# LEGAL REGULATION, TECHNOLOGICAL MANAGEMENT AND THE FUTURE OF HUMAN AGENCY

William Lucy*

**Abstract:** This essay examines the role of human agency within two competing regulatory paradigms: law and technological management. Section 1 sketches both paradigms and suggests that the direction of regulatory travel in familiar jurisdictions is from the former and toward the latter. Section 2 examines the possible effect of this transition upon human agency. It defends a general account of agency, distinguishing that notion from autonomy, and shows that that account informs the legal regulatory paradigm. It then considers whether agency, so conceived, can persist and flourish within a technological management regulatory context. It does so by reference to a thought-experiment. That experiment, and two of three responses to it, assumes that agency can be quantified and the final part of section 2 shows how this can be done. It concludes that a transition from legal regulation to technological management will reduce the amount of human agency in the world and imperil other important values.

**Key Words:** legal philosophy, regulation, agency, technological management

"Technology . . . the knack of so arranging the world that we don't have to experience it":
M Frisch, *Homo Faber* (1957/2006) 178.

## 1. Competing Regulatory Paradigms

There are undoubtedly many ways in which humans can attempt to influence, and to some extent regulate, the conduct of fellow human beings. Law is one such general regulatory form and, as we now know it, it has an interesting and significant distinguishing feature: its mode of operation. This regulatory form's mode of engagement with the social world characteristically looks like this: it sets standards, attempts to enforce compliance with them via incentives and threats, while also declaring and sometimes actually imposing costs for

* Durham Law School; w.n.lucy@durham.ac.uk. Many thanks to the *OJLS* referees and various friends and colleagues (including Karam Chadha, Phil Handler, Johanna Jacques, Deirdre McCann, John Murphy, Sean Pattinson, Scott Veitch and Kara Woodbury-Smith), as well as audiences in Durham, HKU and Liverpool, who provided properly sceptical thoughts and questions. Also, I owe especial thanks to Professors Veitch, Murphy and the Paul KC Chung fund of the Hong Kong University Faculty of Law, and to Thomas Yeon of the HK Bar, for providing wonderful hospitality and a convivial working environment during my time as Paul KC Chung Visiting Professor at HKU, where I presented an early version of this paper.

non-compliance.[1] Modern law has many other distinguishing features, but focussing on its quintessential mode of operation illuminates the nature of its engagement with its regulatees. The standards it sets, usually general and presumptively applicable to all, are communicated and that communication, via information about enforcement and the costs of non-compliance, is intended to affect both the deliberation and conduct of its addressees.

Emphasising modern law's mode of operation might seem to mistake all law for statute law, which is an obvious embodiment of this way of intervening in the social world. But many jurisdictions also have judge-made law and some might be tempted to regard this as quite different to statute law. In some ways it is, but judge made law – certainly in the common law world – nevertheless almost always bears the three marks of law's mode of operation. So, judicial decisions are published, being communicated to the parties and the wider world; they are presented, by the judges who make them and those that publish them, as having a 'holding' – a rule as to how this particular dispute and, usually, disputes of this broad type, should be decided in future; and a specific legal consequence is declared to flow from that in the form of remedy (awards of damages, injunctions and the like), penalty (fine, imprisonment etc), or other legal 'event'.

That law's mode of operation is a communicative engagement with the practical reason of its addressees implies a great deal and some of those implications can fruitfully be bundled together. Two such bundles seem particularly salient: the agency bundle and the rule of law bundle.[2] Taken together, they constitute what I call, in subsection A below, the legal regulatory paradigm. Its competitor, sketched in subsection B, is the technological management paradigm. The remainder of section 1 highlights a fundamental difference between them, namely, the role that human agency plays in each, and gives reason to think that the technological management paradigm might now or soon will be ascendant. Section 2A offers a general account of the nature of agency, while section 2B shows (i) that that account is immanent within the legal regulatory paradigm and (ii) considers what might become of human agency, so understood, within a technological management regulatory

[1] In what follows, I regard the terms 'standards', 'directives', 'rules' and related terms as synonyms. That does not imply there are no important distinctions to be drawn between them for some purposes: on the difference between legal rules and principles, see R Dworkin, *Taking Rights Seriously* (Duckworth 1978) at 22-28 and N MacCormick, *Legal Reasoning and Legal Theory* (OUP, rev. ed., 1994) at 231-232 and 245 for different views. My use of the term 'paradigm' here has no Kuhnian overtones; it refers only to a pattern, exemplar or model.

[2] For an earlier sketch of both, see my *Law's Judgement* (Hart 2017) chs 2 and 3.

context. That discussion focusses on a thought experiment and evaluates an intuitively plausible response to it, noting that the response assumes agency can be quantified and demonstrates how that might done. Overall, I show that the competition between these two regulatory paradigms threatens some of our most fundamental commitments, particularly the value we attach to agency and its close relatives, autonomy and freedom.

*A. The Legal Regulatory Paradigm*

The agency bundle is implicated by the initial, guiding assumption of the legal regulatory paradigm, namely, that law is a means of subjecting human conduct to the governance rules.[3] That assumption brings others in its wake. One is about the nature of law's addressees: given law's guiding (and guidance) assumption, they must in principle be able to understand rules. Addressees must, therefore, share the language of the rule-maker and that language must be used intelligibly by the rule-maker. The norm, in most jurisdictions with which we are familiar, is that legal rules and directives take written form, but this is not perhaps a necessary precondition of legality.

Besides having language, law's addressees must also be assumed to be able, in general terms, to control their conduct in accord with the law. If addressees permanently lack this capacity, then issuing behavioural directives is literally pointless. I can command the trees in my garden to drop their leaves on the 14th of October every year, but I know that this command is silly. Trees lack both the capacity to understand my commands and to regulate their leaf-fall on my say so. This discloses something important about the nature of rule by law for, if addressees are in principle capable of modifying their conduct so as to accord with legal rules, then those rules must be more than merely intelligible. They must also be knowable in advance and possible to comply with. If rules of law are not knowable in advance, then I cannot take steps to comply with them. But, even when known in advance, compliance with a rule that is contradictory or otherwise impossible is itself impossible. A legal rule that commands its addressees to hover, unaided, above ground for eight hours every day is not one with which any human being could comply.

There is another assumption in play here about law's addressees. For, having the capacity (i) to understand legal rules and (ii) to modify one's conduct in accord with them

[3] LL Fuller, *The Morality of Law* (Yale UP, rev. ed., 1969) at 46, 53, *passim*.

implies (iii) some capacity for deliberation and reasoning. This implication follows if we assume, as we do in all normal instances, that the cognitive capacity of being able to understand language can and does inform the capacity for conduct modification: the medium by which the one 'informs' the other is deliberation and, of course, the rational capacity that that assumes. This capacity might well be shared by beings other than human agents, since the processes of being aware of one's environment, being able to glean information from it, and acting appropriately in light of both, seem to be displayed by the behaviour of numerous animals.[4]

This picture is shallow, telling us little of interest about actually existing persons. There should be no surprise in that, since we know the law is populated by abstractions, not the least of which is the legal person. That category often maps onto that of natural persons but certainly not always and rarely exactly. Natural persons also usually display, alongside many other capacities, the three (understanding, behavior modification and deliberation) just noted, the absence of one or more of them undermining their claim to be regarded as a full legal person, a bearer of the complete range of rights, duties, liabilities, immunities and powers which mark that status.[5]

Some of the rule of law implications that arise from law''s mode of operation have already been sketched: law's must be communicated to their addressees, possible to comply with and non-contradictory. They imply the three capacities constitutive of the agential bundle and also flow from law's guidance function: in order to be a means of subjecting human conduct to the governance of rules, these three requirements must be complied with, along with at least five others. Four of the five – that legal rules should be fairly constant, be both non-retrospective and reasonably clear, and also be general as opposed to *ad hoc* individual directives – are clearly integral to law's guidance function. *Ad hoc* individual directives can certainly guide those to whom they are directed – consider the command 'Quick march!' – but lack the generality and constancy necessary to guide large groups. Those out of earshot cannot act on the command and it is hard to discern whether or not it has long-term applicability. By contrast, retrospective directives provide no guidance at all, to either

[4] Tool use by some primates seems to be an obvious example.

[5] See, for example, the UK Mental Capacity Act 2005 for some of the conditions which can suspend or remove that status. Full legal status can also be lost on many other grounds: see, for example, the Insolvency Act 1986, sections 426A, 427 and 429 and schedule 4, as well as the Insolvency (England and Wales) Rules 2016, ch 2.

individuals or groups, unless taken as perverse guides to future conduct. Unclear directives – 'Quick march, a bit!' – also provide little guidance, generating puzzlement and requiring guesses from their addressees. The fifth requirement insists upon congruence between the rules applicable to regulatees' conduct and actual enforcement decisions about the rules (such as whether or not to enforce them, or how to interpret them). A lack of such congruence will undermine the guidance power of those rules.

Two points are worth noting about the rule of law bundle. First, it should not be thought that this bundle embodies a questionable or controversial conception of the rule of law that can stand alongside the allegedly many other controversial conceptions. That is because the eight desiderata of the rule of law bundle constitute the concept of the rule of law, the argumentative plateau from which all discussions of what else the rule of law might entail or require begin. Those discussions can be regarded as offering competing conceptions of the rule of law which have in common the concept of the rule of law, namely, the eight desiderata. If an alleged account of the rule of law does not accommodate those desiderata, then it is not about the rule of law.[6]

Second, the underpinning assumption of the rule of law bundle, that law is a means of guiding conduct, might seem unduly optimistic to some. For it might be assumed, wrongly, that using law as a means of subjecting human conduct to the governance of rules means that those rules must be morally respectable. While we would all hope for that, we are also aware of legal systems that have had and do have morally abhorrent rules; numerous philosophers and jurists have reminded us that legal systems as a whole might have morally troubling functions – they might, in some societies, be a means by which the economically dominant class upholds its position of economic domination. But none of this is incompatible with the claim, which I think undeniable, that law is a means of subjecting human conduct to the governance of rules. Using law in that way does not guarantee that the rules selected are always and ever morally proper.[7]

[6] I unpacked this argument in 'Access to Justice and the Rule of Law' (2020) 40 *OJLS* 377-402 at 385-389. We can at least add 'control of power', as delineated in GJ Postema, *Law's Rule* (OUP 2022), chs 1, 2 and 12, to the eight desiderata and law's guidance function without moving into the realm of contested conceptions.

[7] Nor is clear that Fuller thought differently, despite some critical interpretations of his work. While he held that evil cannot often stand the light of day, and hence that doing evil via the eight desiderata would be more difficult than not, he also accepted that there was a "continuum of legality" (D Dyzenhaus, *The Long Arc of Legality* (CUP 2022), 32), there being better and worse levels of compliance with the desiderata.

*B. The Technological Management Paradigm*

Roger Brownsword has ~~illuminated and~~ developed the notion of technological management in a series of important essays and books,[8] characterising it as

> "typically involving the design of products or places, or the automation of processes . . . [which] seeks to *exclude* (i) the possibility of certain actions which, in the absence of this strategy, might be subject only to rule regulation [and/] or (ii) human agents who otherwise would be implicated in the regulated activities".[9]

Technological management thus has a broader range of regulatory targets than the legal regulatory paradigm: products, places and processes join (usually human) agents as regulatory quarry. The rationale for expanding the range of regulatory targets flows from technological management's guiding assumption that regulation is or should be principally first aa matter of prohibition and elimination, seeking to prevent certain problems, forms of conduct or action ever arising by making them impossible. The best form of regulatory response to any particular social problem, on this view, is to ensure that it does not or cannot arise. Since success is not guaranteed, technological management entails more than simply creating "a designed environment (and/or controlled regulatees) with a required pattern of behaviour".[10] For, having identified the required pattern, regulators must then "monitor whether the control system is producing the required pattern; and . . . respond (by fixing the problem) where the system needs to be adjusted".[11] A particularly promising way of achieving a specific pattern of behaviour, for technological managers, is to foreclose alternatives by architecture and design. Barriers that one can pass through in only one direction and software

[8] Some of the key Brownsword essays are: R Brownsword, 'What the World Needs Now: Techno-Regulation, Human Rights and Human Dignity' in R Brownsword (ed), *Human Rights* (Hart ~~: Hart~~ 2004); R Brownsword, 'Code, Control and Choice: Why East is East and West is West' (2005) 25 *Legal Studies* 1~~-21~~ (hereinafter 'Code'); R Brownsword, 'In the Year 2061: From Law to Technological Management' (2015) 7 *Law, Innovation and Technology* 1~~-51~~ (hereinafter '2061'); R Brownsword, '*Law as a Moral Judgement*, the Domain of Jurisprudence, and Technological Regulation', ch 7 of P Capps and SD Pattinson (eds), *Ethical Rationalism and the Law* (Hart ~~: Hart~~ 2017); and R Brownsword, 'From Erewhon to Alpha Go: For the Sake of Human Dignity Should We Destroy the Machines?' (2017) 9 *Law, Innovation and Technology* 117~~-153~~. The themes of these essays are expanded in his *Law, Technology and Society* (Routledge~~Routledge:~~ 2019) and *Rethinking Law, Regulation, and Technology* (Edward Elgar 2022) and contained in capsule form in *Law 3.0* (Routledge~~Routledge:~~ 2021). The notion's antecedents are in L Lessig, *Code Version 2.0* (Basic Books~~: Basic Books~~ 2006) 72-74, 323-324 and Appendix, as well as his 'Law of the Horse' (1999) 133 *Harvard LR* 501~~-546~~.

[9] 2061 at 8 (my emphasis and inserts ~~mine~~). This and the next two paragraphs are derived from my 'The Death of Law: Another Obituary' (2022) 81 *CLJ* 109-138, 114-115.

[10] Brownsword, Code, above n 8, 7.

[11] Code, 8.

that will not work unless one accepts the terms and conditions of use are quotidian instances of technological management.

By contrast with the legal regulatory paradigm, the technological management paradigm makes few assumptions, beyond thinking that the best regulatory response to a perceived problem is to solve it, which often~~usually~~ means: stop it arising.[12] That stands alongside an assumption of parsimony: that the least demanding and least complex regulatory response is usually~~likely~~ the best. Such a regulatory response need not be the simplest or the cheapest; it should, rather, come closest to complete prevention. None of the assumptions that animate the agential and rule of law bundles of implications are necessarily in play in the technological management paradigm, except indirectly: human agents are usually the designers of particular technological management regulatory responses. For such designers, undoubtedly human manifestations of the agential bundle, components of that and the rule of law bundle might be valuable as part of a parsimonious response to a particular problem, but that is a purely contingent matter; they might not. The paradigm therefore has no pre-commitments about the standing of either bundle.

It would be a mistake to think that technological management paradigm is new, since human beings have always had recourse to places, processes and products as a means of regulation. Pyramids look like an instance of technological management, as do fences, safes and vehicle immobilisers. But, as hinted at in Brownsword's use of 'automation', the regulatory techniques now available in our societies are not just machine-based and computer-driven, but also potentially ubiquitous. Various rubrics have been used to characterise deeply networked societies in which everyday devices (internet enabled fridges, lightbulbs, coffee pots, thermostats and cars) are linked to other more obvious web-connected information processing devices (our laptops, watches, health monitors), all of those capable of being integrated with larger networks of surveillance and monitoring (*inter multos alia*: traffic, travel and movement monitoring systems; health, consumption, trading and market behaviour tracking; internet use and preference monitoring). Such societies are 'Smart' or 'digital', exemplars of the Internet of Things or illustrations of 'everyware'.[13]

[12] I therefore claim that contemporary technological managers are undoubtedly 'solutionists' in E Morozov's sense: see his *To Save Everything, Click Here* (Penguin 2014) ch 1.

[13] The Chinese Social Credit system is touted as a significant step toward realising this goal: see D Mac Sithigh and M Siems, 'The Chinese Social Credit System: A Model for Other Countries?' (2019) 92 *MLR* 1034-1071.

These societies portend ubiquitous regulation via technology without the direct involvement or even explicit awareness of regulatees, whose conduct can be subject to regulation in a frictionless way. Their environment and the opportunities it offers can be pre-determined by everyware.[14] Furthermore, such societies will and perhaps already do embody what Mirielle Hildebrandt has called 'data-driven agency' (*DDA*).[15] That

> "refers to a specific type of intelligence, capable of perceiving an environment and acting upon it, based on the processing of massive amounts of digital data. Data-driven agents can be more or less embodied, ranging from robots (drones, self-driving cars or even companion robots) to software bots (search engines, advertising auctions, smart energy-grids). ... Currently data-driven agency informs a host of invisible adaptations of our online and 'offline' environment, and the rise of a so-called 'cyber-physical infrastructure' indicates that the distinction between online and offline is becoming increasingly artificial, if not redundant. A cyber-physical infrastructure basically entails turning devices, homes, public and private transport, bridges, hospitals and offices online, to enable persistent monitoring and surreptitious adaptation".[16]

When one Data-Driven Agent is integrated with others in a multi-agent system, the system can become smart in the sense that, unlike other automatic and semi-autonomous systems,

> "it contains a more fundamental measure of unpredictability as to how . . . [it] achieves its goals. It contributes to solving problems that its programmers could not foresee. Here we are talking about a set of interacting artificial agents . . . which are executing their own programmes and negotiating with each other to achieve their own goals (predefined by the system developer).[] These interactions generate systemic effects or emergent behaviours at the level of the system that have not been planned or directed from a central point . . .. The system as a whole thus develops what has been called 'global agency', meaning that it begins to behave as a unity of action within its environment".[17]

Thus, a regulatory paradigm at least as old as the legal regulatory paradigm can now – or soon will be – implemented at such a scale and in so many ways that its range and capacity will be massively extended. Everyware and *DDA*, facilitated by the data-processing,

[14] Code, 3-4.
[15] When speaking of data-driven agents, as opposed to data-driven agency, I spell the term out in full.
[16] M Hildebrandt, 'Law as Information in the Era of Data-Driven Agency' (2016) 79 *MLR* 1-30 at 4.
[17] M Hildebrandt, *Smart Technologies and the End(s) of Law* (Edward Elgar 2015) 26.

pattern-recognition and outcome generating capacities of current Deep Neural Networks, Recurrent Neural Networks and generative Large Language Models, can "make calculated *predictions* about future behaviours and states of the world"[18] at a speed and in a quantity that far supersede human abilities. The knowledge those predictions generate allow "highly targeted, dynamic interventions in new markets and other social relations yet at a global scale and in real time", an on-going automated re-organisation of social life.[19] That spectre – regulation *by* AI, not regulation *of* AI[20] – informs what follows.

*C. So What?*

Before outlining what is at stake in the competition between these two regulatory paradigms, we must first make explicit what has been implicit: that these paradigms are ideal-types occupying more or less opposite points on a regulatory spectrum. By 'ideal-types' I mean exactly what Max Weber meant: they are abstractions, accentuating features of empirical reality, but nevertheless rooted in that reality.[21] They should, therefore, resonate with features we recognise in our social, political, legal and cultural context; one invoking an ideal-type should be able to point to actual instances in that context which exemplify one or more features of the ideal-type. Ideal-types are not ideal in the sense of being perfect or commendable, but in the sense that they highlight and curate some features of the social, political, legal and cultural context. A properly constructed ideal-type, in the Weberian manner, should have an historically deep and rich, often comparative, empirical base and I provide nothing like that here.[22] Furthermore, it is eminently possible for scholars to offer alternative or expanded versions of existing ideal-types. In the juristic-cum-regulatory

[18] K Yeung and M Lodge, 'Algorithmic Regulation: An Introduction', ch 1 of K Yeung and M Lodge (eds), *Algorithmic Regulation* (OUP 2019), 10; emphasis in original.
[19] *Ibid*.
[20] Two prospective attempts at regulation are the European Union AI Act (https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf) and the US President's Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/).
[21] This breezy characterisation simplifies a great deal but captures the ideal type of Weber's ideal type: see *Economy and Society*, Vol 1 (University of California Press 1978), G Roth and C Wittich (eds) at 3-24 for his own characterisations. For elucidation and discussion, see B Nefzger, 'The Ideal-Type: Some Conceptions and Misconceptions' (1965) 6 *The Sociological Quarterly* 166-174; T Burger, *Max Weber's Theory of Concept Formation* (Duke UP, expanded ed., 1987) chs III and IV; and F Ringer, *Max Weber's Methodology* (Harvard UP 1997) ch 4.
[22] I made a start on providing such a basis for the legal regulatory paradigm in *Law's Judgement*, n 2 above.

context, for example, Laurence Diver developed a much broader ideal-type of technological management than that offered here, although they are not actually incompatible. Diver's view is that the technological management paradigm is found far beyond the regulatory contexts in play in this essay and he is surely right about that.[23] But, for present purposes, adopting that broader view obscures rather than illuminates the particular issue I address.

What is at stake in the competition between these paradigms? Are they genuinely incompatible or actually complementary? A standard hypothetical often used to illustrate the contrast between them relates to road traffic regulation.[24] The way legal-regulators attempt to achieve the goal of reducing the speed of traffic in a locale would be: (i) set a maximum speed limit and publicise that; (ii) set penalties for exceeding the maximum and publicise them; and (iii) ensure some means of enforcing those penalties against those who exceed the limit. A technological management approach to this goal would have no pre-commitments as to how to achieve it, except for displaying a preference for the most efficient method. And that method could well ignore each of the three steps that a legal-regulator would regard as indispensable. But if, as now seems to be the case, cars can be designed so as not to exceed the speed limit in any particular area, technological managers are most likely to adopt that means to achieve their goal: the goal is designed-in to the means – the vehicle – by which breach of the goal used to be possible, albeit via human agency.[25]

The most vivid point of contrast in the hypothetical is the role of agency in each approach. The technological management approach eliminates human agency as a means of breach, allowing it a role only in the construction of the solution (insofar as human agents are involved in the process of designing and building the vehicles and the necessary traffic infrastructure). The agency of drivers with regard to both knowledge of the regulatory goal and the choice to comply is rendered redundant. For legal regulators, the agency of regulatees is paramount in both dimensions: legal regulators see themselves as duty-bound

[23] L Diver, *Digisprudence* (Edinburgh UP 2023) chs 1-3. Related attempts to characterise our current regulatory context are Karen Yeung's notion of 'algorithmic regulation' (see Yeung and Lodge, n 18) and Fleur Johns' broader idea of 'governance by data': see 'Governance by Data' (2021) 17 *Annual Review of Law and Social Science* 53-71.

[24] This example and paragraph are from my The Death of Law, n 9, 115.

[25] This possibility has allegedly been realised, albeit not seamlessly, in San Francisco: see https://www.washingtonpost.com/technology/2024/01/23/san-francisco-lawsuit-robotaxi-waymo-cruise/ and https://www.washingtonpost.com/technology/2024/01/25/cruise-investigation-doj-sec/ (note also https://www.tesla.com/support/autopilot). For a quaintly dated observation on related matters, see M van Hees, *Legal Reductionism and Freedom* (Kluwer 2000) 100.

to engage and inform regulatees in their efforts to affect the latter's conduct, yet non-compliance is always an option. How can these two approaches possibly be reconciled?

It is easy to see how a conscientious 'smart' regulator could seamlessly move from one point to the other on this regulatory spectrum.[26] For, in addition to doing all the things a legal-regulator would do to reduce traffic speed, a conscientious regulator might also exploit other strategies as a means to that regulatory end. They might, for example, engage the community (drivers, other road users, pedestrians) in dialogue about the problem and the means of solving it; signs, in addition to speed limits, might be installed reminding drivers of their context and responsibilities ('Kids Play Here!'; 'Quiet Traffic Zone' etc etc); and physical changes to the road layout could be adopted (speed humps, carriageway narrowing, reversing the traffic flow etc etc). If a technological fix subsequently becomes available which ensures cars cannot exceed the speed limit, then that is an obvious next regulatory step: the problem will be eradicated.

Charting this process does not, however, show that starting and end points are actually combinable as regulatory strategies. It serves, rather, to emphasise how different they are and, indeed, that the difference is so significant as to be qualitative rather than one of degree. One regulatory paradigm is agency-respecting, probably by definition, and it might, as a result, be agency-enhancing. The other might or might not be agency-respecting and might or might not be agency-enhancing. Hence agency is at stake in the choice and competition between these two regulatory paradigms. But if this is indeed the stake here, then it might be thought that there is still no genuine choice or competition because all human agents would favour the agency-respecting and possibly agency-enhancing option. Matters are not, however, so simple, since there are ostensibly good reasons for agents not to respect agency.

Regulatory efficiency is one. Note how parsimonious, effective and therefore appealing the technological management response to traffic regulation seems. Assuming the technology is available at reasonable cost, how could we object to a regulatory response that makes it impossible for vehicles to speed and therefore guarantees compliance with the law?

[26] 'Smart' regulators "embrace[..] [*inter alia*] flexible, imaginative, and innovative forms of social control which seek to harness not just governments but also business and third parties": N Gunningham, 'Enforcement and Compliance Strategies', ch 7 of R Baldwin, M Cave and M Lodge (eds), *The Oxford Handbook of Regulation* (OUP 2010) 131. This almost antique sense of 'smart' is quite different to that mentioned in n 17, above, although it could now be supplemented by AI, computation and massive data processing.

Furthermore, making non-compliance impossible is obviously more than a matter of just ensuring that the law is not breached: it will bring about a decline – perhaps at some point the complete eradication – of road traffic injuries and deaths.[27] Such a result seems good in and of itself and, when we consider other associated savings that will flow from preventing these losses – the reduction of cases in the courts and all the related costs of litigating road traffic injuries – that impression is reinforced.[28] The benefits appear so obvious and the costs so low that this solution, like most technological management solutions, looks irresistible.

It is also the case that technological fixes such as this look like morally costless interventions in our world. Ensuring that cars cannot speed is a matter of product design and raising questions about agency and associated notions here seems crass, making a mountain of a molehill. Of course, a product that does less harm than it otherwise would looks like a morally better product that should appeal to the conscience of both designers and purchasers. When technological management solutions provide such a moral 'bonus' at no apparent moral cost, they become eminently salient, especially when they are just one further step among a myriad of other similar regulatory interventions: vehicles that continually remind drivers to fasten their seat belts, train doors that do not open without central control, medicines that monitor their consumption.[29] The fact that technological management solutions can often be implemented without recourse to seemingly cumbersome and inefficient law-making and law-enforcement mechanisms might further add to their lustre.[30]

These considerations suggest that, far from being an unattractive option, technological management is the obvious "direction of regulatory travel" in many societies

---

[27] Do not ~~Nor should we~~ assume, despite some contrary indications ~~i~~(code is "perfect control" (Code, n 8 at 4) and "code is law" (*ibid*, at 5)), that code-cum-technological management regulatory solutions are either inescapable or impossible to ~~challenge/~~subvert: see C Reed and A Murray, *Rethinking the Jurisprudence of Cyberspace* (Edward Elgar~~: Edward Elgar~~ 2018) at 86-101 for discussion. ~~In terms that will become fully intelligible only after reading section III B, below, the attempt to 'conduct' conduct – even by code – often creates 'counter-conduct'.~~

[28] That adjudication is regarded as problematic (unnecessarily expensive, inducing fractiousness and the fraying of social ties, inhibiting humane conduct and interaction) is, I think, part of a broader process of 'pathologising' law: see my The Death of Law, n 9 at 118-123

[29] On the latter, see I Goold, 'Digital Tracking Medication: Big Promise or Big Brother?' (2019) 11 *Law, Innovation and Technology* 203-230.

[30] On the potential irrelevance of one or other public/private distinction within the context of technological management, see The Death of Law, n 9 at 115-116 and, for the broader issue, W Lucy and A Williams, 'Public and Private: Neither Deep Nor Meaningful?', ch 2 of K Barker and D Jensen (eds), *Private Law: Key Encounters with Public Law* (CUP~~: Cambridge UP~~ 2013).

with which we are familiar, the path they seem bound to pursue.[31] If that path indeed beckons and is taken, the question of human agency – can it flourish in an environment in which technological management is the principal regulatory mode? – is one of many that arises. To answer it, we must first clarify agency's contours in general terms, so as to see how, if at all, it relates to the kind of agency animating the legal-regulatory paradigm. And, second, we have to consider how we might measure or otherwise assess agency, so as to evaluate its (so to speak) well-being or state of health: might agency flourish in some regulatory contexts and wither in others? I tackle both issues in what follows.

## 2. Agency: Nature and Prospects

### *A. Agency in General*

There is an ordinary common-sense view of what it is to be an agent which is also, I think, philosophically respectable. In a variety of situations, ordinary people speak about the importance of having some degree of control over their conduct and context. Women undergoing obstetric care, people involved in medical and social care and related processes, as well as victims in criminal justice systems, make a roughly similar refrain: at the very least, they want not only to know what is happening, but also to be involved, as co-authors, in the processes that envelop them.[32] Failures to consult and engage those affected by official or expert action, as in the case of unauthorised medical examinations, or those whose suffering has triggered official action, such as in the case of criminal proceedings, evidence a lack of respect for those apparent agents.[33] And, in part, those apparent agents characterise such baleful situations as ones in which they lack control. The inability to control major points

[31] 2061 at 5.

[32] Good starting points for the substantial literature in these areas include: V Smith-Oka, S Rubin and L Dixon, 'Obstetric Violence in Their Own Words: How Women in Mexico and South Africa Expect, Experience, and Respond to Violence' (2022) 28 *Violence Against Women* 2700-2721; H Keedle, W Keedle and H Dahlen, 'Dehumanized, Violated, and Powerless: An Australian Survey of Women's Experiences of Obstetric Violence in the Past 5 Years' (2022) *Violence Against Women* Online First (https://journals-sagepub-com.ezphost.dur.ac.uk/doi/full/10.1177/10778012221140138); R Holder and E Englezos, 'Victim Participation in Criminal Justice: A Quantitative Systematic and Critical Literature Review' (2024) 30 *International Review of Victimology* 25-49; S Irazola et al., 'Keeping Victims Informed: Service Providers' and Victims' Experiences Using Automated Notification Systems' (2015) 30 *Violence and Victims* 533-544; S Hitlin and G Elder, 'Agency: An Empirical Model of an Abstract Concept' (2006) 11 *Advances in Life Course Research* 33-67; and G Elder and M Johnson, 'The Life Course and Ageing: Challenges, Lessons, and New Directions' ch 2 of R Settersten, Jr. (ed.), *Invitation to the Life Course* (Routledge 2002).

[33] An interesting empirical study of the connection between respect and agency is W Schirmer, L Weidenstedt and W Reich, 'Respect and Agency: An Empirical Exploration' (2012) 61 *Current Sociology* 57-75.

during one's life course – decisions about retirement or health care, for example – is also characterised in a similar way, as a failure of respect and self-authorship.

We could elaborate this view of agency in abstract terms as a matter of being in the world, by which is meant an agent is (i) aware and has knowledge of the world, including its other inhabitants as well as the constitutive institutions and practices of the social world; (ii) that their agency is a matter of interaction and deliberative engagement with the social and natural worlds; and (iii) that their agency includes the capacity to control their way of being in those worlds and the manner of their deliberative engagement with them.[34] It is too much to hold that agency is a matter of complete self-authorship on this view, but it certainly presupposes a capacity for intentional action and the sub-capacities internal to that:[35] an ability to plan, to project a plan or plans into the future, to monitor how those plans are going and to make behavioural changes in light of such assessments. Plans in this sense need not be grandiose goals, since there is a plan involved in my intention to go to the supermarket, just as there is in my aim of becoming a heart surgeon or a novelist. According to some philosophers, the pursuit of plans in part explains the psychological continuity we ascribe to most normal human agents and which makes sense of their agency across time.[36]

The capacity to control one's way of being in the world integral to this view of agency is, obviously, a *capacity*: it need not therefore always be exercised nor need it always yield the outcomes toward which it is exercised, since there is no guarantee that one's conduct always brings about the situations or goals at which one aims. One can have the capacity to control both one's way of being in the world and the nature of one's deliberative engagement with it, without one's life or the world turning out the way one intends. This, at least, can be

[34] In capsule form, "[a]gency represents a human capacity to influence one's own life within social structured opportunities": Hitlin and Elder, *ibid,* 56-57 or, in another term from social psychology and sociology, it is a matter of "self-efficacy": A Bandura, *Self Efficacy: The Exercise of Control* (Worth 1997). Philosophical accounts of agency which emphasise the three components mentioned include G Watson, *Agency and Answerability* (OUP 2004) Part I, J Shepherd, *The Shape of Agency* (OUP 2021) chs 1-5 and the sources in notes 35, 36, 39 and 41 below. The broader philosophical issues here are examined in E Mayr, *Understanding Human Agency* (OUP 2011) ch 11; M Schlosser, 'Agency' (2019) *Stanford Encyclopedia of Philosophy* (https://plato.stanford.edu/entries/agency/) and C List and P Pettit, *Group Agency* (OUP 2011) ch 1.

[35] The centrality of intentionality to agency, although the latter does not explicitly feature, is the crux of GEM Anscombe's *Intention* (Blackwell 1957) and is also emphasised by D Davidson, *Essays on Actions and Events* (OUP 1980) chs 1-5.

[36] This is Michael Bratman's view: see his *Intention, Plans, and Practical Reason* (Harvard UP 1987) chs 1-3; *Faces of Intention* (CUP 1999) chs 1- 4; and *Structures of Agency* (OUP 2007) chs 2 and 10. Some social psychologists and sociologists of the life course call this 'planfulness': see M Shanahan, S Hofer and R Miech, 'Planful Competence, the Life Course, and Aging: Retrospect and Prospect', ch 5 of S Zarit, L Pearlin and K Schaie (eds.), *Personal Control in Social and Life Course Contexts* (Springer 2003).

so in some instances of agency, since the general capacity to do X is compatible with attempting and failing to do X on some occasions. Similarly, one can have the general capacity to act intentionally without acting intentionally on all occasions.[37]

I suggest we dub this view of agency 'the thin view' for two reasons. First, there is nothing in it which necessarily confines agency to humans: many non-human animals seemingly have awareness and knowledge of the world, deliberative engagement with it, and a degree of control over their manner of being in it. They also seem able to act intentionally. Second, it excludes other, more complex but related notions, such as autonomy. Being able to distinguish these two notions is an important philosophical advantage of the thin view for, while autonomy and agency are assuredly connected – that connection presumably being the reason why some treat them as synonymous – there is a significant difference.[38] That difference becomes plain, somewhat counter-intuitively, when we attempt to distinguish autonomy from heteronomy.

These are ostensible opposites and part of their opposition resides in the notion of control, specifically over one's wants and one's conduct. Heteronomous beings lack such control, either completely or to a degree: they are therefore often 'victims' of their wants and desires and their conduct manifests that. Harry Frankfurt dubbed such beings 'wantons', a term that captures the idea of beings uninterested in evaluating their impulses-cum-first-order desires.[39] By contrast, Frankfurt claimed *deliberative* control is a hallmark of autonomous beings. Such beings are able to act on the basis of second-order volitions, those volitions being the result of having evaluated and achieved some critical distance from their first-order wants and desires. They are therefore able to decide which first-order desires are to be their will. That ability, for Frankfurt, is the crux of autonomy.[40] He is not alone in thinking

[37] On general and particular capacity, see T Honore, *Responsibility and Fault* (Hart 1999) ch 7.

[38] Some instances of the two being either elided or insufficiently distinguished: PP Verbeek, 'Subject to Technology' ch 2 of M Hildebrandt and A Rouvroy (eds.), *Law, Human Agency and Autonomic Computing* (Routledge 2011) at 29-33 (on 'autonomy' but agency seems to be meant); J de Mull and B van den Berg, 'Remote Control', ch 3 of Hildebrandt and Rouvroy, *ibid*, 47-54 (ditto); A Alghrani *et al*., 'The Mental Capacity Act 2005 - Ten Years On' (2016) 24 *Medical Law Review* 311-317 at 312 (slippage between capacity and autonomy) and J Coggan, 'Mental Capacity Law, Autonomy, and Best Interests: An Argument for Conceptual and Practical Clarity in the Court of Protection' *ibid* 396-414 at 398-403 (a series of quick moves between capacity and autonomy). For interesting discussion of 'machine autonomy' (which looks like agency), see H Chia et al., 'Autonomous AI: What Does Autonomy Mean in Relation to Persons or Machines? (2023) 15 *Law, Innovation and Technology* 390-410.

[39] H Frankfurt, *The Importance of What We Care About* (CUP 1988) 16.

[40] *Ibid*, ch 2 and 164-176.

that, since the thought that the ability to formulate, test and sift – or to critically evaluate and subsequently endorse or reject – one's wants, goals and desires is central to a large number of accounts of autonomy, although the language used to characterise it often differs.[41]

The ability to decide which desires to desire, to critically evaluate and prioritise one's wants and to act upon them, is clearly related to agency. Agency is implicated in the attempt to act upon and realise one's second-order volitions. Doing so in such a way as to maximise one's chances of success requires knowledge of the world and of those in it, as well as a level of deliberative engagement with both such as to allow one to recognise and pursue appropriate means to one's chosen end. But since it is conceivable that one's second-order evaluations require no or almost no world-directed conduct from oneself – I have decided, for example, that a life of pure contemplation is the one for me and am supported in that by the efforts of others – then that must mean autonomy does not necessarily require agency. At least, not in this 'skimmed' form, although 'full-fat' autonomy – realising one's second-order desires in the world – clearly does. Furthermore, the fact of agency implies nothing, in and of itself, about the kind of second-order evaluation characteristic of all forms of autonomy. Heteronomous beings are agents, at least insofar as actual pursuit of their wants is only possible with the knowledge and capacities that are hallmarks of agency. Having low levels of impulse control or zero interest in evaluating one's wants implies neither confusion about one's environment, nor about one's wants and goals, nor an inability to realise them.

Just as agency, on the thin view, has close connections with autonomy, so both notions also implicate the idea of freedom, albeit not in exactly the same form. Understood as the capacity to formulate second-order volitions, autonomy is a matter of what Frankfurt calls 'freedom of the will': an autonomous being must be able to evaluate their wants and desires on their own terms and at their own bidding. It makes little sense to claim that a being who has been forced to come to certain second-order volitions is autonomous unless, by chance, they subsequently endorse those volitions without coercion or manipulation. We can, however, plausibly claim that a being who foregoes any kind of active engagement in the

[41] This is most evident in moral-philosophical work inspired by Immanuel Kant. An admirable account of the significance of Kant's view of autonomy within the history of moral philosophy is J Schneewind, *The Invention of Autonomy* (CUP 1997) chs 22 and 23. Some contemporary instances of broadly Kantian conceptions of autonomy – autonomy as moral self-governance – are: O O'Neil, *Constructions of Reason* (CUP 2012) chs 3 and 4; C Korsgaard, *Creating the Kingdom of Ends* (CUP 1996) Part I, *The Sources of Normativity* (CUP 1996) chs 1-4 and her *Fellow Creatures* (OUP 2018) chs 3, 4 and Part II.

world is autonomous if that withdrawal is a second-order volition. Unlike the thin version of agency, instances of 'skimmed' autonomy such as this neither require nor presuppose active engagement with the world. And such engagement is usually ~~only~~ valuable, from either the first- or second-person perspectives, only when it is also an instance of freedom of action.[42] The conceptual cartography here, then, looks like this: the thin view of agency (i) implies freedom of action; and (ii) is connected to freedom of the will if and when that freedom entails realising certain goals or bringing about particular situations in the world (which we dubbed 'full-fat' autonomy). The eradication or reduction of agency could therefore reduce the amount of freedom of action in the world and/or limit opportunities for full-fat autonomy. I assume that the occurrence of either would be a cause for concern.

### *B. Agency's Future and Measure*

"Prediction, obviously, is about the future, yet it reacts back on how we conceive the future in the present":
H Nowotny, *In AI We Trust* (Polity 2021) 4.

It is clear that the thin view of agency has much in common with the bundles of assumptions, both agential and rule of law-related, which underpin the legal regulatory paradigm. It is no exaggeration to say that those bundles more or less instantiate the thin view within the particular social-cum-institutional context of law. Awareness and knowledge of the social world, accompanied by the capacity for both deliberative and actual engagement with that world, as well as some degree of control over those forms of engagement, are implied by the assumptions of both: neither the set of agential nor rule of law assumptions make sense without those broader agential capacities being in place. We can therefore say that agency not only survives within the regulatory environment constituted by the legal regulatory paradigm, but that it is to some extent respected and might even flourish there. Neither is likely within a technological management regulatory context simply because agency has no presumptive weight there: it is neither presupposed nor necessarily accommodated by that regulatory framework. A wholesale or retail switch from a legal regulatory to a technological

[42] "[T]o deprive someone of his freedom of action is not necessarily to undermine the freedom of his will [or autonomy] . . .. When we ask whether a person's will is free we are not asking whether he is in a position to translate his first-order desires into actions. . . . . The question of the freedom of his will does not concern the relation between what he does and what he wants to do": *Importance*, n 39 at 20.

management framework would therefore seem to reduce the incidence of agency. Furthermore, we might think such an outcome intuitively and absolutely obvious.

Imagine, for example, that we were able in this moment to replace large swathes of private law in any contemporary jurisdiction with technological management analogues: think, for instance, of replacing contract, tort, trusts and land law as we now know them.[43] All transactions with anything like a contractual component, such as conveyances of property, the creation of trusts and quotidian contracts for services, goods and the like, could be replaced with self-executing blockchains or related technology, while the boundaries of property, personal and social space could be policed and immediately enforced by advanced geo-fencing technology. Personal injury and related negligence claims could be eradicated by, for example, infallible self-driving vehicles and AI-driven nurses, surgeons, accountants, building inspectors, law-enforcement agents and the like. In addition, a digitised and blockchain-secure register of all property and land interests would provide an utterly incorrigible guide to who owns what.

Would there be less agency in this regulatory system than in current legal systems? Does expanding the range of technological management reduce the range of agency? *If everything else in the jurisdiction and its wider social and cultural context remained the same*, then it is tempting to answer both questions in the affirmative. The spaces for agency that had existed within private law will have been either reduced or eradicated: decisions about contractual performance are made redundant by blockchain, careless driving is impossible, as is substandard performance of myriad other duties. So, too, is discussion about and formulation of appropriate behaviour intended to be satisfactory performance of a duty, at least as between human agents. One might discuss the different options for a medical procedure with one's robot surgeon, but it seems unlikely that the surgeon will initiate that discussion, unless specifically programmed to do so. Furthermore, unless space to question records is built into both the digital register of all land and other interests and the geo-fences which police personal and property boundaries, then the agential capacities entailed in contesting them will be redundant.

[43] For a measured assessment of how the two regulatory paradigms might combine in some of these contexts, see Brownsword, *Law, Technology and Society* n 8 at chs 10 and 11.

I label this answer to the two related agency questions just raised the 'intuitive response' and examine it further below. It merits sustained treatment because it raises two central and pressing issues of contemporary moral, legal and political philosophy – the extent of the moral community and the quantifiability of freedom – which also animate everyday moral, legal and political discourse. The discussion consists of an evaluation of some objections to the response.

*C. Interrogating the Intuitive Response*

There are at least three reasons why the intuitive response might be regarded as mistaken. I examine the first two together, in subsection (I) below, while the third is dealt with in subsection (II). The first two reasons are connected in that they concern the redistribution of agency, while the third raises an issue that also implicates those two reasons.

(I) Redistributing Agency

The first reason why the intuitive response is mistaken is this: the reduction or elimination of agency within various branches of private law might be counterbalanced by an increase in agency elsewhere. As humans are freed from the responsibilities that previously burdened them and which were undoubtedly occasions for the exercise of agency, they can exercise agency in other, perhaps more fulfilling contexts. Mundane exercises of agency constitutive of everyday economic and legal life could be replaced by those allowing me "to do one thing today and another tomorrow . . . hunt in the morning, rear cattle in the evening, criticize after dinner, just as I have a mind, without ever becoming hunter, fisherman, shepherd or critic".[44] If the overall quantity of agency in the world remains the same, then this kind of emancipation – "franchising out our [mundane] choices"[45] – must be a good thing. There is the same quantity of agency in this world as the world it replaces, but the agency here is of better quality.

Such an outcome might indeed come to pass and is consistent with a long and optimistic strand of thought-cum-futurology about automation.[46] If it could be guaranteed,

[44] K Marx and F Engels, *The German Ideology*, edited and introduced by CJ Arthur (Lawrence and Wishart 1974), 54.

[45] D Runciman, *The Handover* (Profile 2023) 253.

[46] A helpful overview of contemporary and historical work on the alleged benefits and disadvantages of a long-promised age of automation is A Benanav, *Automation and the Future of Work* (Verso 2020). A classic and still

then it would be a clear advantage of the transition from the legal regulatory to the technological management paradigm and ~~. It~~ would also prove the intuitive response mistaken. However, the fact that the commercial enterprises currently driving most significant everyware and related technological advances have so far displayed little or no respect for either the agency or privacy of their users, suggests that this beneficial outcome is unlikely. The pessimistic strand of thought about automation and technological development might give a more accurate indication of our destination.[47]

~~———~~Second, the intuitive response could be thought mistaken because it focusses solely upon *human* agency, which might indeed be reduced were our thought experiment to become real. But, even if it were, *data-driven agency* (*DDA*) could be increased in such a scenario. Replacing private law with technological management analogues removes agency from humans, reallocating it to the various automated, AI-driven systems that constitute those analogues. There might, then, be no diminution at all in the overall amount of agency in that world as compared to the one it replaced, merely a redistribution. Is there a problem with that claim? There are perhaps two, the first being the objection that *DDA* is not 'really' agency. The second objection, which many human agents would highlight, consists of the complaint that the claim both redistributes agency in the 'wrong' direction and assumes that human agency and *DDA* are comparable.

The first objection is difficult to establish insofar as the claim that *DDA* is not *really* agency depends upon showing its dissimilarity to human agency. In fact, if the thin version of agency is the best version, characterising agency for all beings and things capable of it, then it looks very similar to the agency many claim exists in autonomous machine learning/artificial intelligence systems. This claim is not made by technological evangelists keen to inflate the success and advantages of technology, but by restrained sceptics such as Hildebrandt. Her general claims about agency, which fit perfectly with the thin version, are (i) that "an agent is an entity that acts . . . [and that] action requires some form of intention"; and (ii) that can

unfulfilled prophesy about capitalism and automation is JM Keynes, 'Economic Possibilities for Our Grandchildren' in his *Essays in Persuasion* (MacMillan 1931) 358-373 at 368-369.

[47] For the (at best) ambivalent attitude of current technology enterprises to agency, privacy and related notions, see S Zuboff, *The Age of Surveillance Capitalism* (Profile 2019), part III and, for a more limited but equally informative discussion, E Mik, 'The Erosion of Autonomy in Online Consumer Transactions' (2016) 8 *Law, Innovation and Technology* 1-38. The leading contemporary technological pessimist is perhaps Yanis Varoufakis: see his *Technofeudalism* (Penguin 2023) chs 3-6.

"perceive and respond to changes in the environment . . . [and] endure as a unity of action over a prolonged period of time".[48] Since "agency . . . [is therefore] the ability to observe an environment and to act upon it based upon . . . observation",[49] there is no difficulty in saying that it extends to Data-Driven Agents:

> "[d]ata-driven agency is a type of agency where observations are limited to digital data and actions are informed by the computational processing of such data. This brings any determinate data-driven decision-system under the concept of data-driven agency, for instance an application that determines when to start the central heating based on a decline in temperature, or one that determines social security benefits based on input of relevant data and the specified decision tree. Next to deterministic systems, which are in principle predictable, we now have systems that apply machine learning (ML), meaning that the system updates its own operational rules based on the feedback it receives".[50]

The principal features of a Data-Driven Agent's agency are therefore much the same as the agency we expect to find displayed by humans. The most we could say is that these two forms of agency are not in all respects identical, but that cannot, without more, license the claim that they are qualitatively different. *DDA* differs from human agency in ways similar to which some instances of non-human animal agency might differ from human agency, yet they are significantly similar, just as bicycles can differ considerably from one another yet remain the same type of thing. Of course, we might wish to argue that human agency is more valuable than other forms of agency and there is no *a priori* block to that. Such an argument cannot, however, begin from the claim that human agency is not the same as, or is radically different from, either *DDA* or non-human animal agency, and therefore more valuable than them. At least, it cannot do so if the thin version of agency is the best one available.

The two parts of the second objection are connected, since the complaint that agency is redistributed in the 'wrong' direction is likely to be informed by a denial that human agency and *DDA* are comparable. The 'wrongness' of the redistribution depends upon the claim that

[48] *Smart Technologies*, n 17 at 22.
[49] M Hildebrandt and K O'Hara (eds.), *Life and the Law in the Era of Data-Driven Agency* (Edward Elgar 2020), Introduction at 1.
[50] *Ibid* at 1-2.

human agency is more valuable or more important than *DDA* and that, of course, is a denial that they are comparable.

It is difficult for us, as human beings, not to be anthropocentric about ourselves and hence an intuitive rejection of the comparability of *DDA* and human agency should be expected. We often view the world as constituted by a hierarchy of species, humankind at the apex. While this view can collapse into speciesism – the unjustified assumption either (i) that only human kind has moral value and standing, other species having none or (ii) that other species are always less morally important than humans[51] – the fact that Data-Driven Agents are the product of our hands makes it hard for us to regard them as our equals and even harder to accept them as members of the moral (human and non-human animal) community.

That does not mean that non-animal things cannot have either moral or legal standing, but only that such things are usually what humankind take to be significant parts of the natural world, such as rivers, forests and the like, or objects that we regard as having deep cultural significance (like some buildings or art works).[52] Tools, a class of things of which ML/AI systems are surely members, are rarely granted that status.[53] They are, at most, commemorated as testaments to human ingenuity – think, for example, of the recreated 'Manchester Baby' (or Small Scale Experimental Machine) on display in the Manchester Science and Industry Museum – and used both as mirrors into epochs remote from ours and as educational tools. Their value is therefore purely instrumental, in the sense that their significance is solely a result of their utility for us. By contrast, other human beings and, quite possibly many or all non-human animals, have what some moral philosophers call non-instrumental or unconditional as well as instrumental value. There are more and less complicated ways in which this idea can be unpacked, but a unifying theme in all is that things with unconditional value are not valuable just because they bring about or uphold other

[51] The most influential statement of this position is P Singer's in his *Animal Liberation* (Harper Collins 1975) chs 1, 5 and 6 (it features in all subsequent versions of the book, including *Animal Liberation Now* (Vintage 2023)). See S Kagan, 'What's Wrong with Speciesism?' (2016) 33 *Journal of Applied Philosophy* 1-21 for a careful evaluation of Singer's view and, for penetrating general observations, B Williams, 'The Human Prejudice', ch 13 of *Philosophy as a Humanistic Discipline (*Princeton UP 2008).

[52] A classic starting point here is C Stone, *Should Trees Have Standing?* (OUP 3rd ed., 2010), ch 1 of which was originally published in (1972) 45 *Southern California Law Review* 450-501.

[53] For the suggestion that such tools might be regarded as slaves, see Runciman n 45 at 253. For some legal issues involved in giving robot versions of such 'slaves' rights, see B Bennett and A Daly, 'Recognising Rights for Robots: Can We? Will We? Should We?' (2020) 12 *Law, Innovation and Technology* 60-80 at 63-71.

valuable states of affairs or goals. The basis of unconditional value is sometimes regarded as a particular type of rational capacity – for example, "[t]he capacity to propose an end to oneself"[54] – possession of which marks out some members of the natural and social world for special attention. Their standing in those worlds is thereby enhanced and consequently there are some things that cannot be done to them and some ways in which they cannot be treated.

If we accept this claim about non-instrumental value, then drawing a significant moral line between tools, on the one hand, and human and non-human animals, on the other, appears a good deal easier than drawing such a line between human and non-human animals.[55] The claim that both of the latter groups are potential sources of non-instrumental value provides a starting point for an argument which shows that they are more morally valuable than Al/ML systems, even though all three groups are capable of agency. It is thus not agency that makes the key moral difference here, but the capacity (or feature or characteristic) upon which unconditional value is said to rest. Agency does, however, have moral significance, since it needs to be in place if beings are to turn the exercise of that morally charged capacity (proposing an end to oneself, for instance) into conduct. Agency is a condition, as noted above, of full-fat autonomy.

The second reason to think the intuitive response mistaken is, then, unpersuasive. At least, that is so if all forms of agency do not have equal standing, since the re-distribution of it from one group of putative agents (humans) to another (Data Driven Agents) is not always morally insignificant. Indeed, the claim that there is an important moral difference between these two groups of agents is the core of the objection to the comparability claim.

## II. Quantifying Agency

The third reason which might cast doubt upon the intuitive response also underpins concern about the discussion of the first two reasons and about the formulation of the intuitive response itself. For the way that response and those reasons were expressed undoubtedly implied that agency, like human beings and trees, is something that can be counted. Disquiet about that idea is the third reason to doubt the intuitive response. Just as some notable

[54] The phrase is Kant's, quoted in full in Korsgaard, *Creating the Kingdom of Ends* n 41, 110.

[55] For scrupulous examination of some of the difficulties involved in drawing the latter line, see Korsgaard, *Fellow Creatures* n 41, chs 1-3; Singer, n 50 and his *Practical Ethics* (CUP 1979; 3rd ed., 2011) ch 3; and M Nussbaum, *Justice for Animals* (Simon and Shuster 2023) chs 1-6.

thinkers have doubted that freedom can be measured, holding it to be a non-quantifiable variable, so we can imagine the same doubt being raised about the prospect of quantifying agency.[56] However, I suggest that the doubt about quantifying agency can be removed in exactly the same way as the doubt about the quantifiability of freedom or liberty (which I will regard as synonyms here). The pioneering work of Hillel Steiner and Ian Carter shows how this can be done.

I concentrate mainly upon the latter's work in what follows because, although clearly inspired by Steiner, it constitutes the principal book-length analysis of the problems and possibilities of quantifying freedom.[57] Furthermore, lest this approach to quantifying agency be thought eccentric – it seemingly diverts us head-on into all the complexities inherent in the discussion of liberty – remember that any plausible account of agency necessarily implicates a notion of freedom. At the very least, such an account assumes a picture of freedom of action, conduct being the culmination of agency, and, sometimes, also a conception of freedom of the will. This much is plain from some existing efforts to quantify agency which nevertheless avoid explicit discussion of either form of freedom.[58] Given the close connection between the notions, such avoidance can only ever be temporary and Steiner's and Carter's work reminds us of that. Thus there is nothing odd in 'working back' from their accounts of how liberty might be quantified, so as to see how agency can also be quantified.

Carter defends and commends an "empirical conception of freedom" (*MF* 7), by which he means a view of freedom "according to which the extent of … [one's] freedom is a function of the extent of action available to … [one], in 'sheer quantitative terms'" (*MF* 170). He maintains, following and extending Steiner, that quantitative assessments of the extent of

[56] Two notable doubters about the quantifiability of freedom are Ronald Dworkin (regarding liberty as a "commodity", the amount of which can be measured, "seems bizarre": *Taking Rights Seriously* n 1 at 270) and Charles Taylor ("a quantitative conception of freedom is a non-starter": 'What's Wrong with Negative Liberty' in A Ryan (ed.) *The Idea of Freedom* (OUP 1979) at 183).

[57] *A Measure of Freedom* (OUP 1999); I refer to this work as '*MF*' in text and notes, with accompanying page numbers (all emphases are in the original). Its antecedents are H Steiner 'The Natural Right to Equal Freedom' (1974) 83 *Mind* 194-210; 'Individual Liberty' (1974-5) 75 *Proceedings of the Aristotelian Society* 33-50; 'How Free: Computing Personal Liberty' in A Phillips-Griffith (ed), *Of Liberty* (CUP 1983) 73-89; and *An Essay on Rights* (Blackwell 1994) chs 2 and 3. Two important engagements with the Steiner-Carter view are van Hees, n 25, part 2 and M Kramer, *The Quality of Freedom* (OUP 2002) chs 3-5. An interesting application of the view is K Oberman, 'Freedom and Viruses' *Ethics* 132 (2022) 817-850 at 821-837.

[58] See, for example, Hitlin and Elder, Agency, above, n 32 and A Donald et al., 'Measuring Women's Agency' (2020) 26 *Feminist Economics* 200-226.

both individuals' and groups' liberty are not only possible, but also necessary if we are to make sense of the everyday tendency to offer quantitative freedom judgements such as, for example, the claim that the Danes are 'freer' than Russians, or that the wealthy enjoy greater liberty than the penurious. Carter argues that such claims cannot be plausibly cashed out on alternative, non-empirical accounts of freedom, particularly those that are value-based (*MF* part II). That is not to say Carter thinks freedom is not a value; rather, he argues that it is '"non-specifically valuable'", valuable "not only because of the specific things it allows us to do, but also because of the mere fact of our having freedom" (*MF* 34).

If we accept that freedom is non-specifically valuable on the empirical view, how do we go about calculating how free a particular person is overall? Carter invokes Steiner's formula here, which holds that "the extent of an agent's overall freedom can be represented ... [by] the value of" the difference, in terms of a fraction, between the number of a given list of actions one "is free and unfree to perform respectively" (*MF* 172).[59] For the formula to work, that list must be capable of precise specification: at the very least, we need to know exactly what 'actions' are, including where they begin and end, if we are to itemise or list or individuate them. Furthermore, it must also "be a list of all the actions which [an agent] can reasonably be described as either free or unfree to perform" (*MF ibid*). That the formula represents one's overall freedom as a fraction is a result of Steiner's intuition that measuring a person's freedoms alone is insufficient, since freedoms can always be accompanied by unfreedoms. Hence "a person's overall freedom must be a function not only of the number of her specific freedoms, but also of her specific *unfreedoms*, that function consisting in the ratio of the number of actions she is free to perform . . . to the number of actions that she either free or unfree to perform" (*MF*, *ibid*).

For our purposes – the quantification of agency – the possibility that both Carter and Steiner hold out and attempt to realise, that of being able to count both freedoms and unfreedoms, is vital. This possibility implies that we know exactly what counts as a 'doing' or an action, since action is constitutive of both freedom and of constraints upon freedom (they do not regard 'natural' or non-human limits on freedom as genuine constraints).[60] For us, being able to say what is and is not an action, and also being able to specify the exact limits

[59] For Steiner's statement of the formula, see How Free, n 57 at 74; for discussion and elaboration, see Kramer, *Quality*, n 57 ch 5.
[60] See *MF*, ch 8 and *Essay*, n 57 at 8.

of any particular action, allows us to quantify agency insofar as an instance of action is also necessarily an instance of agency.

The way Carter recommends we count and identify actions will not tell us the full extent of agency in the world, since agency can be realised in our refrainings as well as our doings. Yet refrainings, like mental acts, "cannot themselves be members of the set of acts that are taken into account by measurements of overall empirical freedom" (*MF* 206). Why? Because, unlike refrainings and mental acts, only actions are capable of being "spatio-temporally located particulars" (*MF* 176),[61] such particulars providing "us with a criterion of act-identity – a criterion which allows us to say when two act-descriptions are descriptions of the *same* action" (*ibid*). That is important with regard to the issue of quantifying both freedom and agency because we need a resolution to the philosophical problem of multiple act descriptions which, for some philosophers, brings in its wake a genuine multiplicity of acts.[62] Half of Steiner's 12 entry list (taken from Eric D'Arcy) of descriptions of a supposedly single act illustrates the issue:[63]

1. He tensed his forefinger.

2. He pressed a piece of metal.

…

4. He pulled the trigger of a gun.

…

7. He shot a bullet towards a man.

…

10. He killed a man.

…

12. He saved four lives.

How many actions, and thus instances of agency, are there here? Presumably not 12 (or even 6). But, if we're inclined to think that there are fewer than 6 actions here and, perhaps, possibly only one (that of saving four lives?), how can we reduce the multiplicity of descriptions of action to fewer than that or even to just a single action?

[61] The influence here is Davidson: see n 35, chs 1, 3 and 8.
[62] Carter ascribes this view to Alvin Goldman: *MF* 177.
[63] How Free, n 57 at 75.

Carter's answer is, in part, that we can do this if and when we can identify a basic action, "an action that we do not perform *by* performing another action" (*MF* 177).[64] By contrast, a "non-basic action is an action that we perform by performing a basic action" (*ibid*).[65] The basic action in Steiner's list is, presumably, the tensing of a forefinger: it is that which brings about all the remaining and allegedly different actions or events: there is nothing else that remains for the agent to do, if they are attempting to save hostages held by an assailant in that particular context. Reference back to that bodily movement (which need not be intentional) or action (which must), not only allows us to regard all the other, different descriptions of what has occurred here – pulling a trigger, aiming a gun, killing a man, saving four people – as different descriptions of the same 'thing'. It also shows the spatio-temporal specificity of the action or bodily movement in question: it was at that specific place, at that precise time, that X tensed their forefinger, initiating a series of events which culminated in the death of one person and the saving of four others.

To those troubled by what might seem to be a long looking causal chain connecting the tensing of the forefinger with the saving of four lives – it gives rise to the thought that these two things cannot really be either the same 'action' or 'event' – Carter replies by invoking a foreseeability constraint. He holds that basic actions, and those other 'actions' or 'events' which they causally generate, count as one action only insofar as the causally generated 'actions' "could in principle be foreseen at the time at which the agent has the degree of freedom under investigation" (*MF* 189), that time being, for our purposes, the time of the basic action or bodily movement in question.

There is another component to Carter's answer to the multiplicity of act descriptions and hence the supposed problem of a multiplicity of actions: the idea of the compossibility of actions or bodily movements:[66]

> "For two things to be compossible, they must both be members of a single possible world, which is to say that they must be possible in combination ... . If a set of actions is compossible, then there is a possible world in which they all occur. For any set of compossible actions, we can ask whether that set of actions is prevented or unprevented for (unavailable or available to) a given individual" (*MF* 180-1).

[64] Carter draws on Arthur Danto here: 'Basic Actions' (1965) 2 *American Philosophical Quarterly* 141-148.
[65] Steiner's answer is quite different: see How Free, n 57 at 75-76.
[66] The idea is Steiner's, as Carter acknowledges: H Steiner, 'The Structure of a Set of Compossible Rights' (1977) 74 *Journal of Philosophy* 767-775 and *An Essay*, n 57, ch 3.

When we calculate the degree of an agent's freedom or agency, then, we count only the set of actions or bodily movements that can occur in conjunction with one another: my raising my arm and you raising yours cannot occur if we are both crammed tightly together in a crush of people, but we can both shout for help if the crush is not too great. The compossibility constraint might seem to do little or no explicit work with regard to Steiner's list, but that is because the list is on its face already compossibility-compliant. In other contexts, it will limit the multiplication of actions by ruling out of our calculation those that cannot occur together.

Notice that in his characterisation of the compossibility constraint, Carter speaks of a 'set' of actions. This is not just because there will necessarily have to be more than one action in a compossible set, but also because he (alongside other quantifiers of liberty) holds that we must count act-types rather than act-tokens:

> "An act-type is a kind of action, which can be instantiated, or carried out in different ways – different events correspond with it … . The act-type of buying a book can be particularised in different ways – buying a copy of *Winnie the Pooh* in my local book store today, purchasing *The Critique of Pure Reason* at my university bookstore tomorrow, etc. Each of these more particular kinds of acts can be seen as instantiations of the act-type of buying a book".[67]

Act-types are sets of instantiations – a list of particular act tokens – of that broader category. And, while the number of instantiations may indeed be multitudinous, the number of act-types within which they are subsumed will be nowhere near as numerous.

So, how might we quantify agency? Following Carter, we must start with the basic act types (actions and bodily movements) and their foreseeably causally related 'consequences' that are available to agents: these acts occupy distinct parts of physical space and time. If those act-types are available to agents, in the sense that they are members of a compossible set of act-types, then we should count them. This seems straight-forward when we consider a slight variation on Carter's simple example, which concerns three actions: (i) walking down the street at time t+1; (ii) taking Z's glass of orange juice from her premises on the same street and drinking it at time t+2; and (iii) walking away from Z's premises at time t+3.[68] Assume that I can perform all three actions, but you cannot, because a geo-fence makes action (ii)

[67] van Hees, n 25 at 96.
[68] *MF*, 181.

impossible when you appear on the scene at time t+2. It seems intuitively plausible to say first, as Carter does, that "I am freer than you (in a purely empirical sense) in terms of the availability of these actions" (*MF* 181) and, second, as I affirm, that I have one more opportunity for agency than you. I can do actions (i), (ii) and (iii) while only (i) and (iii) are available to you. I therefore have more freedom and more agency than you.

For Carter, Steiner and other quantifiers, this process must also be accompanied by a count of all the basic act-types and their foreseeable causally related consequences which are unavailable to agents as the result of the basic act-types (actions and bodily movements) and their foreseeably causally related consequences of other agents. When we subtract the total of unavailable actions from the total of available actions, we are able to say how free any particular agent is overall; furthermore, this calculation can in principle be carried out within and between groups (*MF* ch 9). As quantifiers of agency, we might stop after the first step outlined in the previous paragraph, content simply to know how much agency there is at a particular time and place: if we return to the scenario that provoked the intuitive response, we could count the instances of agency within the context of a legal regulatory regime and do likewise within an (imagined) context of a technological management regime and compare the totals. But, if we are interested in assessments of the overall amount of agency available to agents, then we need also count the barriers to or constraints upon agency in both regulatory contexts, just as Steiner, Carter and others count constraints on freedom.

We can, then, set aside the third objection to the intuitive response on this ground: quantifying agency is in principle possible insofar as the quantification of freedom is possible.[69] Talk about the reduction, increase or redistribution of either notion is therefore not mistaken. Of course, showing that the quantification of either liberty or agency is in principle possible, which in this context means 'intellectually coherent and plausible as an idea', provides little guidance as to how to go about realising that idea in practice. The complexities involved in quantifying either notion in real world conditions are significant. An awareness of this leads Carter to suggest using two surrogate metrics to quantify freedom within the context of actually existing societies – something akin to the UNDP's Human

[69] There is no sign of Steiner or Carter giving up on the claim that it is in subsequent work: see I Carter and H Steiner, 'Freedom Without Trimmings', ch 13 of M McBride and V Kurki (eds.), *Without Trimmings* (OUP 2022); I Carter, 'The Myth of 'Merely Formal Freedom'' (2011) 19 *The Journal of Political Philosophy* 486–495 and his 'Choice, Freedom, and Freedom of Choice' (2004) 22 *Social Choice and Welfare* 61–81.

Freedom Index as a freedom metric and an exchange value metric (*MF* ch 10) – and, because of the close connection between agency, action and empirical freedom, both could be used by an agency quantifier. Somewhat ironically, the as yet not fully realised kind of densely networked society in which technological management would flourish would also be one in which freedom and agency could more easily be quantified: if everyone's every movement is monitored, then presumably every movement can be counted, too.

This topic – how the quantification of freedom and agency might be realised in actually existing societies – is one that must be pursued in depth elsewhere. My claim is only this: that something like the conceptual architecture outlined by Carter and Steiner for the quantification of liberty is also invaluable for the quantification of agency. The quantification of agency is on our agenda because the idea underpins the intuitive response and the two attempts to reduce or deny its plausibility. But the idea has wider currency: it is not just implied by what some might regard as an eccentric intuition invoked in a project charting the tension between two regulatory paradigms. Attempts to quantify agency have been made in other fields and are often accompanied, not just by the claim that agency can be quantified, but that its quantity can and should be increased to the benefit of human beings in various contexts.[70] That suggests, at the very least, that the idea animating the intuitive response is relatively widely shared and therefore worth taking seriously.

**3. Conclusion**

This essay gives reasons to think that a society which underwent a wholesale transition from the legal regulatory to the technological management paradigm would not be one in which human agency flourished, at least in the forms we currently know. Such a society looks, at first glance, like one in which there would be less rather than more human agency. This conclusion depends upon an ability to measure or assess agency's state of health and a subsidiary claim about agency's value and significance. While I have unpacked some of the steps involved in the first issue in the second half of this essay, I have said relatively little

[70] See the sources in n 58 above and, for example, A Tapal et al., 'The Sense of Agency Scale: A Measure of Consciously Perceived Control over One's Mind, Body, and the Immediate Environment' (2017) 8 *Frontiers in Psychology* 1-11. Note also that the Amartya Sen-Martha Nussbaum capabilities and functionings approach to freedom and justice is also, in effect, an attempt to measure agency (for one starting point, see A Sen, *Development as Freedom* (OUP 1999)), which is a central concern within development economics.

about the second, save to show that agency is closely connected to both our notion of autonomy and some of our understandings of freedom.

Is this conclusion significant? Yes, for three reasons. First, it is important if we value agency and, second, if we think it can be quantified or measured. Third, it is significant if we accept as true, as unfolding before our eyes, the story which was unpacked in the first half of this essay. That is a story about the direction of regulatory travel – from the legal regulatory to the technological management paradigm – in the jurisdictions with which we are familiar. That story constitutes a threat to agency as we now know it and it offers no guarantee of commensurate increases in other similar, or even qualitatively different, instances of agency. Furthermore, a threat to agency is also often a threat to those notions and ideals, like freedom and autonomy, with which agency is connected.

The stakes therefore seem high, but some might be tempted to reduce them. It could be maintained, first, that a reduction of the instances of agency we have focussed upon is no cause for regret since those instances of agency are unworthy. For they seem almost invariably to consist of actually doing, or having the opportunity to do, 'bad stuff' in this sense: acting contrary to the law. Why lament the reduction of or loss of opportunity to do that? Perhaps we could do so only if the importance and quality of human agency, either in general or in particular instances, is not solely determined by the moral quality of agential outcomes.[71] This entails a non-instrumental account of agency's value and the view that it can be good – morally significant – even when it is bad (because it yields outcomes that are morally objectionable). Second, some Others, quite rightly, might correctly point out that futurology is a fraught business and that few or no outcomes are guaranteed. Just as the switch from horse- to steam-power in the early stages of the British industrial revolution was not certain to occur, neither is the transition from legal regulation to technological management. Worrying about the future when it is far from certain might therefore seem silly. But it could also help us cope better with the changes that do come about, even if quite different to those that exercise us in advance.

[71] And, possibly, if there is a right to do wrong. A classic starting point is J Waldron, 'A Right to Do Wrong' (1981) 92 *Ethics* 21-39.