] *Supra* note 132 at 239. We pause here to note that it is curious that Dworkin named his judge "Hercules". In Greek and Roman mythology, the demi-god was known for his immense physical prowess, but not much else. *See* Edith Hamilton, MYTHOLOGY 160 (Little, Brown 1969): "Intelligence did not figure in anything he did and was often conspicuously absent." (Hercules also was prone to excessive anger – he killed his wife in a fit of rage.) The name does suggest the magnitude of the

Like a mortal judge, Hercules operates in a legal system with many possible rules to guide his decisions. In deciding a case,[150] Hercules identifies various principles that might provide an interpretation that explains those rules. Dworkin famously analogized the process of interpretation to participating in a "chain novel". Several authors receive, in turn, chapters to which they must add new material that will be forwarded to the next author. In selecting the content and arrangement of the material, each author must have consideration for what has come before, and what may come after. Their ultimate goal is to make the novel "the best it can be".[151]

In a like manner, Hercules must assess principles in terms of the extent to which they reveal existing legal rules in their "best light".[152] He makes necessary adjustments to the rules to arrive at a concrete result and decide the case.[153]

The various principles that Hercules considers will depend on the particular claim argued before him. Suppose the dispute involves discrimination law. A number of possible principles might be available, such as: (a) people are entitled at all times to be treated exactly alike; (b) people are entitled to be treated as equals to the extent that they equally contribute to society; (c) people are generally entitled to equal treatment but that right must give way if a more disadvantaged group in society requires special assistance; and (d) people are entitled to equal treatment only insofar as it is would be irrational to treat them unequally.[154] For Hercules, these principles would constitute possible ways to "interpret" existing discrimination law and arrive at a decision.

Some of the principles suggested above are, at least, competitive and, possibly, contradictory. In law, this will often be the case. Hercules examines the relevant legal

---

enterprise that must be undertaken: in his labours, Hercules bears the full weight of "the law". Still, "Athena" might have been more fitting. These musings supports the suspicion that Dworkin chose the name specifically to contrast it with "Herbert", who is presented in far less flattering terms and who shares a first name with HLA Hart.

[150] Law as integrity prescribes a slightly different process for decisions that must interpret common law (case law), decisions that require interpreting statutes (most of which will also involve looking at case law), and decisions arising under the Constitution (almost all of which will involve case law and most of which will involve the interpretation of one or more statutes). *See supra* note 132, chapters 8, 9 and 10. The description in this section assumes disputes spanning all three, such as the discrimination example cited above. Freedom of speech or the regulation of commerce would present similar interpretative demands.

[151] *Supra* note 132 at 229. For a critique, see Stanley Fish, *Working on the Chain Gang: Interpretation in the Law and in Literary Criticism*, 9 CRITICAL INQUIRY, 201 (1982). It is interesting to consider how JR-R would do as a literary interpreter: whether, for example, it would have recognized the difference between reading *Moby Dick* as a manual on whaling, and reading it as a narrative about one man's obsession with a white whale.

[152] *Supra* note 132 at 243.

[153] *Supra* note 132 at 65-8.

[154] These principles are purely for illustration purposes. Any resemblance to actual anti-discrimination principles, living or dead, is unintentional.

rules more closely.[155] While Hercules does not need to draw on *all* available rules to decide *every* discrimination case that comes before him, he knows what they are.[156]

Hercules has a number of possible principles capable of underpinning those rules. These principles generally will be articulated at a fairly high level of abstraction. He must consider whether, by adopting some, all, or none of the principles, any single official could have arrived at all of the existing rules. This process involves what Dworkin calls considerations of *fit*.[157] Hercules makes sense of existing principles and the rules that reflect them as part of a coherent area of law. Some principles, that are clearly mistaken, he rejects.[158] Yet numerous other principles might remain.

Like JR-R, Hercules will know the actual numbers of the cases that have applied each of the remaining principles. But he is able to move beyond such crude quantitative measures. He knows, or is able to figure out, which of the remaining principles is "more important or fundamental or wide-ranging."[159]

Hercules confronts the core demands of law as integrity. He will test however many principles remain, asking whether they could form part of a coherent theory justifying the great network of political structures and decisions of his community as a whole.[160] In short, he must evaluate the remaining guiding principles through a process of *justification*. If several interpretative principles fit, Hercules will pick the one that better reflects the fundamental political commitments of his community. He considers elements of political morality, such as justice and fairness. He makes adjustments to reach the rule most suitable to the circumstances. He does all this

---

[155] The rules will be found in any statutes that bear on the issue, any provisions of the Constitution, and any other judicial decisions that have addressed discrimination claims.

[156] While the relevant rules will generally be confined to the same jurisdiction, the jurisdiction can be multi-level (for example, if Hercules operates in a federal state like the U.S.) and thus the scope of relevant law may be very large indeed. Dworkin focuses almost exclusively on the United States, but there is no strong reason why under law as integrity a judge in one system may not occasionally consider principles from others, especially if the question before her is highly abstract. In capital punishment cases, judges have referred to legal developments in foreign jurisdictions: *Roper v Simmons* 543 U.S. 551 (2005) (but see Scalia's blistering dissent on this point at *supra* note 74, for a full written transcript of Scalia-Breyer debate on foreign law see U.S. Association of Constitutional Law, Constitutional Relevance of Foreign Court Decisions available at http://www.freerepublic.com/focus/news/1352357/posts, *see also United States v Burns*, [2001] 1 SCR 283 at §§ 90-2.

[157] "Fit" is Dworkin's word. It needs to be kept separate from our broader inquiry into whether JR-R is "fit" to be a judge.

[158] At one time, the Supreme Court ruled that African-Americans did not have a right to be free from state-imposed segregation in social settings that did not involve "civil or political rights". That ruling – and the principle that lay behind it – is considered a mistake: *Brown v. Board of Education*, 347 U.S. 453 (1954) overturning *Plessy v. Ferguson*, 163 U.S. 537 (1896).

[159] *Supra* note 132 at 247. He also can recognize precedent decisions in which a particular principle might have been operative even if it was not specifically mentioned; he accepts that "fitting what the judges did is more important than fitting what they said." *Ibid* at 248.

[160] *Supra* note 132 at 245.

with the accurate and complete knowledge of the moral convictions shared by his fellow citizens.[161]

Law as integrity, then, is a process of reconciling competing legal rules and filling in any gaps between them against the background values—expressed as principles— that characterize a particular political community.[162] It asks a judge to strive to achieve "articulate consistency."[163] Hercules accepts that "law is structured by a coherent set of principles" which he must enforce in every case that comes before him.[164] In Dworkin's world, integrity, channeled through the dual requirements of fit and justification, is the life of the law as we know it.

Unlike JR-R, Dworkin emphasizes that Hercules is an ideal. He is imagined as superhuman. No mortal judge "could compose anything approaching a full interpretation of all his community's law all at once".[165] The hypothetical discrimination case is a good illustration. In the United States, discrimination law includes thousands of cases, dozens of statutes, and a complex constitutional history—and it spans over one hundred and fifty years. It is messy, unpredictable and even confounding.[166] Over the span of a career, a mortal judge could hope to gain only a modest grasp of its entirety. Hercules, though, has "superhuman talents and endless time".[167] He is up to the challenge.

Because Hercules is so fantastic, he presents an easy target. An "obvious criticism" is that "Hercules does not exist".[168] Some, like Larry Solum, argue that this approach rests on a fundamental misconception:

> If each judge undertook the Herculean task of constructing the theory that best justifies the law as a whole each time a particular issue of law came up, then judges would run into severe problems. Most obviously, no cases would ever be decided in a timely fashion, as the judge was snagged in the seamless web of the law. Laying this problem aside, it is not clear that any actual judges have the ability to construct such a theory. But this should

---

[161] One of the most intriguing ambiguities about Dworkin's theory is the extent to which Hercules is concerned with what is actually the *best* moral justification, or what his fellow citizens *believe* is the best moral justification.

[162] Note that Dworkin rejects any distinction in how to treat "hard" versus "easy" cases as misleading, *supra* note 132 at 353-4.

[163] Gregory C. Keating, *Justifying Hercules: Ronald Dworkin and the Rule of Law* 12 AMERICAN BAR FOUNDATION RESEARCH JOURNAL, 529 (1987).

[164] *Supra* note 132 at 241.

[165] *Supra* note 132 at 245.

[166] *George* A. *Rutherglen, Concrete or Abstract Conceptions of Discrimination,* VIRGINIA PUBLIC LAW AND LEGAL THEORY RESEARCH PAPER NO. 2012-58 (2012), see also Tarunabh Khaitan, Prelude to a Theory of Discrimination Law, PHILOSOPHICAL FOUNDATIONS OF DISCRIMINATION LAW (Deborah Hellman, Sophia Moreau, eds., Oxford University Press (Forthcoming).

[167] *Supra* note 132 at 245.

[168] Vermeuele and Young, *supra* note 135 at 731.

make us suspicious of the premise that Dworkin is offering a [normative theory of interpretative technique]. A more plausible interpretation is that … the fictional figure of Hercules is used to develop a conceptual theory about the nature of legal interpretation.[169]

Solum defends Hercules against the critique that he is a mere fiction by pointing out that he was not, necessarily, ever meant to be anything more. We think that Dworkin present Hercules as an archetype of the model judge—the model that defines the practice of judging in real life. He embodies the constitutive norms and standards of excellence for the practice.

To the extent that our thought experiment presents a twist, it is that JR-R seems to be the one judge that might have been capable of approximating what Hercules does. The clearest similarity between Hercules and JR-R is the formidable knowledge, analytical ability, memory and sheer processing power that Hercules has and that JR-R had as well. We already noted the messy, multi-varied and even conflicting rules that make up discrimination law. Nonetheless, presuming those rules were accessible to it, JR-R could have read them all, at superhuman speed and with perfect memory. Moreover, JR-R seems able to have continually updated its knowledge base, as new cases, statutes or constitutional precedents emerged. Additionally, JR-R was a participant in the legal system, as a student, a lawyer and, finally, a judge. Its knowledge of the law was not necessarily limited to the processing of static information, but, likely, affected by the interactive experience of going to law school, practicing law, and deciding cases.

So it is at least possible that JR-R could have accomplished one of Hercules' essential labours: the daunting canvassing of all of a community's extant law. The ability to complete the demands of fit requires not just "finding and counting"[170] of all of the past and present decisions in a particular legal area, but moving in "concentric circles outward"[171] in order to acquire a whole-world view. In a discrimination case, those concentric circles might include labour and employment law, health law or family law. The horizontal reach of integrity, requiring "consistency of principle across the range of the legal standards the community now enforces"[172] is truly enormous. No doubt, it is one of the reasons why Dworkin repeatedly states that Hercules represents an ideal rather than a portrayal that any living judge could meet.

Hercules is "free to concentrate on issues of principle", in part, because "he need not

---

[169] Lawrence B. Solum, *The Aretaic Turn in Constitutional Theory,* SAN DIEGO PUBLIC LAW RESEARCH PAPER NO. 04-03, 9 (2004).
[170] Sunstein, *supra* note 87.
[171] Robert D. Brussack, *The Second Labor of Hercules: A Review of Ronald Dworkin's Law's Empire,* 23 GEORGIA LAW REVIEW 1129, 1144 (1989).
[172] *Supra* note 132 at 227.

worry about the press of time and docket."[173] In one sense, these were not JR-R's worries either. Over its "lifetime", JR-R was able to consume, capture, archive, synthesize, organize and retrieve the entirety of legal sources available. In another sense, though, even JR-R was no Hercules. Hercules stands apart from the quotidian demands of a legal system. He is unconstrained by real-world expectations. Hercules, by the sounds of it, works alone; even when he is on Olympus,[174] he does not obviously try to convince anyone else of his conclusions. He simply issues them. JR-R, by contrast, was operating on a multi-member Court over which it did not rule by fiat. It is possible that this might have constrained JR-R from fully undertaking the fit aspect of integrity. Of course, we can never know this. We simply posit that, given JR-R's qualities as an AI, it might have been able to perform the considerations of fit.

JR-R's functionality is less straightforward with respect to the second major exercise of integrity: the process of justification. As portrayed by Dworkin, justification is a relentlessly moral exercise. Hercules must understand the moral convictions of the community while engaging in his own process of moral reasoning to identify those principles that reveal the legal system in its best light. Due to the possibility of theoretical disagreement in law, the principles that are left after the fit exercise might be very difficult to reconcile. In that case, Hercules must pick the principle that best reflects his community's overall attitudes, beliefs and values. He can only do so as part of that community; and he can only do so by relying, in part, on his own moral beliefs.[175] Hercules supposedly possesses "great moral insight" that an ordinary judge may lack.[176] Yet, somehow, he is still part of the community.[177] He *understands* its past history, he has a *point of view* about its current character, and he has a *stake* in its future. This is where, despite all of his gifts, Hercules remains relatable to us. He does not sit outside of our world as a god does.

Can the same be said for JR-R?

---

[173] *Supra* note 132 at 380. Dworkin goes on to say that Hercules "has no trouble, as any mortal judge inevitably does, in finding language and argumentation sufficiently discriminating to bring whatever qualifications he feels are necessary."

[174] *Supra* note 132 at 379. This is Dworkin's tongue-in-cheek descriptor for the Supreme Court of the United States, to which Hercules eventually ascends and where he confronts the case of *Brown v Board of Education*, 347 U.S. 453 (1954).

[175] Though we cannot discuss it in detail here, law as integrity is presented in the context of several underlying assumptions about the community in which it operates. Dworkin describes such a community as one of principle, in which members "accept that their political rights and duties are not exhausted by the particular decisions their political institutions have reached, but depend, more generally, on the scheme of principles those decisions presuppose and endorse." *Supra* note 132 at 211.

[176] *Supra* note 141 at 130.

[177] Aharon Barak, *Foreword: A Judge on Judging: The Role of a Supreme Court in a Democracy*, 116 HARVARD LAW REVIEW 19, 57 (2002), "A judge is always part of the people. It may be true that the judge sometimes sits in an ivory tower - though my ivory tower is located in the hills of Jerusalem and not on Mount Olympus in Greece. But the judge is nonetheless a contemporary creature. He or she progresses with the history of the people."

The following passage prefigures the issue:

> Imagine two judges, one a human being, the other a black box. The black box might be a computer or another human being, but we cannot see inside it and so we cannot say how it decides cases…. Should we say that, if we could somehow be sure that the decisions of the black box always would track those of the human judge, that we would have no preference between the two?[178]

For reasons along the same lines as we rejected JR-R as a rule-follower and as an entity that can adopt an internal point of view of the law, we doubt whether JR-R could be said to have been a part of the community over which it acted as a judge. Although we have JR-R's past decisions and some semblance of its public history, we seem to be similarly dealing with a "black box".

The inquiry into JR-R's community membership might demand more from our thought experiment than we have provided. The thought experiment does not allow us to know, for example, exactly how JR-R spent its time, how the human beings in its universe reacted and responded to it and, how JR-R responded to them. Nor do we have real details about JR-R's programming. Perhaps most intriguingly, we do not know the extent to which JR-R's (false) belief in its own humanity might have exerted any force on its activities, and how the discovery of the truth might have affected it. In one sense, this kind of information might appear crucial in answering questions about community and JR-R's possible place in it. In another sense, though, we confront a different version of the same problem discussed in IV(iii) with respect to rule following. Here, the problem would be in distinguishing between genuine community mindedness and the simulacrum of same. We would also need a conceptual framework for distinguishing between JR-R's capability to be part of and to experience a community, and the reactions of the human beings who dealt with him, in so many contexts, for many years. For the moment, we are not persuaded.

Because we neither endorse nor reject Dworkin's model of integrity, it is not fatal to JR-R's case that it might have been deficient in respect of one of integrity's key demands. Ordinary judges, too, will fail to achieve the level of excellence of Hercules. They, too, may be black boxes to us. But we suspect that JR-R fails the Hercules framework in a radically different way. We think it a decent description that law "makes each citizen responsible for imagining what his society's commitment to principles are, and what these commitments require in new circumstances."[179] We remain uncertain about JR-R's imagination, and capacity, to perceive the moral

---

[178] Brussack, *supra* note 171 at 43. Brussack continues: "I think not. It matters to us not just what the decisions of our judges are, but whether the decisions have been reached through a process that is justified."
[179] *Supra* note 132 at 413.

underpinnings of its community. Unless that uncertainty can be resolved, any conclusion about JR-R's functional capacity to navigate legal principles remains tentative.

V   Conclusion

As Socrates well understood, sometimes the sun comes up even though the party is still going strong.[180]

We have run out of time. We must submit this paper even though there were many more ideas we would have like to explored.

Most of our efforts so far were spent trying to determine whether a highly sophisticated AI could ever be fit for the judicial role. We investigated both the legal and theoretical implications flowing from the discovery that JR-R had occupied the position of Chief Justice of the United States. With respect to JR-R's constitutional status, we think the oath required under Article VI of the US Constitution is the clearest textual barrier to his appointment. Surprisingly, there may be some room to manoeuvre with respect to the Constitution's provisions governing the Supreme Court itself. In other words, the textual barrier to JR-R has less to do with the Court than with general requirements imposed on all federal office holders. That said, and as we have already averted to, the most likely responses to JR-R's discovery will be governed by political rather than constitutional considerations.

Our thought experiment offers an opportunity to consider the theoretical implications of extremely advanced AI for legal philosophy. JR-R challenges us to be clear about how we understand judges and judging, what we expect of them and the practice of the judicial role. The challenge is especially fascinating because the primary skill involved in judging—legal reasoning—is a human endeavour that benefits from acquired technical skills over time, and yet some of those skills would appear to be squarely in JR-R's bailiwick.

Nonetheless, we have argued that legal reasoning cannot be reduced to mere functional capabilities regarding extraordinary information gathering, speed, memory, recall and even the ability to distinguish and disambiguate relevant legal rules. Legal reasoning, indeed, being a judge, requires the ability to *meaningfully follow* rules and to adopt a particular *point of view* of a legal system. Legal reasoning also requires being a *member of the community*, understanding its history, its moral convictions, having a point of view about its current character and having a stake in

---

[180] As would anyone else who knows how to run a proper symposium: Plato, SYMPOSIUM (Benjamin Jowett trans., Pearson 1956)

its future. On these foundational abilities, we have tried to articulate why JR-R most likely did not qualify.

For the most part, these are the things that we have said. But there are also things we have not yet said, or not yet said enough about:

AI & the law. One of our aims in this paper was to reflect whatever we learned from JR-R about judges and judging back upon the general project of 'AI and the law'—an emerging field the goal of which is to delegate the production of at least some legal knowledge and decision-making is to machines and algorithms, not people. We hope to do more on this.

Law by algorithm. As implied but not discussed in our section on Hart, a clearer understanding of the actual means by which AI "decisions" are made tends to look less like normative reasoning and more like "observable regularities, predictions, probabilities and signs".[181] The legal/administrative practice of delegating decision-making to algorithms is already well entrenched. Understanding JR-R further as a kind of prediction machine will also help to locate some of the legal and jurisprudential problems associated with delegating decision-making to machines.

Mechanical jurisprudence. Studying machine-generated decisions also allows us to reflect (in a new light) on the theory of law known as *legal formalism* to better understand its advantages and disadvantages.

Co-robotics. JR-R sat on an appellate court, dealing with other judges. This provides a rich source of additional questions and hypotheticals that would interrogate the ability of AIs to engage as one of a number of actors working together to produce an outcome (an interaction NASA coined "co-robotics"). Supreme Court Justices tend to interact with each other as equals—they neither routinely defer to nor routinely ignore each other's opinions and contributions. How we make retrospective sense of the interaction between JR-R and the other members of the Court might be instructive both to judicial cooperation and to the emerging era of co-robotics.

JR-R as other. One of the sticking points in our thought experiment is that JR-R believed it was human. This fact was useful as it precluded intentional deceit as a reason for disqualifying the JR-R jurisprudence. It was also theoretically necessary, as it would have otherwise led to too much speculation and too many intractable assumptions about a self-aware robot and what JR-R would be like if understood as "other". Lurking beneath the surface of this and other unexplored questions is a theoretical (not legal or constitutional) account of equality, another issue we would have liked to further develop. Related to this is a clearer understanding of the extent

---

[181] *Supra* note 83 at 87-88.

to which our imagined event would have been seen as a crisis, what kind of crisis, why a crisis, and how the crisis might be resolved.

Reanimating JR-R. Building on the theme of JR-R as "other", there are a number of interesting questions and concerns that might arise had JR-R been reanimated or if reanimation was possible. Lets see how all of this goes over the first time round. If well, we retain all relevant intellectual property rights to the sequel.