JOURNAL OF
CROSS-DISCIPLINARY RESEARCH
IN COMPUTATIONAL LAW [CRCL]

Evolutionary interpretation: law and machine learning

Simon Deakin* Christopher Markou†

Abstract

We approach the issue of interpretability in artificial intelligence and law through the lens of evolutionary theory. Evolution is understood as a form of blind or mindless 'direct fitting', an iterative process through which a system and its environment are mutually constituted and aligned. The core case is natural selection as described in biology but it is not the only one. Legal reasoning can be understood as a step in the 'direct fitting' of law, through a cycle of variation, selection and retention, to its social context. Machine learning, insofar as it relies on error correction through *backpropagation*, is a version of the same process. It may therefore have value for understanding the long-run dynamics of legal and social change. This is distinct, however, from any use it may have in predicting case outcomes. Legal interpretation in the context of the individual or instant case depends upon the generative power of natural language to extrapolate from existing precedents to novel fact situations. This type of prospective or forward-looking reasoning is unlikely to be well captured by machine learning approaches.

Keywords: legal evolution, precedent, artificial intelligence, machine learning, interpretability

Replier: Masha Medvedeva, University of Groningen. m.medvedeva@rug.nl.

Journal of Cross-disciplinary Research in Computational Law

DOI: pending

www.journalcrcl.org

*Professor of Law, University of Cambridge. s.deakin@cbr.cam.ac.uk.
†Affiliated lecturer, University of Cambridge. cpm49@cam.ac.uk.

Introduction

Artificial intelligence (AI) promises to either replicate, emulate or simulate legal reasoning through a suite of statistical learning and inference-making techniques referred to as machine learning (ML).[1] While the short-term aim of AI advocates involves leveraging these techniques to complement, enhance or extend the capabilities of judges and legal practitioners, it appears that the long-term goal is replacing them altogether.[2] Thus, the rise of ML and automated decision-making is self-evidently a significant challenge to legal modes of thought and action. While this current generation of 'connectionist' AI is rich in data and wields increasingly ferocious computational horsepower with which to crunch it, the same explanatory gaps and 'penumbras of doubt' that led to the stagnation of an earlier generation of AI-leveraging models — referred to at the end of the 20th century as 'legal expert systems'— remain largely unaddressed or explained away as irrelevant.[3]

It helps to keep this not-too-distant history in mind as a new generation of 'legal tech' start-ups and their tools are unleashed upon law firms and legal systems worldwide.[4] This is particularly so with respect to those aspects of legal tech that are framing current debates around 'explainable AI' and what some call the 'seductive diversion' of solving the black box problem: finding a way to have elaborate and opaque algorithms not just show their work but justify their methods, whether this comes in the form of a 'decision tree' which allows the causal antecedents in a model to be isolated and assessed — most often in terms of the 'weight' given to a particular statistical variable — or alternative strategies such as model-agnostic explanators that can identify rules that give insights into why a model provides a specific outcome for a specific input.[5]

Translating non-linear equations and probabilistic inference into clearly defined tributaries of 'reason' is both a technical and epistemic problem. It is also a paradox that limits the very 'power and promise of computers that learn by example'.[6] This has proved especially problematic in the context of law, as the trajectory of the legal expert systems debate made clear in the past. 'Legal knowledge' seems to be more than the sum of what the most 'learned' and 'experienced' lawyers and judges 'know'. Legal reasoning — while bearing many algorithmic features — is ultimately made possible through the tremendous generative power of natural language. If law is to operate as a basis for social cooperation (in systems-theoretical terms, a 'control system' for society), it is by stabilising social expectations through the medium of language. Law operates, in other words, through the potentially infinite linguistic transformations afforded by natural language to cognise new social referents and describe the differences they make, legal or otherwise.[7] Understanding legal evolution as a process which operationalises the generative capacity of natural language will be important for arriving

1 Ethem Alpaydin, *Machine Learning: The New AI* (MIT Press 2016); David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51(2) UC Davis Law Review 653; Christopher Markou and Simon Deakin, '*Ex Machina Lex*: Exploring the Limits of Laws Computability' in Simon Deakin and Christopher Markou (eds), *Is Law Computable? Critical Reflections on Law and Artificial Intelligence* (Hart 2020).

2 Benjamin Alarie, Anthony Niblett, and Albert Yoon, 'Regulation by Machine' [2016] SSRN.

3 For a first-hand historical account and critique of the LES project: Philip Leith, 'The Rise and Fall of the Legal Expert Systems' (2010) 1(1) European Journal of Law and Technology; *cf.* John Zeleznikow and Dan Hunter, 'Rationales for the Continued Development of Legal Expert Systems' (1992) 3(1) Journal of Law and Information Science 94; Andrew Greinke, 'Legal Expert Systems: A Humanistic Critique of Mechanical Legal Interface' (1994) 1(4) Murdoch University Electronic Journal of Law.

4 Markus Hartung, Micha-Manuel Bues, and Gernot Halbleib, *Legal Tech: How Technology is Changing the Legal World* (CH Beck 2018); Robert Dale, 'Law and Word Order: NLP in Legal Tech' (2018) 25(1) Natural Language Engineering 211; Gabriele Buchholtz, 'Artificial Intelligence and Legal Tech: Challenges to the Rule of Law' in Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Springer 2020).

5 Riccardo Guidotti and others, 'Local Rule-Based Explanations of Black Box Decision Systems' [2018] arXiv; Ioannis Mollas, Nick Bassiliades, and Grigorios Tsoumakas, 'LioNets: Local Interpretation of Neural Networks through Penultimate Layer Decoding' [2018] arXiv; Christophe Labreuche and Simon Fossier, 'Explaining Multi-Criteria Decision Aiding Models with an Extended Shapley Value' in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (2018). For some well-aimed scepticism towards the idea of AI explainability, see Scott Robbins, 'A Misdirected Principle with a Catch: Explicability for AI' [2019] (29) Minds and Machines 495.

6 Royal Society, 'Machine Learning: the Power and Promise of Computers that Learn by Example' (2017).

7 Charles Stevens, Vishal Barot, and Jenny Carter, 'The Next Generation of Legal Expert Systems – New Dawn or False Dawn?' (Max Bramer, Miltos Petridis, and Adrian Hopgood eds, Springer 2011).

at a realistic appraisal of the capacity of ML to 'predict' case outcomes.

Our approach in this paper is as follows. We firstly develop an account of legal reasoning as a form of evolutionary learning (section *Legal interpretation as evolutionary learning*). Specifically, we view 'legal reasoning' as an emergent cognitive, socio-cultural and linguistic 'acquisition' process, whereby the legal system gathers knowledge about its environment (data collection) and learns through experience (*backpropagation*). In this sense, individual cases serve as 'training data' for a process of evolutionary learning akin to what we might think of as a 'legal acquisition model'.[8] This is, of course, not the only way in which legal reasoning can be understood; it is useful, however, to think of it this way, if we are to understand to what extent legal reasoning is similar to ML and to what extent it differs from it.[9]

The next step in our analysis is to make this comparison in more detail (section *Machine learning and law: 'direct fit' to society?*). Within ML, backpropagation — short for 'backwards propagation of errors' — is a widely employed algorithm in supervised learning involving artificial neural networks (ANNs) using gradient descent.[10] In short, given an ANN and a specified error function, backpropagation calculates the gradient of the error function with respect to the relative weights accorded to specific factors in a mathematical model. It achieves this by generalising the delta rule for perceptrons to multilayer feedforward neural networks; in this approach, the connections between nodes *do not* form a cycle or feedback loop.[11] We explore how far backpropagation can be analogised to the 'error correction' function performed by aspects of legal process, including the role given to appellate courts in correcting decisions of lower ones and the use of serial litigation to challenge rules which impose private and social costs. We show that both functions can be described in evolutionary terms as aspects of a variation-selection-retention (VSR) mechanism through which systems are aligned with their environments.

However, a comparison between law and ML as modes of learning also points up differences between them. We conclude (section *Conclusion: rethinking interpretability and explanation in law and AI*) that the prevailing connectionist ML paradigm is incapable of capturing the entirety, or arguably the essence, of law as a mode of social learning. Juridical reasoning employs the generative power of natural language to adjust legal rules in the face of an unstable and changing environment, while retaining the information content of previous adaptations. Thus, as currently constituted, law is simultaneously forward- and backward-looking. ML, at least in its present connectionist form, is well suited to modelling the long-run dynamics of legal change — the 'direct fitting' of law to its social context over the course of multiple iterations — but not its adjustment in the instant case. It follows that there is a role for ML in modelling law but not the one which, to date, has garnered most attention. Rather than using ML to predict case outcomes, which is likely to prove either impractical, for litigated cases, or unnecessary, for non-contested ones, it should be used to model the long-run learning process inherent in legal reasoning.

Legal interpretation as evolutionary learning

Evolution and legal theory

We are not the first to suggest that evolutionary concepts and ideas are at least useful — and with some questions indispensable — for understanding the nature of law in

8 By analogy to, or as an extension of, the 'Language Acquisition Model' proposed by Noam Chomsky, *Aspects of a Theory of Syntax* (MIT Press 1965).

9 Mireille Hildebrandt, 'Algorithmic Regulation and the Rule of Law' (2018) 376(2128) Philosophical Transactions of the Royal Society A; Paul Nemitz, 'Constitutional Democracy and Technology in the Age of Artificial Intelligence' (2018) 376(2133) Philosophical Transactions of the Royal Society A.

10 Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning* (MIT Press 2016) pp. 200-220.

11 Andreas Zell, *Simulation Neuronaler Netze* (Addison-Wesley 1994) p. 73.

its various aspects,[12] that is, as a social existent, a form of order, behavioural practice, linguistic discourse, a guide to action or repository of knowledge. However, evolutionary concepts are currently peripheral to accounts of legal reasoning and interpretation. This makes it difficult to give a coherent account of legal interpretation which is not simply descriptive, on the one hand, or unmoored from historical and social context, on the other. 'Evolution' is not synonymous with 'change', the sole descriptor of change or the only metric of it but, using a more or less precise meaning of the term, can help clarify the issues at stake.

'Evolution' in the sense that we are using it here has a meaning which Richard Dawkins summarises as 'the non-random survival of randomly varying coded information'.[13] This conveys the essence of what is sometimes called the VSR algorithm: evolution occurs through a cycle of variation (or mutation), selection (or survival) and retention (or inheritance). The process is not unique to nature. It can be observed in any context where

> information is stored and transmitted over time through certain forms ... those forms respond to changes in their environment, in the process altering the content of the information that they preserve; and ... the resulting process leads to a series of alignments between function and form, on the one hand, and form and environment, on the other'.[14]

Note that this definition does not give priority to the units in which information is stored and through which it is carried over the environment; rather it is their interaction, and mutual alignment, which is stressed. It is characteristic of numerous human and social institutions including those of the legal system.[15]

Evolution and rule-learning

George Priest's influential 1977 paper 'The Common Law Process and the Selection of Efficient Rules'[16] is notable for its lack of any substantive consideration of legal reasoning. The omission is instructive. In his model, 'judicial decision making may be described as random'[17] in difficult or contested cases, insofar as these cases generate random variations in outcomes. This is to suppose that when faced with a novel question — a century ago, whether a manufacturer of consumer goods was liable for harms caused to the ultimate user, or today, whether a computer algorithm is a

[12] Recent contributions include Robert C Clark, 'The Interdisciplinary Study of Legal Evolution' (1981) 90(5) Yale Law Journal 1238; Michael BW Sinclair, 'The Use of Evolution Theory in Law' (1987) 64(3) University of Detroit Law Review 451; Yuri V Balashov, 'On the Evolution of Natural Laws' (1992) 43(3) British Journal for the Philosophy of Science 343; Mark J Roe, 'Chaos and Evolution in Law and Economics' (1996) 109(3) Harvard Law Review 641. At the end of the 19th century and the turn of the 20th there was a similar wave of interest in evolutionary ideas among legal scholars, in which evolution was associated with successive, and increasingly progressive, stages of society: Peter Stein, *Legal Evolution: The Story of an Idea* (Cambridge University Press 1980). It may be noted that Darwin himself denied any association between evolution and social progress (Charles Darwin, *The Descent of Man* (Murray 1871) pp. 166-167 and that the modern evolutionary synthesis in biology likewise rejects teleological interpretations, in both its 'Dawkinsian' and 'Gouldian' variants (see respectively Richard Dawkins, *The Blind Watchmaker* (Norton 1986); Stephen Jay Gould, *The Structure of Evolutionary Theory* (Belknap Press 2002)).

[13] Richard Dawkins, 'Man vs. God' Wall Street Journal (12 September 2009).

[14] Simon Deakin, 'Law as Evolution, Evolution as Social Order: Common Law Method Reconsidered' in Stefan Grundmann and Jan Thiessen (eds), *Recht und Sozialtheorie im Rechtsvergleich / Law in the Context of Disciplines* (Mohr Siebeck 2015).

[15] The VSR algorithm is central to the idea of universal or 'generalised' Darwinism which has been applied in numerous social science contexts including institutional economics (Geoffrey Hodgson and Thorbjorn Knudsen, *Darwin's Conjecture: The Search for General Principles of Social and Economic Evolution* (University of Chicago 2010)), information theory (Eric D Beinhocker, 'Evolution as Computation: Integrating Self-organization with Generalized Darwinism' (2011) 7(3) Journal of Institutional Economics 393) and social ontology (JW Stoelhorst, 'The Explanatory logic and ontological commitments of Generalized Darwinism' (2008) 15(4) Journal of Economic Methodology 343), as well as the economics of law (Simon Deakin, 'Evolution for our Time: A Theory of Legal Memetics' (2002) 55(1) Current Legal Problems 1). Systems theoretical approaches have also made use of the idea that linguistic concepts allow for the retention of information, within an evolutionary framework drawing on cybernetic ideas of self-reference and co-evolution (Gunther Teubner (ed), *Autopoietic Law – A New Approach to Law and Society* (De Gruyter 1987); Niklas Luhmann, *Law as a Social System* (Fatima Kastner and others eds, Klaus A Ziegert tr, 2004); and see further section *Learning and legal concepts* below). For an argument in favour of synthesis, and conversely against fragmentation, of these contemporary theorisations of legal evolution, see Simon Deakin, 'Legal Evolution: Integrating Economic and Systemic Approaches' (2011) 7(3) Review of Law & Economics 659.

[16] George L Priest, 'The Common Law Process and the Selection of Efficient Rules' (1977) 6(1) Journal of Legal Studies 65.

[17] ibid p. 68.

'product' for the purposes of consumer protection law — different courts will arrive at a range of different solutions. Some will find for the plaintiff, some for the defendant. If there were enough decision points — or *neuronal layers* to use the equivalent term in the context of ANNs — we might plausibly imagine, according to Priest, that in around half the cases the judges opt for one outcome and that in the other half they opt for the other. This gives us the element of mutation which is needed to trigger the evolutionary process.

Selection in Priest's model is supplied by litigation, the basis for which 'builds on a model of litigation and an assumption about transaction costs that are simple and realistic'.[18] Priest's point is that litigation is skewed towards cases that impose private costs on parties. Without these costs, parties would lack incentives to challenge them. This skewing effect means that rules which are broadly acceptable in terms of the results they achieve are unlikely to get challenged.[19]

Conversely, the population of cases that *do* get litigated consists disproportionately of those which impose private and (by extension) social costs. It is only necessary to posit random decision-making coupled with litigants' private incentives to see that over time the less efficient rules are going to be purged from the system. The efficient rules are those which 'survive'. It is not so much that the efficient rules are selected *in* as much as the inefficient ones are selected *out*, leaving the former in place. The process continues through successive cycles of elimination until there are no longer any inefficient rules to be challenged, and the population of remaining efficient rules is stabilised. Priest's insight can be described more formally using the mathematical model of a Markov chain – a stochastic model which describes a sequence of potential events where the probability of each event depends on the state attained in the previous event. A *Markov process* is a stochastic one that satisfies the *Markov property* — often referred to as 'memorylessness'. Stated more simply, it is a process about which predictions can be made about future outcomes *solely* on the basis of its *present* state. These predictions are regarded as being just as useful as those that could be made with the benefit of knowing the full history of the process.

Priest's model is parsimonious, which is no doubt part of its attraction, but from the point of view of evolutionary modelling it is excessively reductive in according no role for inheritance or retention. Once we bring inheritance into the picture, we not only get a model which is more complete from an evolutionary perspective but also one that is somewhat more realistic and therefore useful in the sense of capturing features of legal decision-making additional to those Priest was able to incorporate. The key to mapping the evolutionary concept of inheritance onto law lies, we suggest, in understanding the emergent nature and practice of legal interpretation.

As a result of the insights of legal realism, the idea that legal reasoning — applying rules and concepts to the determined facts of a dispute — might actually decide the outcome of a complex case of the kind that results in a considered judgment of an appellate court has been displaced by variants of 'rule scepticism'. Law and economics scholars, following Priest, tend to take the view that economic efficiency drives outcomes; critical theory, in a mirror image of this position, emphasises the role of politics. 'Pragmatic' judges, including some of the most eminent, maintain that they arrive at decisions on the basis of what they consider to be right or just (or possibly efficient) in the instant case and then retrofit them to legal authority (precedent, statute or constitution, as the case may be) using the concepts at their disposal.[20]

Juridical language, from this point of view, is little more than a means to an end. Yet it remains the case that judges *do* routinely express their decisions through the medium of the distinct linguistic forms that we characterise as 'legal' or 'juridical'. At the very least we can say that conceptual or discursive reasoning remains the medium through which legal decisions are expressed, even if other forces

[18] Priest (n 16) p.66.

[19] Priest concedes in a footnote that 'although efficient rules may remain unchallenged where judges clearly have manifested hostility to efficient rules, the settlement of disputes arising under such rules may approximate inefficient outcomes. As with Holmes, the model in this paper construes the law to mean '[t]he prophecies of what the courts will do in fact.' ibid p. 72.

[20] Richard A Posner, *How Judges Think* (Harvard University Press 2008) p. 371.

('efficiency', 'politics', 'principle') may ultimately be driving outcomes.

The anthropologist Bronislaw Malinowski put the point this way:

> Law and order arise out of the very processes which they govern. But they are not rigid, nor due to any inertia or permanent mould. They obtain on the contrary as the result of a constant struggle not merely of human passions against the law, but of legal principles with one another. The struggle, however, is not a free fight: it is subject to definite conditions, can take place only within certain limits and only on the condition that it remains under the surface of publicity. Once an open challenge has been entered, the precedence of strict law over legalized usage or over an encroaching principle of law is established and the orthodox hierarchy of legal systems controls the issue.[21]

Malinowski's account is instructive for present purposes, particularly in its conceptualisation of 'law and order' as a 'struggle', implying an ongoing and not necessarily determinative process, albeit one that contains finite outcomes. An implication of Malinowski's account is that there is more to legal reasoning than the elimination of error, important as that is. Legal argument uses the generative capacity of natural language to cognise social existents 'in real time' through individual cases; in other words, it is as much forward- as backward-looking. At the same time, the retention function performed by conceptual reasoning prevents 'socially useful' knowledge — what we might otherwise call valid code — from being purged from the training model prematurely or without appropriate re-weighting.

Edward Levi's canonical account of legal reasoning[22] is generally taken to have explained why common law reasoning — and by extension *any* form of legal reasoning which uses adjudication in individual cases to arrive at general rules, which is a feature of the civil law too if under somewhat different conditions — cannot be described as the application of known, general rules to emerging, diverse facts. According to Levi, this is because the rule of law cannot be known until it is applied. Through the doctrine of precedent, 'a proposition descriptive of the first case is make into a rule of law and then applied to a next similar situation'. This is a three-step process: 'similarity is seen between cases; next the rule of law inherent in the first case is announced; then the rule of law is made applicable to the second case'. Far from general rules 'once properly determined' remaining 'unchanged' if imperfectly applied in later cases which are differentiated from the original source, 'the rules change from case to case and are remade with each case'.[23] This is because 'the scope of a rule of law, *and therefore its meaning*, depends upon a determination of what facts will be considered similar to those present when the rule was first announced'.[24] Thus, legal reasoning proceeds on the basis of what Levi called a 'moving classification system', whereby 'the classification changes as the classification is made'.[25] If we put this in computational terms, we could analogise legal reasoning to the process of adjusting a model in real-time on the basis of emergent, learned and retained inputs (cases), one modifiable by subsequent iterations (decisions) and normalised by error correction (appellate review and re-litigation).

From Levi's account it might seem that analogical reasoning, the core of common law interpretation, is not far removed, in terms of its effects, from the 'memoryless' decision-making posited by Priest and exemplified by Markov processes. The judge in the instant case, Levi writes, 'will ignore what the past thought important; he will emphasise facts which prior judges would have thought made no difference'.[26] At the same time, the judge does not escape the constraints of precedent. There is still a 'classification' whose contours and boundaries fall to be defined each time the rule is applied. Nor are the outcomes

21 Bronislaw Malinowski, *Crime and Custom in Savage Society* (Malinowski Collected Works. Volume 3: Crime and Custom in Savage Society, reprinted by Kegan Paul, Trench, Trubner & Co. 2001, Routledge 1926) pp. 110-111.

22 Edward H Levi, 'An Introduction to Legal Reasoning' (1984) 15(3) University of Chicago Law Review 501.

23 ibid pp. 501-502.

24 ibid p. 502 (our emphasis).

25 ibid p. 503.

26 ibid p. 502.

random. The law may dynamically adjust to novel technologies and new social concerns but it does so through the medium of an interpretative process which is recognizably a form of reasoning, if still, as Levi put it, 'imperfect'.[27]

If Levi was a rule-sceptic in the tradition of certain strands of legal realism, this scepticism did not extend to the point of arguing that rules have no causal influence on decisions. Rather, Levi's position is more nuanced: insofar as rules decide cases, cases *simultaneously* decide rules. We might describe this as not only a stochastic process but one inextricably bound up with law's recognition of and accommodation to an emergent social reality. The respect paid to juridical terms means that 'words which have been found in the past are much spoken of, have acquired a dignity of their own, and to considerable measure control results'.[28]

The appearance of axiomatic reasoning can thus mislead; legal interpretation is not 'simply' or 'purely' deductive. Yet we should accept that it is partly so, at least in the sense that an outmoded concept will only be discarded once a slow process of linguistic decomposition has been completed. The process Levi describes in some detail, namely the disintegration, in English and US tort law, of the concept of the 'inherently dangerous' product, played out over the course of an entire century.[29]

If Levi stressed the decomposition of precedents and the rules they inform in the light of technological and social change, Karl Llewellyn[30] had already shown that precedent also has an affirmative dimension. The technique of 'distinguishing' can be used, he suggested, not just to 'cut down' earlier decisions but to 'build up' precedents found useful for the instant case and thereby for future decisions. These two versions of precedent exist 'side by side' in the same judgment.[31] Their effect is that at the point of adjusting the law to a novel social or technological context, the court rationalises its decision by reference to an anterior classification which thereby assumes a new content. Explicitly using evolutionary language, Llewellyn wrote that this is the means by which the law achieves 'at once stability and change'.[32]

Learning and legal concepts

What then of the role of legal concepts? By one view they are precisely what is retained when a rule changes. The rule can be modified in terms of its meaning and scope of application, even as the concept continues apparently unaltered. This difference seems to be precisely what Gottlob Frege helped to parse out by drawing a distinction between what he calls 'sense' and 'reference'. A 'reference' of a word is the object or concept it is meant to designate; whereas the 'sense' of a word is said to be the way in which the words tie us to the object or concept.[33]

The stability of concepts has been recognised since at least the celebrated description given by Oliver Wendell Holmes in *The Common Law*,[34] according to which a rule or 'formula' (possibly a synonym here for 'concept') which has been designed for a 'primitive time' outlasts its initial justification, requiring 'ingenious minds' to find a new rationale for it in terms of present-day policy. So adjusted, the rule or concept 'adapts itself' (a phrasing which seems to anticipate later 'cybernetic' understandings of law's self-reference) to 'the new reasons which have been found for it and enters upon a new career'. As the form is given a new

27 Levi (n 22) p. 503.

28 ibid p. 506.

29 ibid pp. 507-519.

30 Karl N Llewellyn, *The Bramble Bush: On Our Law and Its Study* (Oceana Publications 1930).

31 ibid p. 69.

32 ibid p. 67. This is also captured by Baudrillard's notion of 'simulation and simulacra' which capture the sense in which representation systems operate in process of interaction with their context, with the result that 'the territory no longer precedes the map, nor does it survive it. It is nevertheless the map that precedes the territory — precession of simulacra — that engenders the territory': Jean Baudrillard, *Simulacra and Simulation* (Glaser Sheila tr, University of Michigan Press 1983) p. 41

33 Gottlob Frege, 'On Sense and Reference. Translated by Max Blach' in Peter Geach and Max Black (eds), *Translations from the Philosophical Writings of Gottlob Frege* (Über Sinn und Bedeutung, first published in *Zeitschrift für Philosophie und philosophische Kritik*, volume 100 (1892), pp. 25-50, New York Philosophical Library 1952).

34 Oliver Wendell Holmes, *The Common Law* (Little, Brown, and Company 1881).

content, it 'modifies itself to fit the meaning which it has received'.[35]

Adapting Holmes, we may say that concepts as distinct from rules adjust only very slowly to a changing context. They *do* change, but only over time, and with significant lag. Thus, concepts supply a kind of 'inheritance' mechanism or 'social memory retention'[36] function within the process of legal evolution. The validity of a legal norm is time-limited by an uncertain or unknowable future. There is no way of being absolutely certain that anyone in the future will abide by a particular rule. Law's normativity creates a presumption that what is valid or legal today was valid yesterday and will be tomorrow. The law achieves a stabilisation of expectations through the use of linguistic categories which are roughly translatable into observable practices. The resulting alignment of cognitive frames and social practice may be conceived in algorithmic terms as coefficients of the 'objective function' of law.

The question then becomes: *what* exactly do legal concepts 'retain'? Applying evolutionary language, the answer is: information about rules and the contexts where they apply, to whom and when. This information, however, takes a specific form. Concepts are not phrased as commands or executable instructions, in the manner of rules within a linear programming environment but as categories or (to use Levi's term) 'classifications'. Generalisation or abstraction becomes possible when the discursive limits of a term — such as the notion of 'reasonableness' — are reached. A linguistic mutation is needed to cognise the world as it *is*, not as it *was* before those discursive limits were reached. This process is observable when lawyers generalise or 'abstract' from the more detailed, fact-specific content of 'rules' to higher level categories. 'Abstraction' might be thought of then as a linguistic technique for 'coding' (as Dawkins puts it) complex information into a condensed form or as a 'cognitive frame', establishing what James Walsh calls a 'mental template that individuals impose on an information environment to give it form and meaning.'[37] In the cybernetic or information-theoretic terminology used by Niklas Luhmann, concepts 'store' the distinctions on which the courts rely to operationalise a rule; they are 'historical artefacts, auxiliary tools for the retrieving of past experiences in dealing with legal cases'.[38]

Avoiding the evolutionary analogy, Levi considered it unhelpful 'to dispose of the [legal] process as a wonderful mystery possibly reflecting a higher law, by which the law can remain the same and yet change'. And there is certainly no sense in which the legal system *has* to be this way. There is no law of nature, let alone of society, which dictates that social institutions *have* to mirror material processes to be efficient or functional, let alone just. If legal systems possess mechanisms akin to the replicator dynamic observed in nature, it is only because similar generic processes are at work in any context where a system has some capacity to receive feedback from its environment and adjust to changes through a cycle of information retention, intertemporal variation in the transmission of that information and subsequent error correction through both variation and selection.

When considering concepts as a form of inheritance, it is also relevant to consider the role played by written text as a medium of information retention and transmission. Text, in this sense, can be distinguished from convention, on the one hand, and digital code, on the other.[39] 'Convention', as Lewis explains,[40] is at one and the same time a basis for social order and a way of encapsulating social knowledge. Conventions are assumptions or beliefs that are widely enough held within a sufficient population to represent the 'common knowledge' of the group.

The knowledge being shared here does not simply relate to what is deemed appropriate or fair behaviour in a particular setting; it is also the knowledge that this knowledge is widely shared ('everyone knows that everyone knows'). Knowledge of this 'second order' type depends not just on

35 Holmes (n 34) p. 8.

36 Pierre Nora, 'Between Memory and History: Les Lieux de Mémoire' [1989] (26) Representations 7.

37 James P Walsh, 'Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane' (1995) 6(3) Organization Science 280.

38 Luhmann (n 15) p. 340.

39 Mireille Hildebrandt, 'Code-driven Law: Scaling the Future and Freezing the Past' in Simon Deakin and Christopher Markou (eds), *Is Law Computable? Critical Reflections on Law and Artificial Intelligence* (Hart 2020).

40 David K Lewis, *Convention: A Philosophical Study* (Harvard University Press 1969).

observation but on certain practices becoming routinised to the point where they acquire a taken-for-granted quality; where they become, in other words, 'custom'.

In the terminology of evolutionary game theory, conventions are capable of becoming 'correlating devices' which indicate to agents the 'first best strategy' for responding to the environment in which they find themselves.[41] The implication of this is that correlating devices are 'public representations' of social knowledge.[42] The information they convey is, in summary form, information about what others have done in the past and therefore what the recipient is expected to do in the present instance and, unless the environment changes, into the future.

The common law's meta norm — 'like cases must be decided alike' — assumes the role of maintaining interpretative order in a system which is perpetually on the verge of disintegrating under the cognitive demands placed upon it. The embedding of law in written text makes it possible for information 'storage' to take place on a vastly greater scale than was previously possible. Through the accumulation of individual instances, the law is called on to process an ever-greater volume of material drawn from social and commercial life. At the same time, societal complexity and diversity mean that the range of questions which the law is called on to resolve is also subject to an exponential increase. Precedent frames the process of information retrieval precisely by 'cutting down' what can be considered a valid rule and 'building up' future ones by reference to existing categories.[43] Similar meta-rules in the spheres of statutory and constitutional interpretation perform the same function.[44]

Yet it is not just the order provided by precedent which enables law to perform its task of stabilising normative expectations. As Levi emphasised,[45] it is precisely the defeasibility and openness of legal language which makes legal reasoning not just possible in its own terms but 'indispensable to peace in a community'. The legal mechanism which allows for 'differences of view and ambiguities of words' is what permits rules to be contested otherwise than by recourse to violence. It is also how a society adjusts its rules, taking 'the first step in the direction of what otherwise would be forbidden ends'. Yet it is also possible that these features of legal reasoning, apparently so vital for society's functioning, are no more than by-products of a legal process which has quite other origins and may have a very different end. Written, positive law may be less an adaptation to the needs of modernity than an exaptation[46] or fortunate accident arising from the conjunction of certain economic and technological forces. And if that is so, the arrival of digital code as a technology to rival, incorporate and possibly displace written text is an event which requires the utmost attention and care of analysis.

Machine learning and law: 'direct fit' to society?

There is a small but growing literature examining the limits which ML techniques are likely to encounter in the legal domain. Trevor Bench-Capon,[47] writing from a computer science perspective but on the basis of considerable experience of earlier (pre-ML) attempts to apply artificial intelligence techniques to legal reasoning, questions whether ML is capable of capturing the features of legal language to which Levi drew attention. It is not just that legal language has to be defeasible in order to allow for contestation. Because legal concepts operate as 'moving classification' systems, any model which makes predictions on the basis of past cases, which ML unavoidably has to do, is liable to

41 Herbert Gintis, *The Bounds of Reason: Game Theory and the Unification of the Behavioural Sciences* (Princeton University Press 2009).

42 Masahiko Aoki, *Corporations in Evolving Diversity: Cognition, Governance, and Institutional Rules* (OUP 2010).

43 Llewellyn (n 30) p. 75.

44 Thus, evolutionary mechanisms are not confined, within common law systems, to judge-made law; nor are they confined to common law systems, being found also in civil law regimes in various forms, including general clauses: Deakin, 'Evolution for our Time: A Theory of Legal Memetics' (n 15).

45 Levi (n 22) p. 501.

46 On the idea of exaptation in legal evolution, see Deakin, 'Evolution for our Time: A Theory of Legal Memetics' (n 15). The concept originates with Stephen Jay Gould and Elisabeth S Vrba, 'Exaptation: A Missing Term in the Science of Form' (1982) 8(1) Paleobiology 4.

47 Trevor Bench-Capon, 'The Need for Good Old Fashioned AI and Law' in Walter Hötzendorfer, Christof Tschohl, and Franz Kummer (eds), *International Trends in Legal Informatics: Festschrift for Erich Schweighofer* (Weblaw: Bern 2020).

miss a critical dimension of legal reasoning. This is that the past is always being reinterpreted in the light of new information; hence it is not just an uncertain future which ML needs to capture, but the retrospective reclassification of existing legal materials.

It should be noted that this critique is distinct from the one most frequently made in the context of ML, which is that it does not offer an explanation of outcomes. In other words, it is not just that ML's results cannot be adequately explained using the types of arguments which lawyers are accustomed to making. ML's predictions are likely to be 'wrong', if to be 'correct' means that they should accurately represent the outcomes in contested cases where a change in the scope of a rule is a likely result. Since, as Priest pointed out, the stock of litigated cases disproportionately consists of precisely these contested cases, ML is likely to be systematically in error at least with respect to predictions of outcomes in cases of the kind which come before courts on a regular basis. From this point of view, ML may be useful for performing other tasks of legal analysis, such as classifying documents obtained through discovery. It follows from Priest's argument that ML would also have some validity in any context where rules are well settled; but this is presumably also the area in which ML will be of least value to legal practitioners and those they advise.

Legal change has a pattern which can observed and a dynamic which can be understood. Predicting the outcome of individual cases on the basis of pre-existing rules may be hazardous in contested (litigated) cases. Predicting such outcomes on the basis of the *capacity* of those rules for adjustment in the light of new fact situations is not so hazardous; it is after all what lawyers are trained to do and do all the time. Thus, the argument that ML is backward looking, whereas law is prospective in its effects, cannot be the whole story. If legal change is explicable according to certain evolutionary dynamics, these should be reflected in historical data. It should then be possible to design ML models which incorporate some of the features of legal reasoning which define it as a learning process: not just its defeasibility and indeterminacy but also its self-referentiality and conceptual continuity.

One way into this line of argument is consider the nature of ML as a type of 'direct fitting' process which has similarities to evolution in nature and, by extension, to forms of social evolution of which law is one. Uri Hasson, Samuel Nastase and Ariel Goldstein[48] advance this kind of analysis for ML by analogy to evolution in nature. Artificial neural networks (ANNs), they point out, are 'formal learning models' inspired largely by a neuronal model of the brain, one involving biological neural networks (BNNs). ANNs are simplified models of the process of synaptic network connection observed in BNNs. Artificial 'neurons' (nodes) are linked to each other through connections ('synapses'), the strength of which ('weights') can be adjusted by learning. The weights are adjusted in response to feedback (backpropagation) so making it possible for the network to 'learn'. Over time this enables the components of the network to be fitted to its objective function, which means providing a more or less accurate fit with the feature of the world which the network is attempting to represent (the human face, human language and so on).

This type of learning does *not* depend on the ML algorithm starting out with a more or less accurate (ideal-fit) model of the world and seeking to extrapolate from that model to features of the world which are revealed from data. In any context involving complex and non-linear interactions between different components of a unity, it is not possible to extrapolate from one part of what Hasson et al call a 'design space' (here, a synonym for 'environment') to another. Rather, the process works through *interpolation*: using dense sampling of a locality or niche to arrive at an emergent understanding of its features. An 'over-parameterised' process can arrive at such an understanding of its world through 'brute-force direct fitting': with enough iterations and plentiful training data, the model achieves prediction without needing to have a fully specified model of the world to which it relates.

ANNs then do not need to learn 'simple, human-interpretable rules or representations of the world'.[49] Inter-

[48] Uri Hasson, Samuel A Nastase, and Ariel Goldstein, 'Direct Fit to Nature: An Evolutionary Perspective on Artificial and Biological Neural Networks' (2020) 105(3) Neuron 416.

[49] ibid.

pretability is just a by-product of a process which, in order to work, has no need of it. Insofar as the model appears to have been designed to fit the world to which it relates, what we are observing is simply the design inherent in the world reproducing itself in the model. A structured world results in a structured representation of that world. In the same way that Darwinian evolution removes the need to posit an intelligent force to guide change, so direct-fit ANNs enable us, Hasson et al. suggest, to dispense with intentional or interpretable rules to guide learning.

What would it mean to apply this understanding to the legal system? If evolution in nature is 'a blind-fitting process by which organisms become adapted to their environment',[50] legal evolution occurs through a similar type of 'direct fit' effect through which legal ideas, formulas, instructions and practices become adapted to their social context.

Examples of this process could be taken from almost any area of contemporary law, but a potentially useful illustration concerns the concepts used by modern legal systems to describe the phenomenon of 'work'.[51] In English law, concepts such as 'labour, 'service' and 'employment' do not just have a history; they have mutated in observable ways, their meaning shifting over the long durée between the end of serfdom in the late Middle Ages and the rise of an industrial (and possibly now post-industrial) society. In the middle of the 19th century, the term 'employee' meant a salaried manager or professional, in distinction from a manual worker in industry or agriculture, to whom the term 'servant' was then applied; a century later, 'employee' had subsumed 'servant' and referred to a wage or salary earner regardless of status or mode of working, as distinct from an independent or 'self-employed' contractor. That shift in meaning occurred alongside parallel developments in the economy (the emergence of vertically integrated forms of production and organisation) and polity (the rise of the welfare state as a means of diffusing and managing social and economic risks). The change occurred both through adjudication, as courts altered the meaning of terms in the light of the disputes coming before them, and through legislation, which gave drafters the opportunity to modify statutory definitions in the light of changes in policy.

The definitions used by employment lawyers today had their origins in a specific mid-20th century context. Does that mean that they are inapplicable to the world of gig work and the platform economy? That could have been the case if the concepts were somehow frozen in time, but that is not what we observe. The core concepts of employment law are being changed in the very process of their application to new conditions.[52] At the same time, they are not being wholly abandoned: the 'retention function' of legal concepts shapes judicial and legislative responses. The law may seem to lag behind technological change, but the information content of juridical rules may not be without value to today's conditions. Despite claims for the novelty of the platform economy, precarious work is nothing new, and solutions arrived in the past may have continuing relevance.

Thus, we can read off something of the structure of the social world from the way in which the legal system represents it. But any such 'reading' involves translation: system, code and environment are not *identical* to one other. Law is not exactly a 'mirror' of the world and, if it is taken to be one, it is liable to misinform and distort. If legal evolution is a process which does not know its own end and which produces design with no pre-design to draw on, its outcomes, whatever they may be — 'efficiency' or 'justice' according to taste or political orientation — are incidental and contingent to the underlying process. Moreover, this process, being 'blind', may produce outcomes which, far from being efficient or just, are dysfunctional, even pathological. Contrary to Priest's model, then, our understanding of legal evolution offers no guarantee of outcomes that can be thought of as optimal or even stable.

At the same time, direct fitting of law to its context is not the whole of legal reasoning. In particular, it does not describe what occurs when a judge decides the instant case. Legal reasoning in the application of rules and concepts to

50 Hasson, Nastase, and Goldstein (n 48).

51 The example that follows draws on Simon Deakin and Frank Wilkinson, *The Law of the Labour Market: Industrialization, Employment, and Legal Evolution* (OUP 2005) ch. 2.

52 Simon Deakin, 'Decoding Employment Status' (2020) 31(2) King's Law Journal 180.

the facts of the individual case uses the generative capacity of human language in ways which are not adequately captured by the notion of direct fit through iteration. Rather than interpolation, this is extrapolation or the prospective fitting of the emergent rule to new facts; as Levi put it,[53] the rule is altered in the very act of its application. Using ML to model this aspect of legal decision-making is not going to produce anything more than a partial account (at best) of the adjudicative process. Where ML may be more useful is in modelling the very long run process through which law adjusts to social, economic and political changes, and influences them in its turn. *This* process, which operates at a deep structural level below that of the individual case and is only observable over extended periods of time, might be well captured by the learning algorithms of ML.

Conclusion: rethinking interpretability and explanation in law and AI

In this paper we have sought to apply an evolutionary understanding of law to the debate over interpretability in AI. We have argued for a model of legal evolution which puts legal reasoning at the core of analysis and stresses its quasi-genetic function as a mode of information retention. Our model is more complete than the standard law-and-economics account, which focuses on variation and selection to the exclusion of inheritance or retention. Our approach also permits a more precise consideration of what legal reasoning and machine learning have in common and how they differ. Law has features of error correction similar to those captured by the backpropagation algorithm in ML. These common features may enable us to use ML to model the long-run coevolution of legal systems and their contexts (commercial, industrial, political and so on). However, ML is less well suited to capturing the type of prospective or forward-looking reasoning which occurs when judges and drafters apply legal concepts, expressing the generative power of natural human language, to novel fact situations.

If our argument in this paper is correct, the search for interpretability in AI is a diversion, at least insofar as it is taken to be synonymous with a search for a human-interpretable rules to explain AI outcomes. But the further implication of our approach is that interpretability is something of a mirage for other systems too, including the law. We can understand legal reasoning as a process without seeking, or indeed needing, to impose upon it a particular goal or end. Law does not find its justification in any particular theory of efficiency or justice. Law is not bound to reproduce these social goods and if it does so it is not as matter of routine or unaided by other institutions.

We may wish to hold law to a higher standard: to reproduce justice and social order while protecting individual autonomy. Our point, however, is that we need to be realistic about what law can achieve, to understand its distinct potential as well as its shortcomings. In particular, in the context of the law-AI debate, we should appreciate what legal reasoning has in common with ML and how it differs from it, and what would be lost if law were to be folded into ML.

Evolutionary models, including those of ML, could help supply us with what we need: a realistic and hence useful theory of law for the information age. So equipped, we can focus our attention on the juridical equivalents of network architecture, data and code. We can see then that there is a role for ML in modelling law, albeit not the one which, to date, has garnered most attention. Rather than using ML to predict case outcomes, which is likely to prove either frustratingly difficult (for litigated cases) or redundant (for non-contested ones), we should be employing it to model the long-run learning process of which legal reasoning is a part. This would imply using ML approaches to look for latent patterns and structures in the coevolution of law and its social context.

References

Alarie B, Niblett A, and Yoon A, 'Regulation by Machine' [2016] SSRN.

[53] Levi (n 22) p. 502.

Alpaydin E, *Machine Learning: The New AI* (MIT Press 2016).

Aoki M, *Corporations in Evolving Diversity: Cognition, Governance, and Institutional Rules* (OUP 2010).

Balashov YV, 'On the Evolution of Natural Laws' (1992) 43(3) British Journal for the Philosophy of Science 343.

Baudrillard J, *Simulacra and Simulation* (Sheila G tr, University of Michigan Press 1983).

Beinhocker ED, 'Evolution as Computation: Integrating Self-organization with Generalized Darwinism' (2011) 7(3) Journal of Institutional Economics 393.

Bench-Capon T, 'The Need for Good Old Fashioned AI and Law' in Hötzendorfer W, Tschohl C, and Kummer F (eds), *International Trends in Legal Informatics: Festschrift for Erich Schweighofer* (Weblaw: Bern 2020).

Buchholtz G, 'Artificial Intelligence and Legal Tech: Challenges to the Rule of Law' in Wischmeyer T and Rademacher T (eds), *Regulating Artificial Intelligence* (Springer 2020).

Chomsky N, *Aspects of a Theory of Syntax* (MIT Press 1965).

Clark RC, 'The Interdisciplinary Study of Legal Evolution' (1981) 90(5) Yale Law Journal 1238.

Dale R, 'Law and Word Order: NLP in Legal Tech' (2018) 25(1) Natural Language Engineering 211.

Darwin C, *The Descent of Man* (Murray 1871).

Dawkins R, *The Blind Watchmaker* (Norton 1986).

— 'Man vs. God' Wall Street Journal (12 September 2009).

Deakin S, 'Evolution for our Time: A Theory of Legal Memetics' (2002) 55(1) Current Legal Problems 1.

— 'Legal Evolution: Integrating Economic and Systemic Approaches' (2011) 7(3) Review of Law & Economics 659.

— 'Law as Evolution, Evolution as Social Order: Common Law Method Reconsidered', in Grundmann S and Thiessen J (eds), *Recht und Sozialtheorie im Rechtsvergleich / Law in the Context of Disciplines* (Mohr Siebeck 2015).

— 'Decoding Employment Status' (2020) 31(2) King's Law Journal 180.

Deakin S and Wilkinson F, *The Law of the Labour Market: Industrialization, Employment, and Legal Evolution* (OUP 2005).

Frege G, 'On Sense and Reference. Translated by Max Blach' in Geach P and Black M (eds), *Translations from the Philosophical Writings of Gottlob Frege* (Über Sinn und Bedeutung, first published in *Zeitschrift für Philosophie und philosophische Kritik*, volume 100 (1892), pp. 25-50, New York Philosophical Library 1952).

Gintis H, *The Bounds of Reason: Game Theory and the Unification of the Behavioural Sciences* (Princeton University Press 2009).

Goodfellow I, Bengio Y, and Courville A, *Deep Learning* (MIT Press 2016).

Gould SJ, *The Structure of Evolutionary Theory* (Belknap Press 2002).

Gould SJ and Vrba ES, 'Exaptation: A Missing Term in the Science of Form' (1982) 8(1) Paleobiology 4.

Greinke A, 'Legal Expert Systems: A Humanistic Critique of Mechanical Legal Interface' (1994) 1(4) Murdoch University Electronic Journal of Law.

Guidotti R and others, 'Local Rule-Based Explanations of Black Box Decision Systems' [2018] arXiv.

Hartung M, Bues M.-M, and Halbleib G, *Legal Tech: How Technology is Changing the Legal World* (CH Beck 2018).

Hasson U, Nastase SA, and Goldstein A, 'Direct Fit to Nature: An Evolutionary Perspective on Artificial and Biological Neural Networks' (2020) 105(3) Neuron 416.

Hildebrandt M, 'Algorithmic Regulation and the Rule of Law' (2018) 376(2128) Philosophical Transactions of the Royal Society A.

— 'Code-driven Law: Scaling the Future and Freezing the Past', in Deakin S and Markou C (eds), *Is Law Computable? Critical Reflections on Law and Artificial Intelligence* (Hart 2020).

Hodgson G and Knudsen T, *Darwin's Conjecture: The Search for General Principles of Social and Economic Evolution* (University of Chicago 2010).

Holmes OW, *The Common Law* (Little, Brown, and Company 1881).

Labreuche C and Fossier S, 'Explaining Multi-Criteria Decision Aiding Models with an Extended Shapley Value' in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (2018).

Lehr D and Ohm P, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51(2) UC Davis Law Review 653.

Leith P, 'The Rise and Fall of the Legal Expert Systems' (2010) 1(1) European Journal of Law and Technology.

Levi EH, 'An Introduction to Legal Reasoning' (1984) 15(3) University of Chicago Law Review 501.

Lewis DK, *Convention: A Philosophical Study* (Harvard University Press 1969).

Llewellyn KN, *The Bramble Bush: On Our Law and Its Study* (Oceana Publications 1930).

Luhmann N, *Law as a Social System* (Kastner F and others eds, Ziegert KA tr, 2004).

Malinowski B, *Crime and Custom in Savage Society* (Malinowski Collected Works. Volume 3: Crime and Custom in Savage Society, reprinted by Kegan Paul, Trench, Trubner & Co. 2001, Routledge 1926).

Markou C and Deakin S, '*Ex Machina Lex*: Exploring the Limits of Laws Computability' in Deakin S and Markou C (eds), *Is Law Computable? Critical Reflections on Law and Artificial Intelligence* (Hart 2020).

Mollas I, Bassiliades N, and Tsoumakas G, 'LioNets: Local Interpretation of Neural Networks through Penultimate Layer Decoding' [2018] arXiv.

Nemitz P, 'Constitutional Democracy and Technology in the Age of Artificial Intelligence' (2018) 376(2133) Philosophical Transactions of the Royal Society A.

Nora P, 'Between Memory and History: Les Lieux de Mémoire' [1989] (26) Representations 7.

Posner RA, *How Judges Think* (Harvard University Press 2008).

Priest GL, 'The Common Law Process and the Selection of Efficient Rules' (1977) 6(1) Journal of Legal Studies 65.

Robbins S, 'A Misdirected Principle with a Catch: Explicability for AI' [2019] (29) Minds and Machines 495.

Roe MJ, 'Chaos and Evolution in Law and Economics' (1996) 109(3) Harvard Law Review 641.

Sinclair MBW, 'The Use of Evolution Theory in Law' (1987) 64(3) University of Detroit Law Review 451.

Society R, 'Machine Learning: the Power and Promise of Computers that Learn by Example' (2017).

Stein P, *Legal Evolution: The Story of an Idea* (Cambridge University Press 1980).

Stevens C, Barot V, and Carter J, 'The Next Generation of Legal Expert Systems – New Dawn or False Dawn?' (Bramer M, Petridis M, and Hopgood A eds, Springer 2011).

Stoelhorst JW, 'The Explanatory logic and ontological commitments of Generalized Darwinism' (2008) 15(4) Journal of Economic Methodology 343.

Teubner G (ed), *Autopoietic Law – A New Approach to Law and Society* (De Gruyter 1987).

Walsh JP, 'Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane' (1995) 6(3) Organization Science 280.

Zeleznikow J and Hunter D, 'Rationales for the Continued Development of Legal Expert Systems' (1992) 3(1) Journal of Law and Information Science 94.

Zell A, *Simulation Neuronaler Netze* (Addison-Wesley 1994).

A reply: Law versus AI. Who will win?

Masha Medvedeva • University of Groningen, m.medvedeva@rug.nl.

The authors present an intriguing comparison of legal reasoning with algorithms behind artificial neural networks that can be used for predicting court decisions. They argue that, while machine learning has a high potential for modelling legal and social change, it is not equipped for making decisions in an instant case. While I wholeheartedly agree with this position, I would like to add and elaborate on some points from a more technological perspective.

I believe that when the authors discuss legal reasoning they mean a decision-making process; predicting court decisions does not necessarily qualify as such, however. The way in which research on predicting court decisions is conducted today does not imply an ability to make new judgments, although it is often claimed to be able to do so. In many applications outside the legal domain the processes of automated prediction and the taking of a particular decision may appear to coincide. For instance, when the phone recognises our thumb print to unlock itself, the application classifies the print as the one saved in the system or not, and based on the prediction makes a decision on whether to unlock the phone. The decision to unlock the phone is made solely on the basis of a prediction, and therefore, in practice, making the prediction equals taking the decision. Those systems generally have higher performance — and more importantly, lower stakes — than a judicial decision-making process. One could argue that making decisions using predictive models in higher stakes scenarios is the same as, for instance, predicting that one football team will win against another with the exact score of 2-1, then cancelling the game, announcing the scores and moving the team with the higher predicted score in the championship. In this scenario the processes are clearly different. I would argue that it is also the case when predicting judicial decisions.

This is not to say that predicting court decisions (or 'categorising' them, as I would rather call it [5]) is not useful in some circumstances, for example to model dynamics of legal change, as the authors suggest (as in e.g. [3]). Additionally, it can be applied for longitudinal legal and linguistic analysis of case law and understanding of the change in interpretation of various words and concepts. In the same way that modern machine learning systems are able to distinguish between *bank* as a financial institution and *bank* of the river, a neural network could be taught through time to distinguish between different meanings and thus to take into account the evolution of legal texts. It could even recognise if the word appears in a new context. However, it is hard to imagine machine learning system to be able to *predict* how various social constructs may change in the future, since it is only able to find patterns in historical data [2, 1]. In light of that, I appreciate the authors' characterisation of the interpretability of AI as a diversion, since even interpretable systems cannot have 'forward-looking reasoning'.

Moreover, the models that are actually able to predict future court decisions (we might call it 'forecasting' [5]) using machine learning are not able to imitate legal reasoning, since they are not actually modelled on it. In order to forecast decisions that have not been made yet, one needs to use documents that are available before the judgment is made, for example submissions made by the parties or decisions of the lower courts, etc. However, those do not normally contain the arguments of the court that is to make the decision, precisely for the reason that the judgement has not been made yet. Thus, models that forecast court decisions are not able to emulate legal reasoning simply because their input does not contain it. It has also been shown that it is a very hard task [7, 4, 6].

The authors also draw similarities between how legal systems correct themselves through appellate courts on the one hand, and backpropagation in neural networks on the other. I have to disagree with some parts of that analogy. Both are clearly designed for error correction; however,

they correct fundamentally different things. The court systems correct *decisions* (e.g. of lower courts) and backpropagation corrects the *weights* in order to achieve a preset known decision during learning, without changing it. Backpropagation is not used during testing of the model or making predictions using new data. Moreover, the original weights of a neural network are most often assigned randomly. Therefore, if the processes were comparable, it would mean that the legal reasoning would take into account the information randomly, often ignoring it completely and coming up with a somewhat random decision at first and then trying to adjust through a long range of appeals.

The authors point out that artificial neural networks are inspired by neuronal model in the brain and can therefore be considered a type of 'direct fitting'. The model then learns through the process of interpolation, by adjusting the weights until it has representation of the data that can be used to make predictions. However, the goal of any machine learning system is to be able to generalise in order to be able to make predictions for new data, rather than just represent the data as is. Therefore, when building machine learning applications much effort goes specifically into making sure the system (e.g. a neural network) does not fit too well (or *overfit*). A common technique when using neural networks is *dropout*, which randomly drops some nodes, allowing the system to *forget* information so as not to just memorise the data. One may argue that while machine learning is fundamentally backward-looking, this is a forward-looking element, because it is designed to make predictions about unseen (future) data. Instead of 'direct fitting' and adjusting to the social context, the system focuses on more general features that are more likely to remain the same through time.

All these arguments, however, do not deny the main purpose of the paper, to demonstrate the differences and commonalities between legal reasoning and neural network architecture. The authors provide a strong argument for why the structure of machine learning algorithms, and artificial neural networks in particular, are fundamentally not designed for judicial decision-making.

References

[1] Richard Berk. *Machine Learning Risk Assessments in Criminal Justice Settings*. Springer, 2019.

[2] Ray Worthy Campbell. 'Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning'. *Colorado Technology Law Journal* 18.2 (2020), pp. 323–349.

[3] Jens Frankenreiter. 'Are Advocates General Political? Policy Preferences of EU Member State Governments and the Voting Behavior of Members of the European Court of Justice'. *Review of Law & Economics* 14.1 (2018).

[4] Daniel Martin Katz, Michael J Bommarito II, and Josh Blackman. 'A General Approach for Predicting the Behavior of the Supreme Court of the United States'. *PloS One* 12.4 (2017).

[5] Masha Medvedeva, Martijn Wieling, and Michel Vols. 'Rethinking the Field of Automatic Prediction of Court Decisions'. *Artificial Intelligence and Law* (2022).

[6] Masha Medvedeva et al. 'Automatic Judgement Forecasting for Pending Applications of the European Court of Human Rights'. In: *Proceedings of the Fifth Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2021)*. 2021.

[7] Bernhard Waltl et al. 'Predicting the Outcome of Appeal Decisions in Germany's Tax Law'. In: *International Conference on Electronic Participation*. Springer. 2017, pp. 89–99.

Authors' response: **Use your illusion**

Simon Deakin and Christopher Markou

We are grateful for Masha Medvedeva's thoughtful feedback and critique of our paper. We greatly benefited from reading her co-authored paper 'Rethinking the field of automated court decisions'[1] and share its call for terminological caution. The response raises several important points; we wish to highlight the important distinction it identifies between *prediction* and *forecasting* in computational approaches to law.

Prediction involves estimating the likelihood of an event using a model of the world. A model is fitted to a training set, and an estimator $f(x)$ is used to derive the likelihood of something happening, or having happened, with further samples of x. Forecasting is a subset of prediction which involves estimating the future value of something given past time-series values. To be useful, a forecast requires an underlying and defensible logic. It might be wrong, but it must be based on *something*.

This brings us to the question of what it is exactly legal technologists are doing when they claim to be *predicting* the outcome of future cases. We agree with Medvedeva et al's observation:

> While researchers may believe they are "predicting court decisions", very infrequently this involves actually being able to predict the outcome of future judgments. In fact, predicting court decisions sometimes ... ended up not being anything other than identifying the outcome from the judgment text.[2]

So what *are* they doing? Trying to guess what a court *might* do on the basis of what it has done before is squarely in the realm of *legal outcome forecasting*. Maybe this sounds a little duller than 'prediction', but the difference is important. Forecasting requires more than statistical correlation if it is to work well.

Results based on statistical correlations are often presented as if they must be true once they pass a certain threshold of significance. Statistical significance, however, is just a measure of the fit between the result reported and the underlying model. Statistical correlations cannot predict change when the models on which they are based do not adequately represent causal structures. This is why Long-Term Capital Management failed,[3] mortgage schemes fail,[4] high frequency trading fails,[5] and Google, Facebook et al. would fail if the illusion of behavioural advertising and analytics was not still working well enough for them.[6]

One of the things which AI models of legal reasoning currently lack is a convincing account of what we call *forward-propagation*. By this we mean the generative ability of natural language to cognise new things in the social world, including behaviours and relationships, and thereby break from (even while claiming to apply) precedent. Computational approaches to law have thus far demonstrated little more than the ability to backpropagate reality to fit a model in the present. We appreciate Medvedeva's point that, in this respect, techniques are moving on, and may be better able to address this issue as the field develops. At least for the time being, however, AI risks instantiating an autoregressive vision of law that, true to the spirit of con-

1 Masha Medvedeva, Martijn Wieling, and Michel Vols, 'Rethinking the Field of Automatic Prediction of Court Decisions' [2022] Artificial Intelligence and Law.
2 ibid.
3 Philippe Jorion, 'Risk Management Lessons from Long-Term Capital Management' (2000) 6(3) European Financial Management 277.
4 Will Douglas Heaven, 'Bias Isn't the Only Problem with Credit Scores — and no, AI Can't Help' in Kirsten Martin (ed), *Ethics of Data and Analytics: Concepts and Cases* (Auerbach Publications 2022).
5 Ricky Cooper, Michael Davis, and Ben Van Vliet, 'The Mysterious Ethics of High-frequency Trading' (2016) 26(1) Business Ethics Quarterly.
6 Shoshana Zuboff, *Surveillance Capitalism* (Profile 2019).

servatism, would ensure that nothing would ever happen for the first time.

References

Cooper R, Davis M, and Van Vliet B, 'The Mysterious Ethics of High-frequency Trading' (2016) 26(1) Business Ethics Quarterly.

Heaven WD, 'Bias Isn't the Only Problem with Credit Scores — and no, AI Can't Help' in Martin K (ed), *Ethics of Data and Analytics: Concepts and Cases* (Auerbach Publications 2022).

Jorion P, 'Risk Management Lessons from Long-Term Capital Management' (2000) 6(3) European Financial Management 277.

Medvedeva M, Wieling M, and Vols M, 'Rethinking the Field of Automatic Prediction of Court Decisions' [2022] Artificial Intelligence and Law.

Zuboff S, *Surveillance Capitalism* (Profile 2019).