The Digital Disruption of Human Rights Foundations

Chapter · June 2019

DOI: 10.4324/9781351025386-6

1 author:

Hin-Yan Liu
University of Copenhagen
28 PUBLICATIONS **72** CITATIONS

SEE PROFILE

5 The digital disruption of human rights foundations

Hin-Yan Liu[1]

Introduction

Contemporary discourses have largely compartmentalised the human rights challenges posed by digitalisation and the internet as impinging upon the right to privacy,[2] or asserted that the freedom of expression must be preserved and defended online.[3] Thus, digitalisation and the internet are primarily perceived as information and communication media which pose human rights questions in relation to the freedoms elaborated within Articles 17 and 19 of the ICCPR.[4] Given the emphasis of human rights law upon state action or omission and notions of victimhood,[5] these types of direct challenges clearly occupy the foreground. But without diminishing the seriousness of the challenges posed to the protection of these discrete enumerated rights, this chapter explores some of the more foundational questions exposed by digitalisation and the internet upon the edifice of human rights itself.

This departure is justified because digitalisation and the internet have far more pervasive effects that perturb the broader social equilibria within which human rights frameworks are embedded. While much of the discourse on the human rights issues of the digital age is introspectively assessing the impact of the technology upon the protection of discrete rights, digital technologies are disrupting large swathes of society and the economy that have significant, yet indirect, impact upon human rights law. In this vein, the transnational nature of the internet, the digitalisation and hence traceability of everything, and the privatisation of governance have been highlighted as conceptual challenges to the universal human rights framework.[6] The disruptive impact of digital technologies also threatens the practical enjoyment of human rights in the real world by exacerbating global inequality, facilitating political repression, and opening new vulnerabilities for abuses.[7] Such lines

1 Associate Professor, Centre for International Law, Conflict and Crisis, and Coordinator, Artificial Intelligence and Legal Disruption Research Group, Faculty of Law, University of Copenhagen. Email: hin-yan.liu@jur.ku.dk

2 United Nations General Assembly (2014) The Right to Privacy in the Digital Age, 21 January 2014. A/RES/68/167.

3 United Nations Human Rights Council (2016) The Promotion, Protection and Enjoyment of Human Rights on the Internet, 27 June 2016. A/HRC/32/L.20.

4 United Nations General Assembly (1966) International Covenant on Civil and Political Rights, 16 December 1966. 2200A (XXI).

5 For example, see Article 34 of the European Convention on Human Rights, and the categories of direct victim, indirect victim and potential victim. See generally, Antônio Augusto Cançado Trindade (2011) *Access of Individuals to International Justice*. Oxford: Oxford University Press, 125–127.

6 Eileen Donahoe (2016) So Software Has Eaten the World: What Does It Mean for Human Rights, Security & Governance?, *Just Security* (18 March), www.justsecurity.org/30046/software-eaten-world-human-rights-security-governance/ (accessed 17.4.2018).

7 Eileen Donahoe (2016) Digital Disruption of Human Rights. *Just Security* (25 March) www.justsecurity.org/30225/digital-disruption-human-rights/ (accessed 17.4.2018).

of enquiry pursue human rights challenges indirectly and tangentially: digital technologies do not necessarily raise human rights questions as their most compelling problems, but their impact upon the enjoyment of human rights are more subtle and pervasive. In other words, the human rights impact of digital technologies may be both secondary and second-order. Thus, the result is that other technological features attract the limelight, and that much of the incursion of digital technologies upon human rights are overlooked and perhaps unintended consequences that are not related in neat and causally connected fashion. The looming threat of this type of shock to the foundations of the human rights law framework is exacerbated by the fact that these are not first and foremost questions of human rights. While this appears tautological, the difficulty enunciated by the societal disruption of digital technologies is such that it would be insufficient to deploy human rights mechanisms to confront them: a problem itself confounded because human rights have largely crowded out competing frameworks for understanding harms.[8]

Instead, this chapter occupies different vantage points to gain alternative perspectives upon the same underlying digital technologies and the capacities that these technologies enable. These perspectives underscore the internet as a *system*, as a *network*, and as a *distributor and dissipater*. The hope is that foregrounding particular features or characteristics of digital technologies and the internet will reveal different ways that the protection and enjoyment of human rights are being disrupted, thus paving the way for framing appropriate solutions.

Digitalisation and human rights from the systems perspective

A holistic approach to digitalisation, information and communications technologies would suggest the prospect for systems behaviour. After all, the internet is inherently an interconnected network that serves a purpose and which maintains an equilibrium over time, so the "system-ness" of digital technologies and the capabilities that it enable seem to be core features of this group of technologies.[9] To clarify, "a system is an interconnected set of elements that is coherently organised in a way that achieves something",[10] and what is pertinent for our purposes is that the behaviour of the system is an intrinsic property:

> A system is a set of things . . . interconnected in such a way that they produce their own pattern of behaviour over time. . . . The system, to a large extent, causes its own behaviour! An outside event may unleash that behaviour, but the same outside event applied to a different system is likely to produce a different result.[11]

This perspective suggests that incursions made into the realm of human rights by digital technologies may neither be intended, foreseen, or indeed foreseeable,[12] and that such incursions may be system behaviours as opposed to the actions or omissions of an agent. Both these observations make it extremely difficult to invoke the concept of human

8 'As a dominant and fashionable vocabulary for thinking about emancipation, human rights crowds out other ways of understanding harm and recompense', David Kennedy (2005) *The Dark Sides of Virtue: Reassessing International Humanitarianism*. Princeton: Princeton University Press, 9.

9 Note also that systems can be nested in other systems, and that the purposes at various levels are able to compete and conflict against each other, Donella H. Meadows (2008) *Thinking in Systems: A Primer*. White River Junction, VT: Chelsea Green Publishing, 15–16.

10 Ibid., 11.

11 Ibid., 2.

12 'In fact, one of the most frustrating aspects of systems is that the purposes of subunits may add up to an overall behavior that no one wants', Ibid., 15.

rights, let alone claim their breach. This is because, from a legally doctrinal point of view, there must be direct causal connection between the actions or omissions of a state or its agents and the concrete enumerated right possessed by an individual for the human rights system itself to become engaged.

First, foreseeability and predictability are key criteria relevant even for the enlarged conceptions of victimhood that have developed in human rights jurisprudence.[13] Even for the seminal cases which expanded the notion of victimhood under the jurisprudence of the European Court of Human Rights during the 1970s and 1980s, the notion of potential victim remain anchored in the legalised threats hanging over an individual's enjoyment of enumerated human rights. Finally, it should be noted that the jurisprudence underpinning potential victimhood have clear origins – a threatening law on the books – as to the source of the potential violation, which no longer holds with the impact of digital technologies on the enjoyment of human rights.

Second, if the impugned behaviour is an expression of system behaviour, rather than state action or omission,[14] then that behaviour appears to be removed beyond the realm of legal human rights protections. This is because part of the justification of human rights law as a mechanism to hold state power to account dissipates with the prospect for human rights violations to be "merely" the outcomes of system behaviours. Furthermore, the access to justice dimension that human rights law serves also becomes neutralised because a finding of a violation would not readily lead to reparations for the victims nor rectification of those violations for the future. In short, the causally- and agent-oriented nature of human rights law mean that it is largely toothless in the face of unforeseeable system "violations".

Another way of phrasing this is to ask what the meaning of human rights is, if human rights violations are relegated to the mere outcomes of systems behaviour. Recourse to the vocabulary of natural disaster and acts of God come close to the outcomes of system behaviours: events that remain outside of the capacity of human understanding and ability to influence. In such contexts, human beings are reduced to patients who must suffer the ill-effects without the ability to challenge or remedy the treatment. But with such externalisation of treatment comes the neutralisation of the very meaning of human rights itself because such treatment is rendered beyond the sphere of rights and responsibilities.

Human rights violations as normal accidents?

Viewing the human rights challenge of digital technologies from a systems perspective also raises the spectre of human rights violations as "normal" accidents, which are a small and exceptional class of technologically induced accident.[15] Thus, systems with potential

13 See the cases that are authority for laws that violate an individual's human rights by virtue of their being on the books, irrespective of their actual implementation or application to an individual, Trindade 127–129.

14 Unless the state itself is also considered as a system, which in reality is perhaps a closer analogy, rather than the contemporary approach of treating the state as an actor. Such a shift might close the gap between systems theory and human rights law. In this respect, it is worth noting that systems can be nested within other systems, Meadows 15–16.

15 Normal accident theory has been applied to understand a range of catastrophic technological failures. The essential claim is that accidents are unavoidable – and hence become normal – where systems are complex, tightly coupled, possess multiple objectives, and are embedded within competitive environments. Charles Perrow (2011) *Normal Accidents: Living with High Risk Technologies*. Princeton: Princeton University Press. For the application of normal accident theory to the governance of artificial intelligence, see Matthijs Michiel Maas (2018) Regulating for "Normal AI Accidents" – Operational Lessons for the Responsible Governance of AI Deployment. *Association for the Advancement of Artificial Intelligence*.

human rights impacts that also possess the underlying features of normal accidents may generate human rights violations as a natural matter of course.[16] In other words, a direct analogy would place breaches of human rights on par with spectacular normal accidents such as Three Mile Island, Challenger and Fukushima.[17] This direct comparison would transform the "risk" from orthodox accidents to risks in terms of human rights breaches. If such a comparison with normal accidents stands, breaches of human rights are endemic within complex, tightly coupled systems that possess competing objectives and which are deployed in competitive settings. Cyberwarfare might provide an example, insofar as technological or organisational vulnerabilities allow for vulnerabilities that result in the compromise or deprivation of basic services that are considered as core human rights, such as to health, water or food.[18]

The normal accident theory remains, however, focused upon discrete catastrophic events suggesting that alleged human rights violations that arise as a result would have a relatively discrete and singular origin. It may furthermore be possible to extend human rights violations beyond "spectacular" normal accidents by drawing an analogy with our recent typology of "boring" apocalypses.[19] The focus of normal accidents remains on the dramatic, but perhaps that is only where failure modes are the most visible. Thus, this focus on spectacular normal accidents need not constrain its application to the idea of human rights violations as a feature, rather than a bug, of technological systems enabled by digital technologies. Removing the threshold of spectacular catastrophe suggests that lower level normal accidents might be much more routine, but which largely evade detection and analysis.[20] If human rights violations inhabit this space of lower-threshold higher-regularity normal accidents, it would seem that normal accidents with human rights implications become more ordinary and mundane. Taken together, a third category of human rights-relevant activity might be described. In addition to intentional or active violations and negligent, reckless or otherwise omission-based passive violations, this third category encompasses human rights violations as attractors in a dynamical system. This third category hints at the possibility that a system may be predisposed towards certain value-states and that such attractors may map onto the sphere of outcomes that have traditionally been articulated in human rights language. In other words, it is possible that a system might lean towards breaches of human rights by its very nature or as a result of its particular configuration.

16 Perrow himself discusses what an accident is, and provides a typology of victimhood focusing on third-party innocent bystander victims and fourth-party foetuses and future generation victims, Perrow 62–71. Such a treatment demonstrates the difficulty of applying human rights law to such normal system accidents, since victimhood in law is generally not so accommodating.

17 Ibid., 15–31.

18 These examples deploy social and economic positive rights, which may be conceptually further from the core civil and political negative rights that are often discussed in the context of digital technologies and human rights. But this very much illustrates the point in that traditional human rights approaches fail to capture the multi-faceted challenges posed by digitalisation.

19 Hin-Yan Liu, Kristian Cedervall Lauta and Matthijs Michiel Maas (2018) Governing Boring Apocalypses: A New Typology of Existential Vulnerabilities and Exposures for Existential Risk Research. *Futures*, Vol. 102, 6–19.

20 This is analogous to claims that, because accidental nuclear wars have not occurred, that the danger of nuclear weapons accidents are minimal, Scott D. Sagan (1993) *The Limitations of Safety: Organizations, Accidents and Nuclear Weapons*. Princeton: Princeton University Press, ch 1. As Sagan himself recognised, the counterargument would be that, given the patterned flaws or structured vulnerabilities, such dynamical systems have failure as attractors. The mere insistence that such an accident has not occurred can then be answered by the success of local patches and serendipity.

This is important in the context of digital technologies because of the types of fixes that are implemented when human wrongs occur. Take, for example, Google Photo's tagging of black people as "gorillas" in 2015: more than two years later, the company has still evaded implementing a more lasting and systemic solution, opting instead to erase gorillas and some other primates from its lexicon.[21] This obviously superficial solution at the point of expression does nothing to rectify the underlying data set or the black box learning algorithms that arrived at such a conclusion. Digital technologies in this context exhibit a predisposition towards discriminatory effects:[22] this predisposition continues to lurk within the underlying data and data-processing mechanisms even as explicit wrongs that emerge are merely patched over.

Taken together, a systems perspective of digital technologies and their impact upon human rights suggest that infringements on fundamental protections might be expressions of system behaviours and normal accidents, placing such infractions beyond the realm and remedy of human rights law. But the digitalisation potentially provides more pernicious wrongs in laying the foundation or infrastructure that prejudices towards infringements against human rights or makes their occurrence more probable. This last point can be explored from the perspective of digital technologies forming networks in the next section.

Digital networks and the aggregation of human wrongs

Digital technologies enable rapid interaction and continuous communication that form the foundations of digital networks exemplified by the world wide web. The network perspective foregrounds two forms of human rights disruption: the patterned and cumulative incursions in human rights-relevant arenas; and the formal "horizontal" equality of the nodes in that network that diffuse the necessary "vertical" hierarchical relations that human rights law requires. The latter observation is discussed as digital diagonality through the lens of distribution and decentralisation in the third section of this chapter.

Human rights law requires an infringement to pass beyond a threshold of severity before a claim will be considered before a court or tribunal, thereby presenting a barrier against violations which occur through an incremental process. Where human rights law requires a significant single-origin action or omission that is causally connected to the alleged infringement, digital technologies enable a series or complex of more minor incidents to be structured in such a way as to amount to a more major interference. In the parlance of risk studies, a human rights "risk" is conflated largely with a human rights "hazard" that narrows consideration to one-hit knock-out interferences that pass

21 Furthermore, "A Google spokesperson confirmed that "gorilla" was censored from searches and image tags after the 2015 incident, and that "chimp", "chimpanzee", and "monkey" are also blocked today". Tom Simonite (2018) When It Comes to Gorillas, Google Photos Remains Blind. *WIRED,* www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/ (accessed 3.5.2018). This solution has also been deployed by other services encountering similar problems, Alex Hern (2018) Google's Solution to Accidental Algorithmic Racism: Ban Gorillas. *The Guardian* (12 January) www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people (accessed 3.5.2018).

22 It is important to note that such biases are often picked up from algorithmic inferences from collated human behaviours, Hannah Devlin (2017) AI Programs Exhibit Racial and Gender Biases, Research Reveals. *The Guardian* (13 April) www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals (accessed 3.5.2018).

the threshold for a violation.[23] Yet, a network enables patterned accretion of outcomes that, when taken as a whole, exhibit greater impact than an examination of the sum of the parts would suggest. Thus, the problem is akin to the structured accumulation of smaller wrongs, each of which would individually fail to surpass the threshold required to register as a human rights-relevant event. This argument is akin to the idea of structural discrimination that interconnected autonomous vehicle networks are likely to exhibit, through minor and subtle biases that lean in the same direction, if direct interventions to the contrary are not implemented.[24] Furthermore, there are cycling or recursive effects at play here: for example, getting turned down for a credit application may reduce your score in other systems, making it less probable in being able to secure an insurance policy, which in turn makes success less likely for a desirable professional position. Such vicious cycles may be precipitated by relatively minor and arbitrary trigger conditions.

A different way of phrasing this is that human rights procedures only filter out larger wrongs that are above a certain magnitude, reflecting the severity appropriate to the seriousness of human rights violations. What happens, however, if the smaller interferences that are not strained out by this filter pass through it and are thus not considered as human rights violations, but which afterwards coalesce into a mass that would not have been able to pass through the gaps in that filter in the first place? In other words, human rights law mechanisms that filter for severity at a single point will remain oblivious to the aggregation of more minor wrongs that lean consistently in the same direction.

Yet another way of approaching this idea is that human rights violations can be framed as a probabilistic function: that there is a given probability that a given system will impact upon an enumerated right of a particular individual. Such probabilities remain relatively constant regardless of whether a human rights violation actually materialises or becomes registered. That an individual's human rights are threatened probabilistically is akin to the expansion of victim status, discussed previously, where a discriminatory law on the books can underpin potential victimhood even if that law is not implemented. The difference in the digital context, however, is that the looming stochastic threats to human rights are not explicit and overt as with archaic legislation. The difficulty in detecting such probabilistic impacts upon one's rights posed through network effects constitutes a significant barrier to challenging and changing statistical infringements upon an individual's human rights.

Digital technologies may be especially prone to this consistent leaning or the structured accumulation of minor biases. This is because it provides the stable infrastructural basis in the form of big data which exhibits disparate impacts.[25] In other words, digital technologies allow for the coding, accumulation, retention and propagation of the biases inherited from previous decision-makers. Furthermore, the aggregation facilitated by networked digital technologies connects hitherto distinct and isolated areas which allow for these

23 The risk calculus is as follows: Risk = Hazard * Vulnerability * Exposure. See Liu et al. (2018).

24 Hin-Yan Liu (2018) Three Types of Structural Discrimination Introduced by Autonomous Vehicles. *UC Davis Law Review Online*, 149–180, available at https://lawreview.law.ucdavis.edu/online/vol51/51-online-Liu.pdf; Hin-Yan Liu (2017) Irresponsibilities, Inequalities and Injustice for Autonomous Vehicles. *Ethics and Information Technology*, Vol. 19, 193; Hin-Yan Liu (2016) Structural Discrimination and Autonomous Vehicles: Immunity Devices, Trump Cards and Crash Optimisation. In Johanna Seibt, Marco Nørskov and Søren Schack Andersen (Eds.), *What Social Robots Can and Should Do*. Amsterdam, The Netherlands: IOS Press.

25 Solon Barocas and Andrew D. Selbst (2016) Big Data's Disparate Impact. *California Law Review*, Vol. 104, 671; Karen Yeung (2018) Algorithmic Regulation: A Critical Interrogation. *Regulation & Governance*, Vol. 12, 505–523; Anthony Danna and Oscar H. Gandy (2002) All That Glitters Is Not Gold: Digging Beneath the Surface of Data Mining. *Journal of Business Ethics*, Vol. 40, 373.

disparate impacts to accumulate in ways that were not possible prior to the interconnectivity enabled by digitalisation.

From the perspective of the individual, the networked nature of databases can become a pernicious problem that places them in algorithmic prisons.[26] These are not literal prisons depriving individuals of liberty and which can be challenged through *habeas corpus*: rather these are algorithmic gatekeepers that influence access to important opportunities and services such as jobs, credit and insurance. While credit scores and other metrics have obviously influenced access to credit, for instance, these metrics have not hitherto been so networked and thus pervasive. An important factor is that it is no longer clear who is presenting these barriers and what behavioural aspects erected these obstacles.[27] Furthermore, there are again the cycling or recursive effects mentioned previously, which in the contextual framework of algorithmic prisons, can become especially problematic.

Indeed, the metaphor of a prison is perhaps overly monolithic since this suggests that there is a single authoritative jailor to be appealed to or challenged. Instead, this aggregation connects many hitherto disparate factors, temporarily uniting them to have the effect of jailing an individual by constraining her liberty and access to opportunities. A different way of putting this is that the algorithmic prison comprises of separate bars which are put in place by different databases: each one, taken individually, may be only a minor nuisance or a temporary inconvenience and be circumvented with some effort. The cumulative effect of these individual bars, taken together, is sufficient to frame a lasting cage.

This presents three human rights-relevant problems for the individual: first, that a whack-a-mole approach to perceived violations cannot provide a lasting remedy; second, that the same underlying data profile might underlie a range of different effects detrimental to individual freedom or well-being; and third that the data is often collected and retained by private entities that do not have direct human rights obligations.

Taking these in turn, the first issue is the absence of a feedback mechanism between the finding of a human rights violation and the prospect of a lasting remedy. Even if the contours of the algorithmic prison can be described and challenged as interfering with an enumerated right, any prescribed remedy will not affect the underlying data profile that was the source of that (and most likely other future) human rights infractions. Hence, wielding human rights law against the constraining forces of networked data can provide little more than a temporary and superficial salve to the latest infraction.

The second issue is very much the reverse side of the same coin: that networked digital technologies are robust and resilient to change. Even with the right to be forgotten and the required purging of data, it remains difficult to confirm compliance given the volume and pace of data collection.[28] Furthermore, data profiles can be reconstructed from data points held in other databases or can be points of convergence. Not only does tackling the outcomes as they emerge become a fruitless strategy, but even earnest attempts to alter the core data profile may be difficult to effect. This is not only because digitalisation provides the basis for near costless deposition of data, but also near costless duplication. This means that even if a human rights challenge to the manifest outcomes of a data profile were

26 Bill Davidow (2014) Welcome to Algorithmic Prison. *The Atlantic,* www.theatlantic.com/technology/archive/2014/02/welcome-to-algorithmic-prison/283985/ (accessed 26.4.2018).

27 'Even if an algorithmic prisoner knows he is in a prison, he may not know who his jailer is. . . . A prisoner might not have any idea as to what type of behavior got him sentenced to a jail term'. Ibid.

28 The right to be forgotten can also be readily circumvented by relocation to jurisdictions that adopt a more laissez-faire attitude.

successfully upheld, and the remedy required systematic eradication of the basis of data profile that gave rise to that infraction, in practical terms that data profile will persist in the network (in the form of duplications, earlier copies, or from convergent data profiles gathered by other entities on the same individual). This robustness underpinned by networked digital technologies ultimately requires a more systemic and continuous onslaught to remove. Human rights law mechanisms, however, provide only piecemeal, unsustained and isolated objections that will be unable to dislodge these disparate impacts arising from networked digital technologies.

Finally, there is the private dimension of data and digital networks: often it is private firms that collect and collate the relevant data that generate algorithmic prisons. This is further exacerbated by the fact that the walls of these algorithmic prisons are often built by other private entities, be these employers, creditors, or insurers. With no legal obligation beyond perhaps merely respecting human rights,[29] the absence of state power or state action makes it difficult to articulate these influences as human rights questions in the first place. A different way of putting this is that, in addition to the digital disruption of human rights, there are the familiar problems produced by private action that remove the immediacy of human rights obligations. Then, to top this all off, the digital network conforms to no jurisdictional boundaries making it even more unclear as to where any such claim might be made. This last observation is more readily apparent if the distributive and decentralising aspects of digitalisation are foregrounded.

Digital decentralisation: the diagonality of human rights obligations

Digital technologies have a flattening role, exemplified in the surveillance shift from Big Brother to Big Other,[30] and "in the complex personal data ecosystem that now exists, it is practically impossible for individuals to provide meaningful, voluntary consent to the sharing and processing activities which algorithmic regulation entails".[31] As digital technologies magnify the reach and impact of private organisations, the law, and especially human rights law, still treat these practically powerful companies formally as private entities.

Beyond exacerbating the public-private divide as it relates to human rights obligations, digitalisation also distributes decision-making authority by interjecting digitalised data-driven processes.[32] Even advisor and recommender style systems, however, can cause problems for human rights procedures for two converging reasons. First, there is the legitimating impetus of digitalisation providing the backdrop of better and more efficient performance.[33] This shifts the norm towards the deployment of algorithmic systems in a way that requires justification to opt out. Second, human agents or organisations might be relegated to bearing

29 John Gerard Ruggie (2013) *Just Business: Multinational Corporations and Human Rights*. W. W. Norton & Company.

30 Shoshana Zuboff (2015) Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, Vol. 30, 75.

31 Yeung 10.

32 Many 'concerns can be avoided if algorithmic systems are configured as recommender systems . . . rather than automated sanctioning systems', Ibid. 12.

33 See for example the contrasting opinions in the context of safety and autonomous vehicles, Aarian Marshall (2017) To Save Lives, Deploy Self-Driving Cars as Soon as You Can. *WIRED*, www.wired.com/story/self-driving-cars-rand-report/ (accessed 28.3.2018); Peter Hancock (2018) Are Autonomous Cars Really Safer Than Human Drivers? *The Conversation*, http://theconversation.com/are-autonomous-cars-really-safer-than-human-drivers-90202 (accessed 28.3.2018).

the brunt of accountability where human-technological assemblages go wrong.[34] This dispersal of decision-making echoes the claim made earlier about the meaning of a human rights violation that arises as an emergent outcome of complex interactions in a system. In other words, this dispersal of decision-making has decentralised the source of a human rights violation and thus obscured its author. Indeed, in holding up a mirror to examine the processes underlying a human rights violation, it may have knock-on effects that demonstrate a general lack of intentional authorship in human rights violations.[35]

The challenge is more than that introduced by mere horizontality, and is not easily remedied through positive obligations imposed upon the state.[36] Rather, decision-making authority persists such that power asymmetry invokes the idea of human rights, but that authority is no longer centralised in a way that is identifiable, and more importantly, challengeable. And the core challenge enabled by digitalisation is not one of horizontality in its orthodox sense, since a form of horizontality is introduced that maintains an asymmetry of power and the possibility of resistance. In other words, it is no longer a situation of orthogonal relations of verticality and horizontality reflecting state power and authority, but instead a *diagonality*, whereby nominally equal entities nevertheless exercise structural power and authority over those with the same nominal legal status. Such diagonal relations are necessarily difficult to contemplate through the lens of human rights law. The refusal or inability to do so, however, will leave unexamined and unchallenged the impact of decentralisation and dispersal enabled by digitalisation.

From one vantage point, digitalisation does not introduce new challenges because private corporations have been accruing wealth and power.[37] Digitalisation merely magnifies corporate power, extending and cementing the trends towards hegemony. From the perspective of digital technologies being distributors and dissipaters, however, a fundamental challenge to human rights protections emerges. As the law treats the corporation as a legal person,[38] the causal chain in any alleged human rights violation remains relatively clear and straightforward since the corporation as a distinct and independent entity is the source of the alleged human rights interference. That the law in its contemporary form largely insulates the corporation from human rights liabilities, so to speak, is a conceptually simple obstacle that is theoretically capable of being overturned. The distributing and dissipating functions facilitated by digital technologies, however, present a different layer of fundamental human rights challenges by breaking this neat causal chain. This compounds the formal legal difficulty of holding corporations to account for

34 M. C. Elish (2016) Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. *WeRobot 2016*; Wendell Wallach and Colin Allen (2008) *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.

35 The dilution of authorship resembles Arendt's famous observation on the banality of evil, Hannah Arendt (1994) *Eichmann in Jerusalem: A Report on the Banality of Evil*. Penguin.

36 This is most obviously because of the trans-jurisdictional nature of digital technology and its impacts, but also because it is not easy to define precisely what it is that the state has a positive obligation to providing or ensuring that would be legally actionable.

37 See for example the spectacular power wielded by the charter corporations, Nick Robins (2006) *The Corporation That Changed the World: How the East India Company Shaped the Modern Multinational*. London: Pluto Press. Indeed, such is the power of the corporation that it can be analogised with God, with corporate law serving as the system of mythology surrounding it. In this sense, corporate law serves the purposes of mediation and legitimation: 'The corporation is essentially a magical and mysterious entity that smooths over the contradictions in our culture and makes inequities seem natural'. Douglas Litowitz (2005) The Corporation as God. *The Journal of Corporate Law*, Vol. 30, 501.

38 *Santa Clara County v. Southern Pacific Railroad* (1886) 118 US 394 (US Supreme Court); *Salomon v. A Salomon* (1897) [1897] AC 22 (UK House of Lords).

human rights with the difficulty of dissipating the sources (the dilution of sources might even render these as just influencing factors) that lead to human rights infractions.

The compounding challenges to human rights posed by digitalisation would suggest that the orthogonal relations that have characterised human rights law must also be supplemented by diagonal obligations that recognise the practical and substantive power differentials rather than merely the legal form that an entity adopts. Yet, this diagonality will merely assuage a prior challenge because it only causally connects the corporation as a source of human rights-relevant activity. Where the sources or influencing factors that precede an outcome are dispersed or dissipated, as made possible by digitalisation, the question of causation itself becomes the core issue that conflicts with the very notion of human rights.

Concluding thoughts: digital infrastructure facilitating human rights disruption

This chapter has highlighted the disruption wrought by digital technologies to the very foundations of human rights. This was made possible by adopting three perspectives that digital technologies enable or facilitate, that are markedly different from the challenges illuminated by a more myopic focus on digitalisation itself. The following table summarises these fundamental challenges posed by digital technologies from these different perspectives.

	Effect	*Human Rights Challenges*
System	- Unintentional, unforeseen and unforeseeable behaviours. - Attractor or equilibrium states. - "Normal" accidents.	- Removal of outcomes from human rights scrutiny. - Systemic violations: system configuration may naturally predispose towards violations. - Neutralising the meaning and stigma of human rights violations. - A small subset of unavoidable violations.
Network	- Relatively minor, but patterned, structured, aggregated, or cumulative effects. - Duplication and replication of databases, convergence of data points. - Stochastic or probabilistic impacts.	- Private actors as a source or influencing factor of violations. - Cycling or recursive effects leading to vicious cycles. - Algorithmic prisons leading to structured denial of services and opportunities. - Causality focus problematic for probabilistic impacts.
Distributor/ Dissipater	- Blurs causation. - Exacerbates power differentials.	- Private actors as a source of violations. - Obscure sources and influencing factors of violations. - Compounds extant challenges of corporate human rights liabilities.

Networked digital communications technology has been claimed to give rise to or enable "a new system of social ordering known as algorithmic regulation".[39] Thus, the digitalisation impulse lays the infrastructure necessary for further forms of human rights

39 Yeung.

disruption that has been variously termed algorithmic regulation[40] or algocracy[41]. While digitalisation in itself does not present these threats, it lays the infrastructure that facilitates these developments. The difficulty of deploying human rights and other legal tools to confront the challenges posed by digital technologies is thus that it is not an action or omission, nor an agent or patient, that is the source of the problem. Instead, digitalisation provides the base conditions which are particularly difficult to engage with legally. Accustomed to interacting with actors and actions, law and regulation remains largely blind to the preconditions – the stage where the play is enacted out.[42]

The infrastructural perspective derives ready analogies with architectural modes of regulation[43] and architectural exclusion.[44] These analogies point to other types of difficulties: first, if digitalisation is akin to architectural or code modalities of regulation, then law is relegated to a parallel modality of regulation.[45] But what this means is that it is the combined effects of these different regulatory modalities which give rise to the regulation of behaviour as a whole, with the effect that the law is not supreme or capable of overriding other effects as legal doctrine would seem to suggest. Second, if the infrastructural perspective is analogous to architectural exclusion, a large initial hurdle is the presumption that artefacts[46] and architecture[47] are neutral. These presumptions undergird the received notion that the infrastructural level is not an appropriate level for regulatory interventions. This translates in the digitalisation context to immunising the infrastructural predispositions from regulatory scrutiny. While the focus upon actors and actions is consistent with legal doctrine and provides the continuity of the legal perspective, it overlooks the possibility that the digitalisation will create or facilitate winners and losers under this new system of social ordering.

Looking towards the horizon, digital technologies may impose impacts upon human rights protections in a more prospective manner. Big data facilitates the profiling of individuals and thus imposes subtle constraints upon personal freedom, as discussed earlier.[48] First, there is the collective judgement objection, that these constraints are imposed through a collective manner. Specific data points in an individual profile are statistically correlated with certain behaviours or outcomes of a larger group which shares those particular data points. Digitalisation thus allows collective probabilities to be transformed into predicted actions for the individual. Second, there is the prospective curtailment of opportunities and possibilities which are a long way from the retrospective nature of human rights violations that requires that the alleged breach had already occurred in the past. It is simply not possible to deploy human rights law to the possibility of probable future breaches.[49]

40 Ibid.

41 John Danaher (2016) The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy & Technology*, Vol. 29, 245.

42 Langdon Winner (1980) Do Artifacts Have Politics? *Daedalus*, Vol. 109, 121.

43 Lawrence Lessig (1999) The Law of the Horse: What Cyber Law Might Teach. *Harvard Law Review*, Vol. 113, 501.

44 Sarah Schindler (2014) Architectural Exclusion: Discrimination and Segregation Through Physical Design of the Built Environment. *Yale Law Journal*, Vol. 124, 1836.

45 Lessig identifies law, social norms, market and architecture as modalities of regulation, Lawrence Lessig (1998) The New Chicago School. *The Journal of Legal Studies*, Vol. 27, 661.

46 But see, Winner 1980.

47 Schindler 2014.

48 Davidow 2014.

49 Beyond those where anachronistic laws hold the prospect for discriminatory impact if enforced, as discussed previously.

This prospective dimension also creates difficulties to contain the possibilities of social manipulation and control engendered by digital technologies. The contemporary developments of China's Social Credit System[50] exemplify the disjuncture between technological realities and human rights law. The social credit system operates largely to restrict or facilitate access to desirable goods and services, and as such these are in the realm of social and economic positive rights (freedom to) that are notoriously difficult to enforce. Contemporary human rights law is much better at addressing civil and political negative rights (freedom from) that are oriented against preventing mistreatment. Compounding these matters further is the fact that the goods and services sought are often not considered within the substantive ambit of existing human rights and are also offered or provided by private actors. Thus, contemporary human rights jurisprudence is badly aligned towards attaining better standards or conditions in the context of social credit systems.

Social credit systems also manipulate individual behaviour, and perhaps individual identity, in the quest for a higher score. On its face, the individual engages with these systems in a voluntary and consensual manner that fails entirely to engage human rights law. But the effect is much more pernicious than mere nudging[51] because it complements a choice architecture with real, significant and consistent consequences. In doing so, such systems mould compliant and malleable individuals in a way that cannot readily be confronted with the paradigm of conflictual and coerced compliance that undergirds contemporary human rights protections. In this way, social credit systems many constitute yet another regulatory modality that runs in parallel to law,[52] that limit the regulatory potential played by law. Insofar as human rights *law* has crowded out other forms of understanding harm,[53] it also curtails the efficacy of human rights as a tool to push back against undue influence.

Digitalisation essentially provides the infrastructural basis for power, coercion, manipulation to flow. As long as the aim of human rights is oriented towards stemming such excesses, it has a role to play in the implementation and deployment of digital technologies. But, as the different challenges made visible from the system, network and distributor or dissipater perspectives make clear, there is significant work that needs to be done before human rights law mechanisms will be fit for purpose.

50 Rachel Botsman (2017) Big Data Meets Big Brother as China Moves to Rate Its Citizens. *WIRED,* www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion (accessed 3.4.2018); Mara Hvistendahl (2017) Inside China's Vast New Experiment in Social Ranking. *WIRED*, www.wired.com/story/age-of-social-credit/ (accessed 3.4.2018); Simina Mistreanu (2018) Life Inside China's Social Credit Laboratory. *Foreign Policy,* https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/ (accessed 26.4.2018). See also, Rogier Creemers (2018) China's Social Credit System: An Evolving Practice of Control (9 May). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3175792

51 Richard H. Thaler and Cass R. Sunstein (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press; Cass R. Sunstein (2015) The Ethics of Nudging. *Yale Journal on Regulation*, Vol. 32, 413; Cass R. Sunstein (2016) *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge: Cambridge University Press.

52 Social credit systems may also constitute elaborate and explicated forms of the social norm regulatory modality.

53 Kennedy 2005.