The conscience of an algorithm

Law, innovation, and the limits of human metaphors

TT Arvind
University of Newcastle upon Tyne

12 June 2018

THE STRENGTH OF HUMAN METAPHORS

- Law defaults to using human metaphors in shaping responses to new social forms
  - Makes sense as a heuristic
  - Range of regulatory tools available based on human metaphors
  - Inherent extensibility of these tools to new social forms
    - rules focused on outcomes ('contractual performance') easily applicable to new social forms
    - likewise for focused on standards of conduct (e.g. 'reasonableness')
- The use of human metaphors has been productive in the past
  - Classic example: private law regulation of the corporation
- But they have limits, which morally significant technologies instantiate

THE LIMITS OF HUMAN METAPHORS

- Human metaphors direct attention away from aspects of the social form that is unlike humans
- Treating these forms *as if* they were human fails to deal with issues specific to those forms
- Programmed systems do not act like individuals
  - Decisions based on *simplified* models (rather than on *accurate* models)
  - 'Satisficing' (rather than finding the optimal course of action)
  - Procedural rationality (rather than substantive rationality)
- Humans can behave otherwise. Systems cannot.
- Reliance on human metaphors draws attention away from specific problems these pose.
- Reframing the issue requires stepping outside the metaphor and considering the phenomenon *de novo*.

(RE)DEFINING THE PROBLEM

- Debates about 'AI consciousness' reflect reliance on human metaphors
- In legal terms, the problem relates to 'autonomous decision-making systems'
  - Systems which make decisions through non-deterministic processes
  - A human knowing the inputs and the criteria will not necessarily predict the decision
  - Algorithmic 'black boxes'
- Systems of this type are in use in a range of areas
  - Loan decisions
  - Parole decisions affecting the liberty of individuals
- Illustrates issues raised by morally significant technologies
- Solution requires a dramatic shift in approach
- Tentative suggestions inspired by natural resources law and by equity

THE CHALLENGE OF AUTONOMOUS SYSTEMS

- Algorithms making decisions which do not conform to legal standards
  - E.g. basing decisions on racial grounds
- Lack of transparency of bases of decisions by autonomous systems
  - Prisoners do not (and cannot) know what they need to do to get parole
- Reviewability of decisions by autonomous systems
  - Impossibility of subjecting algorithms to administrative law standards
- Contract as sole means of access to morally significant technology
  - Absence of contract justifies denial of access

Can the law give algorithms a conscience?

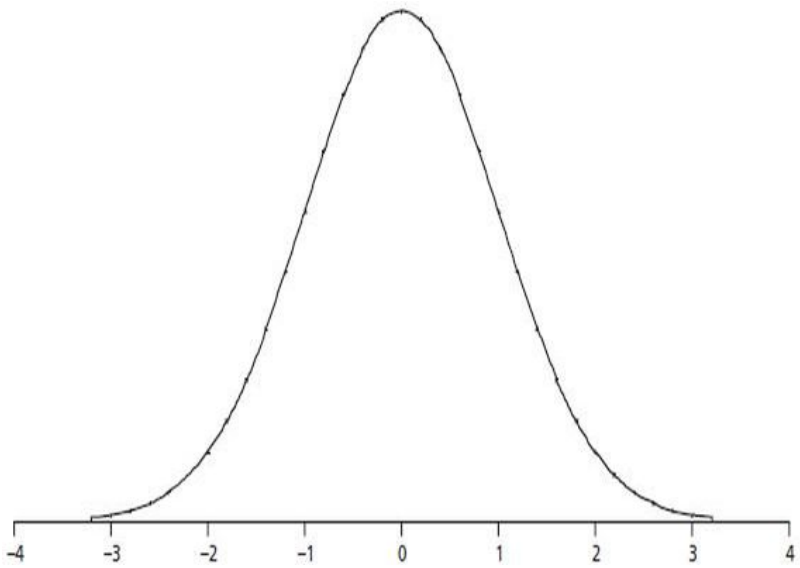
DEALING WITH AUTONOMOUS SYSTEMS

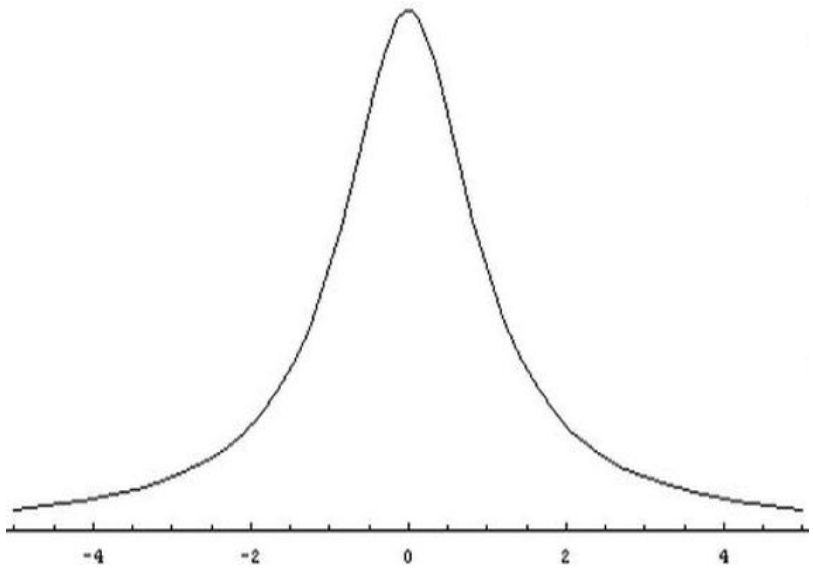
- Regulation suffers from two limits:
  - Regulators lack *technical* capability to audit algorithms and other new technologies
    - Reflects limitations of human metaphors
    - Reviewing human actions is a thing radically different from reviewing algorithms
  - Regulators lack *legal* tools to reframe the issue
    - Ability to reframe the issue is critical to regulatory control, but harder if human metaphors taken literally
    - Cumulative framework of contract and intellectual property entrench and legitimise a very narrow social vision (eg transparency)
    - Cannot move away from this without appreciating where the heuristic utility of metaphors runs out

THE PROBLEM OF REGULATORY CAPACITY

- The capacity to regulate assumes:
  - Ability of regulators to scrutinize the regulated system
  - Regulators having a roughly equal ability as the regulated to evaluate situations
- Both assumptions tend to hold where the actions under review are those of small numbers of humans
- Classic works of regulatory theory (e.g. Braithwaite, Grabosky, Ayers, Scott, Hood, etc.) deal with precisely such sectors
  - Mining safety, chemical industries, taxation, competition law, professions, product safety, etc.
- Neither holds true in relation to autonomous system.

AN ILLUSTRATION





EXPLAINING THE PROBLEM OF REGULATORY CAPACITY

- Underlying issues:
  - Inaccessibility of decision-making processes
  - Incomprehensibility of algorithmic code
  - Complexity of algorithmic code
    - Tracing nature and weight of criteria through routines, functions, library calls...
- Instantiated by:
  - Regulatory inability to scrutinise models during financial crisis
  - Regulatory inability to detect test-optimisation (Volkswagen, Intel...)

THE CHALLENGE OF REGULATORY REFRAMING

- Successful regulation requires the ability to reframe issues:
  - Bringing in aspects of interaction that are left out by the regulated community
  - Matters moved from purely commercial/economic framing to more socially embedded frame
- Braithwaite: Cultures of vice transformed into cultures of virtue
- Morally significant technologies require a reframing:
  - highlighting broader social interests at play
  - creating evaluative frameworks to assess if adequate account has been taken of those interests

HURDLES TO REGULATORY REFRAMING

- Regulators do not always succeed in this (Water industry, financial misselling...)
- Failures particularly common where decision-making is less individual and more system-based
  - Contrast chemical industry and mining with water industry and finance
- Consequence of combination of:
  - simplified models, satisficing, and procedural rationality, with...
  - ...legal tolerance of ruthless pursuit of self-interest

RUTHLESSNESS AND DERELATIONALISATION

- In the absence of a legal duty, persons are allowed to act in ruthless disregard of the interests of others
- Using human metaphors makes the subject of the metaphor the focus of the duty
  - (Not the humans involved in generating that metaphor)
- Law starts seeking to regulate technology rather than makers and users of the technology
- Owners of morally significant technology only regulated vis-a-vis the technology, not vis-a-vis community of individuals affected by the technology
- Evidenced by the absence of duties owed by algorithm-creators to those affected by the algorithms
- In effect, the technology is treated as a subject of law, rather than the underlying social relations between its owners and members of society interested in or affected by it.
- Approach legitimised by intellectual property law and contract law

THE EFFECTS OF DERELATIONALISATION

- Consequences:
  - Moral distancing
    - Autonomy and 'otherness' of the system's decision-making process distance the creator of the systems from responsibility for the outcomes the system produces
  - Entrenching alienated understandings of social relations
    - Law exacerbates rather than ameliorates the derelationalising effects of the intermediation of technology in human relations
- Addressing this requires moving beyond human metaphors, and finding new ways of conceptualising the relations between people and morally significant technologies

(RE)RELATIONALISING THE LAW

- Human metaphors are paradoxical
- Their effect is to distance the law from the actual needs and expectations at issue
- Solution lies in drawing the focus away from them and towards the underlying human needs
- Reach beyond the system to the individuals involved in the system
- Requires a new range of metaphors and regulatory tools to:
  - deal with the challenges posed by the creation of morally significant technologies, and
  - create a legal framework that compels companies and other entities involved in developing and using these technologies to have regard to the needs and expectations of those they affect.

BEYOND THE HUMAN METAPHOR

- Company law is too grounded in human metaphors to be of direct use
- Abandoning that metaphor points to other possibilities:
  - Public law approach: Analogy with natural resources
  - Private law approach: Analogy with fiduciary duties
- Both approaches put the focus of the duty on the relationship between the holder of the morally significant technology and those affected by the technology
- Engagement with law's hortatory function: setting and communicating standards to the regulated community
- Parallels with some developments in company law (e.g. CSA)

PERSONS OR RESOURCES?

- Can we substitute *natural* metaphors for human metaphors?
- Autonomous systems—and morally significant technologies generally—are not persons but resources
- Legally treated in ways analogous to the way in which natural resources are treated
- Energy law: General acceptance of states' power to govern even privately-held resources
  - Regulate use to minimise possibility of social harm
  - Regulate use to maximise and redistribute social benefits
- To classify as a resource is to assert that its significance is social, not private
- Holding a resource generates social responsibilities definable in legal terms

REDISCOVERING EQUITY

- Common law:
  - Starting point is ruthlessness unless specifically restrained by a duty
  - Duties are exceptional, and hence narrowly construed
- Equity:
  - Starting point is in the idea of conscience, responding to imbalances in social relations
  - Duties are purposive, and construed accordingly
- Equity historically played an active role in responding to the emergence of new social forms
  - Account responding to wardship, trusteeship responding to landholding patterns
  - Preoccupation of late 18th and 19th century equity with widows and orphans
- Preserving aspects of relationality challenged by the emergence of new social forms

TECHNOLOGY CREATORS AS FIDUCIARIES?

- Holder of morally significant technology fixed with duties to take account of the interests of individuals affected by that technology
- Developed by analogy to three aspects of the duty of a fiduciary:
  - Acting in good faith in the interests of the affected persons
  - Acting for a proper purpose
  - Not allowing personal interests to conflict with those of the affected persons
- Required to describe how those interests have been adequately protected
- Strict liability for falling short of the standard expected of a fiduciary

IN CONCLUSION

- Morally significant technologies pose a deeper challenge for the law than typically recognised
- The solution lies in understanding
  - the limits of technical regulation
  - the consequences of derelationalistion
  - the role of human metaphors in bringing these about
- Shifting our focus to less studied, but nevertheless promising, legal principles such as those underpinning equitable obligations or natural resources law holds more promise in finding ways forward.