Relational Approaches to Autonomous Systems Ethics

Luke Moffat
Sociology Lancaster University Lancaster, UK
l.moffat1@lancaster.ac.uk

ABSTRACT

Autonomous and/or Intelligent Systems (A/IS) are often conceptualised according to a model of autonomy characterised by an absence of interference, also called negative autonomy. What makes a system autonomous, according to this model, is the feature of independently giving a rule to oneself. Feminist critiques of autonomy, including relational critiques, challenge this negative model by drawing attention to the necessity of interdependence, connection, and entanglement. With that in mind, this paper explores how relational theories of autonomy help to speculate other futures for A/IS. It views A/IS not as discrete isolated individuals governed by negative liberty, but as interdependent, entangled constellations.

Considering A/IS otherwise, not as self-prescribing, isolated nodes, but as vast constellations of material, philosophical and political realities, has far reaching consequences for an individualist ethics that holds only single discrete individuals accountable. This paper explores some of the ways in which this might be possible, through concept of relational autonomy, and semiotics.

CCS CONCEPTS

• ; • **Human-centered computing** → HCI theory, concepts and models; • **Social and professional topics** → Cultural characteristics; • **Applied computing** → Sociology;

KEYWORDS

Relationality, Semiotics, Ethics, Feminist Philosophy, Autonomous and/or Intelligent Systems

ACM Reference Format:
Luke Moffat. 2023. Relational Approaches to Autonomous Systems Ethics. In *First International Symposium on Trustworthy Autonomous Systems (TAS '23), July 11, 12, 2023, Edinburgh, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3597512.3600201

1 INTRODUCTION

Academic literature on the subject of autonomous and/or intelligent systems (A/IS) ethics multiplies by the day. A common theme across this literature is an attempt to characterise the human-technology relation in a normative way.[1] What is sometimes neglected in this process is the more basic question; in what does this human-technology relation consist? The ways that autonomy is conceptualised in mainstream westernised ethics of technology often takes this relation for granted, as pre-defined and one-directional. This paper argues that relational approaches to autonomy, inspired by feminist ethics, can open other ways of doing A/IS ethics, and simultaneously, paint a more convincing picture of what autonomous systems actually are.

The aims of this paper are twofold: 1. to show how autonomous systems are already relational, and in so doing, 2. expand upon the prevailing conceptual and theoretical presentations of autonomous systems, and autonomous systems ethics, according to the negative or one-direction model of autonomy – as freedom from interference.

Relationality, in the context of autonomous systems ethics, seeks to collapse the distinction between human and technology by problematizing it. There is no clear point at which human beings end and technologies begin. Moreover, both are mutually co-constitutive – they create what STS scholars call socio-technical reality. There are two distinct but interconnected registers in which relationality is important here. First, technology does not exist in some abstract space separable from human motivation and actions; it is deeply bound up with human agendas, priorities, desires, and strategies of power. Second, relationality applies on a planetary scale. The material, conceptual, and occupational components that make A/IS possible rely upon a planetary network of natural elements couched as resources, exploitative labour practices, colonizing imaginaries of technological utopianism, and political landscapes which can legitimize these varying forms of violence [7].

There are many challenges in producing research on autonomous systems, including what to call them. In the last four to five decades of scholarship across Sociology, Philosophy, Computer Science, Design, and other disciplines, terminology has abounded for a new proliferation of technological change. Most commonly used today are terms like Autonomous Systems, Artificial Intelligence, along with more specific terms like Machine Learning and Neural Networks. All of these refer, in some capacity, to technological artefacts that are connected, and perform some or all of their assigned tasks, autonomously. There is not enough space here to debate what autonomy means, given its numerous affiliations with vastly different disciplinary perspectives. For the purposes of this paper, I use the term Autonomous and/or Intelligent Systems in an attempt to capture both concrete physical autonomous technologies such as drones, autonomous vehicles, and automated robots, as well as less material technologies such as AI, machine learning algorithms, and neural networks. A/IS can be both singular and plural, referring at times to a single connected device, at others to a series of devices that interact with each other.

Part of the relational perspective in this paper also seeks to build on the already established notion that the relationships between human beings and technologies span the material and non-material

[1] In the sense of normative values being applied to the relation, i.e. goodness, trustworthiness, efficiency etc.

TAS '23, July 11, 12, 2023, Edinburgh, United Kingdom

ACM ISBN 979-8-4007-0734-6/23/07…$15.00
https://doi.org/10.1145/3597512.3600201

registers. An algorithm is not necessarily "less" material than a drone. A/IS seeks to capture this blurred distinction between material and nonmaterial, not by claiming all things are matter, but by insisting that the distinction is performative [3]. While acting as a convenient catch-all, A/IS is also a vehicle of critique, in the sense that systems commonly placed under the labels of autonomous and/or intelligent, rarely are so, at least not in the ways they are presented [5]. In the case of AI, it is still a subject of debate whether intelligence can be defined according to 'correlation techniques' [8]. At the very least, it cannot be conclusively stated that intelligence, including artificial forms of intelligence, is reducible to the statistical methods used to inform it.

Despite these uncertainties around technological autonomy, there is a powerful narrative in numerous domains of computer science, which characterises the autonomy of a system according to the degree of freedom it enjoys from human intervention, the so-called negative model of autonomy.

2 NEGATIVE AUTONOMY

Autonomy has a long history in disciplines like philosophy and political theory. Many of the dominating assumptions about what autonomy is for technologies derive from these westernised traditions. In particular, the notion of negative liberty or autonomy, popularised by figures such as John Stuart Mill, Isaiah Berlin, and Milton Friedman, has had a huge influence on the way in which technologies are deemed to be autonomous. In simple terms, negative autonomy is characterised by freedom from interference. A person acts autonomously so long as their actions are not being deliberately impeded by someone or something else. Many of the philosophical foundations for this model of autonomy derive from Kant's critical philosophy.

For Kant, autonomy is an essential part of deontological ethics. According to this model, ethical decisions can only be made on the basis of a rational rule, which is independent from any outcome. This is an early modern example of the notion that one should do something simply because its right, unaffected by any personal motivations. Acting autonomously, for Kant, means acting according to a rule that one gives to oneself. These rules, or principles, derive from what Kant calls the 'pure' or 'rational' part of cognition. This means that they form the basis for experience, even though they never actually appear *in* experience (the a priori). In the ethical domain, more than any other, autonomy is vital for guaranteeing that actions are not carried out as the result of coercion or manipulation. Kant sometimes calls this 'autonomy of the will', and in the *Groundwork* he claims that 'autonomy is. . .the ground of the dignity of the human and of every rational nature' [12].

The self-legislative, or deontological, model of autonomy finds its way into various registers of A/IS discourse. Gyurky & Tarbell make these connections explicit in their work on foundational synthesis of autonomous systems, even going so far as to refer to a 'Noumenon network'; that set of processes that govern an A/IS' functionality, which are behind or beyond human computation [10] This 'Noumenon network' is precisely the unknowable space from which originate ethical decisions according to Kant. It is intimately connected with the idea that human beings (and here, A/IS too) can only make free decisions when they are not interfered with from outside forces.[2]

There is no real consensus in philosophical discourse, or, it seems, in A/IS discourse, about what counts as outside interference in the case of decision making. Nonetheless, there is a general focus on non-interference as the basic component of negative autonomy. In addition, there is a focus, in both cases, on the importance of rationality. Rationality is often a central assumption for agent deliberation.' [8]. This view of rationality entails that 'agents are expected, and designed, to act rationally in the sense that they choose the best means available to achieve a given end, and maintain consistency between what is wanted and what is chosen' [8].

The main advantage of a rationality assumption is that it can be applied very widely to a broad range of situations and environments. Assumptions of rationality can generate 'falsifiable, and sometimes empirically confirmed, hypotheses about actions in these environments. This gives conventional rational choice approaches a combination of generality and predictive power not found in other approaches.' [8]. For Dignum 'AI modelling needs to follow a social paradigm that can account for the reality in which human behaviour is neither simple nor rational, but derives from a complex mix of mental, physical, emotional and social aspects.' [8].

This involves, for Dignum, some kind of accounting for unforeseen circumstances, with which an A/IS would need to deal. But even current research on precisely this task, still follows a primarily rational, individualistic, negative conception of autonomy.

This view of autonomy often carries over to the ways in which A/IS are conceptualised. While not technically possible in the present, a shared imaginary in A/IS design is the prospect of autonomous agents carrying out tasks completely by themselves, with no or with minimally invasive human intervention. These imaginaries are often couched in terms of utilitarian ends [16], and so come with heavily normative associations. A/IS, particularly in terms of their social interactions, are discussed in terms of usefulness and optimality (ibid.) Making concessions to the human entanglements with technologies they use and are used by, The "human in the loop" is a frequent figure in technical disciplines. This hypothetical figure often fulfils a role of monitoring and evaluating the performance of an A/IS. The trouble is that the human in the loop is 'an impossible figure who can never meaningfully engage the plurality of posthuman doubts lodged within the calculus' [2]. The complexity of decision-making and "situational awareness" with which A/IS have to deal makes it difficult for a human in the loop to have efficacy, especially in situations where 'the steps of a normalized risk calculation protocol' have been followed 'beyond the limits of the calculable' [2].

The preceding applies as much to A/IS ethics as it does to A/IS design and use. Any prospective A/IS ethics that makes a claim for real-world impact needs to, as a minimum, confront its own assumptions about what autonomy is. The assumptions about autonomy being a primarily negative feature of rational decision making have wide-ranging and expansive effects on A/IS imaginaries, programming, and uses.

[2]The 'in-itself' domain of quantum computing, for example, contains within it registers of reality that human computation cannot access, at least currently. Kant explains the in-itself (*ding an sich*) as that which exists independently of all human experience [11]

Models of negative autonomy encounter difficulties when it comes to accounting for uncertainties, emergent features, and edge cases. This is because negative autonomy is by definition, non-prescriptive. It says nothing about what a given agent will or should do, it merely states that whatever that agent does, it should do so with minimal/no outside interference. This works well for providing accessible descriptions of A/IS, but the amount of ongoing intervention required to build in this kind of freedom from interference involves monitoring, modifying, and updating of multiple interconnected systems, such that an A/IS is never truly (meaning negatively) autonomous.

Relational autonomy provides one way of moving beyond purely negative models of autonomy, by focussing on the necessity of this interconnectedness, and its generative capacity. Feminist theories of autonomy that use relationality focus on the notion of human selfhood to make a case that one cannot merely be 'free from', they must also be 'free to'. The notion of being 'free to', which is tied to theories of positive autonomy, is not a defining condition of relational autonomy. Simply stated, relational autonomy means that even the process of recognize oneself as a self, is to simultaneously recognize your relationships to others (I would add both humans and nonhumans). Here, relationship signifies a straightforward connection, as in, involvement. In addition, it also signifies dependence. We don't just involve ourselves with others but we need them. So, even to deem someone/thing as autonomous requires some kind of intervention.

In what follows, I argue that not even a qualified negative autonomy model for A/IS is sufficient, either for capturing the realities created and inhabited by A/IS, or for fulfilling the ethical demands we might want to make of A/IS. A qualified negative autonomy that recognizes the need for intervention still cannot accommodate the need for creating and expanding positive benefits, nor can it account for the generative capacities of A/IS design and use. Relational autonomy opens possibilities for investigating what kind of realities are created when human beings do A/IS, and A/IS ethics. Before diving deeper into relationality, the following section summarises some of the ways that technological intelligence is conceptualised in negative models of autonomy.

2.1 Intelligence by negation

Is intelligence a speculative, or even aspirational, label? As Dignum summarises, current AI techniques, including deep learning and neural networks, succeed in 'perceiving images, written or spoken text', and picking out 'commonalities in these examples' [8]. Dignum continues:

> Many [theories of intelligence] characterise human intelligence as more than an analytical process and to include creative, practical and other abilities. These abilities, for a large part associated with socio-cultural background and context, are far from being possible to be replicated by AI systems, even if these may approach analytical intelligence for some (simple) tasks. [8]

Dignum suggests that there is something *sui generis* about human intelligence, something that is essentially beyond quantification. While this view can sometimes be overstated in discourses about human exceptionalism, the point remains that socio-cultural processes or 'backgrounds', are one space from which A/IS might be qualitatively cut off. Put simply, A/IS occupy other kinds of realities to those that give rise to socio-cultural processes. There are various locations – both physical and figurative – at which humans and A/IS encounter. That much is obvious. But there are some that do not overlap. It is beyond the scope of this paper to assess the validity of this view, but it is worth keeping in mind that at least some of the context of social-cultural backgrounds is beyond the reach of A/IS, while at the same time, A/IS are increasingly featuring in precisely those social-cultural contexts.

It is important to return to the centrality of rationality here. As Dignum writes that 'intelligent systems are expected to hold consistent world views (beliefs), and to optimise action and decision based on a set of given preferences (often accuracy has highest priority)' [8]. As has already been shown, however, human decision making, not to mention intelligence, is bound up with many more complex layers of social, cultural, behavioural, and environmental meanings. The attempt to transpose models of human intelligence onto machinic ones falters here, because the model was incomplete to begin with.

The disadvantages of assumptions of rationality have to do with universalism. Rational approaches to autonomy more generally assume that human decision making is rational all of the time, meaning, made by atomistic individuals, weighed in light of available evidence, prioritising an outcome that is maximally beneficial to the individual, rather than a collective or any other form of external cause. If, as has been argued [2] [5], decision making is not an exclusively rational process in human beings, then then there is no reason to expect that it should be in A/IS.

To investigate this further, it is important to show how AI/S systems 'are used to represent knowledge, what kind of knowledge and whose knowledge they contain' [1]. Speaking about AI specifically, Adam critically reflects on these questions about 'how AI is used and what knowledge it uses, rather than the possibility of true AI', claiming this as an important space for feminists. I take up this invitation for critical reflection as well, by gesturing toward at least some of the relations that are entailed by the design, production, use, and eventual disposal, of A/IS.[3]

2.2 Semiotics of A/IS

In addition to the concept of relationality, I also draw on semiotics as a way to describe how A/IS are imagined. Semiotics has multiple meanings in this context. On a basic level, semiotics is a branch of linguistics, focussed on the study of signs and symbols. As a branch of computational linguistic specifically, semiotics also refers to the programming of AI algorithms that detect, process, and configure spoken languages. There is a third sense in which semiotics is relevant here, as the study not just of sign and symbols, but of cultural imaginaries, and socio-technical assemblages [12]. In this sense, semiotics of A/IS concerns the systems themselves, as well as the ways in which those systems are imagined, speculated about, and presented in the media.

[3] I make the association between AI specifically, and A/IS more generally, on the grounds that many of the assumptions brought to light by Adam are not exclusive to AI, but travel across diverse and seemingly disparate spaces of A/IS production.

Winner's analysis of technological artefacts is helpful here. In his view, technological and political domains are always entangled with each other. He states, 'the available evidence tends to show that many large, sophisticated technological systems are in fact highly compatible with centralized, hierarchical managerial control.' [18]. This evidence, however, is often not counted in ethical debates around A/IS use, especially social use. One symptom of this is that, as Winner continues, 'people are often willing to make drastic changes in the way they live to accord with technological innovation at the same time they would resist similar kinds of changes justified on political grounds' [18].

The larger issue that Winner highlights is one of perceived technological neutrality. According to Winner, 'because technological objects and processes have a promiscuous utility, they are taken to be fundamentally neutral as regards their moral standing [17]. This goes hand in hand with the perceived non-prescriptiveness of negative autonomy. Because negative models of autonomy avoid defining normative rules or conditions for being autonomous, that precise state of being autonomous is also taken to be morally neutral. As mentioned above, the focus on rationality in negative models autonomy contributes to the assumption that, as far as decision-making goes, an action that is determined by rational choices has no ethical charge of its own. To put it simply, autonomy is seen as devoid of ethical value.

Taking a semiotic approach to these issues, as Winner arguably does, makes clear the centrality of assumptions about rationality and the morally neutral quality of autonomy in A/IS ethics. In the following, I present some of the imaginaries associated with A/IS, and A/IS ethics that help maintain these assumptions.

2.3 Technological and cultural imaginaries of rational autonomy

Winner's arguments show how technologies, including A/IS, are deeply bound up with political and culture imaginaries. There is, then, significant scope for engaging with these imaginaries as relational. This means that technologies in general are 'ways of building order in our world' [18]. How are orders created? In technological terms, 'one version claims that the adoption of a given technical system actually requires the creation and maintenance of a particular set of social conditions as the operating environment of that system' [18].

Winner recognises that this argument might be overstating the connections between technology and social conditions. As such, he presents a 'second, somewhat weaker, version of the argument', which holds that 'a given kind of technology is strongly compatible with, but does not strictly require, social and political relationships of a particular stripe' [18]. According to this second argument, renewable energies are more democratic than fossil fuels for example, not because of anything inherent in the technology itself, but because the demands of fossil fuel industries (expropriation of lands, extreme concentrations of wealth, ecological destruction) are less conducive to equality and justice than the demands and outputs of renewables. In Winner's terms:

> Solar energy is decentralizing in both a technical and political sense: technically speaking, it is vastly more reasonable to build solar systems in a disaggregated, widely distributed manner than in large-scale centralized plants; politically speaking, solar energy accommodates the attempts of individuals and local communities to manage their affairs effectively because they are dealing with systems that are more accessible, comprehensible, and controllable than huge centralized sources. [18]

A similar thing could also be said of A/IS. The demands of current models of A/IS design and use, are extremely unsustainable. From the minerals required to produce components, to the massive data exhaust from large tech companies, to the ecologically destructive e-waste practices. The argument could be made that this model, and the larger model of industrial capitalism, are associated in the way described by the second argument.

In both cases, there is either insistence upon, or suggestion of, an affiliation between certain technologies and certain political systems. Winner opens up a spectrum between compatibility and demand. Both of these arguments, then, are versions of the same basic claim, that technologies affect political organisation. What this claim misses though, is the possibility that a given technical system and a particular set of social conditions (including political systems) are co-constitutive, one is generative of the other and vice versa. The model of A/IS production is both constituted by and constitutive of, late industrial capitalism. Winner distinguishes between the internal and external dimensions of this technological-political relation, to describe the way certain political characteristics (like the hierarchy of a factory) exist within given technologies, while others (like solar energy being conducive to decentralisation) exist outside. Relational approaches, as described in the next section, do not insist on this inside/outside dynamic.

3 RELATIONAL A/IS

In feminist ethics, relational autonomy refers in part to the notion of the social self, which represents 'a dissatisfaction with the ideal of individual autonomy' [15] This ideal 'denies the inescapable connectedness of selves and the fact that their immersion in networks of relationships forms. . .their very identities' (ibid.). Adam's feminist critique of AI logics is also motivated, in part, by a need to 'look toward epistemological communities rather than individuals' [1]. While Barclay's work is concerned with human autonomy, and particularly the notion of the self, Adam's is specifically about AI. Epistemological communities need not be exclusively human communities; they can include other living beings, technological artefacts, discourse, organic and non-organic environments. A common thread across the human and AI discussions of Barclay and Adam respectively, is a motivation to examine ways in which knowledge is represented, to make knowledge visible as a production of power, rather than a neutral medium of creation.

The initial departure from negative models of autonomy considers knowledge as constructed, rather than discovered. The truths espoused by the verification of A/IS in public and commercial spheres are not universal or "natural" truths. They are generated by the same sets of techniques that create the technology. As such, Adam points to 'a certain distrust within feminist writing of a discipline which seemingly makes a pretence to neutrality where feminism has declared that none exists' [1]. This pretence to neutrality shares

common features with logics of negative autonomy. The assumption, for example, that everything exterior to human though is neutral space, rather than some with its own life, is widespread in A/IS literature. The impacts, both direct and tacit, of 'belonging to a particular culture' on A/IS production should not be understated. As Adam continues

> 'The crux of both the feminist and sociological arguments is that knowledge is a social, cultural product and epistemologies which rest on an invisible yet universal subject, and by extension AI systems based on these epistemologies, deny such a cultural plurality' [1]

This denial of cultural plurality is rooted in the rationalist logic of negative autonomy. As such, for Adam, 'the epistemology of symbolic AI is based on the Cartesian rationalist view that all knowledge is based on symbolic representations' [1]. The ideal type of representation in this sense is symbolic logic, where 'expressions in logic can be manipulated independently of their meanings. This means that symbolic AI can be seen as a giant Cartesian research programme which has attempted to discover the logical rules which comprise human knowledge and where the rules of logic offer a comforting certainty.' [1].

A lingering question here concerns whether it is possible to 'build systems to reflect these types of knowledge, whether it is feasible to reflect a true plurality of belief and what the implications are for continuing to pursue systems based purely on propositional knowing that knowledge'. [1]. To fully address this is beyond the scope of this paper, but it does provide a way in to understanding how relationality impacts upon A/IS and A/IS ethics.

To understand these A/IS relations more fully, it is useful to discuss some of material registers of A/IS production. "Relations" operate at various levels here. There is what might be called the black box level, where one attempts to visibilise and make sense of the interactions between material components and coded programming in a given system. Very crudely, how does an autonomous drone work? The very attempt to visibilise these relations from a perspective outside of those that create them, i.e. beyond the realms of computer science, software engineering etc. generates a conundrum. To "understand the inside" of a black box and communicate it in a way that people outside of computer science domains can access, requires translation. As Crawford shows, even a single command given to a single A/IS device entails a planetary network of data, material components, political schemas, and cultural imaginaries [5]

Birhane points out some of the ways that these relations are obscured by assumptions. One assumption is that any real-world hiccups in deployment can be solved with a technical fix. Some tweaking or rearrangement of the A/IS programming will cancel out any problem it encounters interacting with people and environments. The problem with this view is that it compresses ethical, social, legal and ecological issues in the tiny box of techno-solutionism – the view that all problems created by technology can be solved with more technology. As Birhane points out, relationality claims that 'neither people nor the environment and context in which they are embedded are static. What society deems fair and ethical changes over time and with context and culture.' [4].

If Birhane's arguments are convincing, and ethical values change with time and context, then relational A/IS ethics needs to make visible, not just the connections between technologies, societies, and planetary impacts, but also the ways in which these relations develop over time. This is by no means an easy task. In the following sections I suggest some ways this might be possible.

3.1 Making A/IS ethics matter

The preceding arguments are made in order to facilitate innovation in the field of A/IS ethics, and explore other ways in which autonomy specifically can be explicated. A relational approach to A/IS ethics has implications both for ethical discourse, and for the design and manufacture of A/IS. One implication is the necessity of transitioning from an ethics *by* design to ethics *through* design (Luján Escalante et al., 2022). Ethics through design implies that ethical values are not static or purely rational; rather they are contextual and responsive to how they are used. This means that ethical values cannot be simply implanted into a A/IS. Instead, ethical values form the basis of ethical *conduct*. Ethical conduct takes into consideration the flexible and dynamic nature of doing ethics, it opposes the idea that ethics is a tick-box exercise or a simple prescribing of rules. Instead, it asks people involved in A/IS production to critically reflect on their own practice.

3.2 Some implications

What does all of this mean for the field of A/IS ethics? In short, the implications of relationality are material *and* conceptual. In a material sense, relationality seeks to make visible to physical entanglements that are required to make A/IS work. These include the practices around sourcing components to manufacture A/IS technologies, as well as the physical enactments of A/IS in the worlds. In a conceptual sense, relationality looks towards the semiotics of cultural and political practices, and argues, following both Winner's and Dignum's claims, that technological, cultural, ethical, and political domains are mutually influential. Together, these domains generate specific types of reality, which create and are created by A/IS production. As Winner points out 'the construction of a technical system that involves human beings as operating parts brings a reconstruction of social roles and relationships' [18].

A/IS ethics, as just one place where A/IS are imagined and discussed, can be further developed by the perspectives mentioned above. By addressing the physical and conceptual entanglements that exist between A/IS design, programming, manufacture, use, and disposal, relational approaches broaden the awareness of costs and implications of A/IS production.

4 REFLECTIONS

At this point, it is useful to take a step back and reflect on relational A/IS ethics and its implications mean. Given the largely conceptual nature of this paper, there could be criticism of the lack of data or empirical findings, which would evidence the value of relationality. It is important to note, however, that relationality is a framework, a set of conceptual resources to analyse and describe A/IS production. It remains to be shown precisely how these resources might manifest in actual A/IS practices. Below are some advantages and potential limits to relation approaches to A/IS ethics.

4.1 Advantages of relationality

Relationality provides multiple lenses for making visible the entanglements of A/IS production. It emphasises the importance of this visibility, specifically by looking at places where A/IS entanglements are not addressed, or obscured. One concern shared between feminist ethics and non-westernised ethics such as indigenous protocols, is disassembling the hierarchy which places humans (meaning white men) at the top and nature at the bottom. In 'Making Kin with the Machines', this is made clear with the opening statement, 'man is neither the height nor the centre of creation' [14] Indigenous epistemologies are grounded in a view of nature that includes humans within it. They are often inherently relational, and relationality is 'rooted in context and the prime context is place'. There is no universal set of rules that can prescribe ethical technologies. Birhane argues that 'any data scientist working to automate issues of a social nature. . .is engaged in making moral and ethical decisions – they are not simply dealing with purely technical work but with a practice that actively impacts individual people' [4]. It can be convincingly argued that data scientists, among others, are indeed working increasingly with issues of a social nature, given the rapid proliferation of A/IS in social spaces. The contextual difference of these social spaces are essential for relational ethics.

In addition to showing or making visible the relations of A/IS, relational approaches also have a moral commitment. As Birhane says, 'relational ethics, at its core, is an attempt to unravel our assumptions and presuppositions and to rethink ethics in a broader manner via engaged epistemology in a way that puts the needs and welfare of the most impacted and marginalized at the centre.' [4]

Putting those most marginalized at the centre sounds straightforward on the surface, and various ethical A/IS initiatives claim this as one of their values. The work of considering peoples' needs and welfare presents its own ethical dilemmas, to do with power relations between those who have more or less agency within their social contexts. To be clear, this doesn't excuse complacency, particularly regarding exploitative labour practices for extracting materials that go into A/IS. It does, however, point towards some of the necessary limitations of relationality as just one of many approaches to A/IS ethics.

Indigenous protocols for AI offer one concrete way of approaching relational ethics. The figure of AI as *āina* is inspired by indigenous Hawaiian traditions, which make clear that 'humans are inextricably tied to the earth and one another' [14]. Instead of treating AI as either a slave or a tyrant, *āina* emphasises treating our relations as mutual connections of care and nourishment. *Āina* asks the question, how can we collectively craft 'beneficial relationships among humans and AI'?

4.2 Limits of relationality

Relationality is not a cure-all. In fact, the value of relational approaches to A/IS is precisely that is refutes the notion of a cure-all for ethical dilemmas. For disciplines that engage explicitly with the design and manufacture of A/IS, this could be seen as a limitation. Relationality problematizes the notion that different stages of A/IS design and use can be isolated from each other. Because of this, it could be argued that relationality makes "too much mess" of the topics it tackles, by connecting everything to everything. Crawford, discussing her own research practice with Vladan Joler, admits that trying to map out the relations of a single Amazon Echo was a hugely more complicated task than they originally anticipated [6].

There is a constant danger hanging over ethics of A/IS, that the newest ethical framework to come along becomes solutionist, and is reduced again to the prescription of rules. It is important to be realistic about what ethics can do and what it cannot. Relationality, as I have argued, makes important contributions to A/IS ethics, but it is not all-encompassing. In addition, there are still dimensions of A/IS production that, probably by necessity, follow negative, individualistic models of autonomy, in order to achieve certain functionalities. This paper has been asking what A/IS production would look like if there were ways to critique or change this model, particularly through a transition from ethical design to ethical conduct.

5 CONCLUSIONS

This paper has argued that relational approaches to autonomy can be useful for the field of A/IS ethics. It has argued that ethics is more than responsibility or accountability; rather ethics is important for scaffolding ethical conduct. One significant conclusion that can be drawn from this is that A/IS ethics, indeed any ethics, is not limited to specific, cordoned off domains. Instead, a relational A/IS ethics proposes itself as a way of doing, a process that can be taken up by computer scientists, designers, programmers, as well as community advocates, policy experts, and ethics researchers.

Beyond the arguments presented in this paper, there is scope for making and refreshing ethical commitments to making A/IS work for all. My aim in this paper was to show how, despite the levels of complexity involved, both in A/IS design and A/IS ethics, there is real urgency and, more, possibility, for doing A/IS ethics differently.

ACKNOWLEDGMENTS

This work is supported by the Engineering and Physical Sciences Research Council [grant number: EP/V026763/1].

REFERENCES

[1] Adam, Adam, A. (1995). A feminist critique of artificial intelligence. European Journal of Women's Studies, 2(3), 355-377
[2] Amoore, L. (2019). 'Doubt and the algorithm: On the partial accounts of machine learning'. Theory, Culture & Society, 36(6), 147-169.
[3] Barad, K., (2007). Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning. Duke University Press.
[4] Birhane, A., 2021. 'The impossibility of automating ambiguity'. *Artificial Life, 27(1)*, pp.44-61.
[5] Crawford, K. (2021a). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.
[6] Crawford, K. (May 2021b) "Kate Crawford on "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence", Youtube, accessed 08/03/23, https://www.youtube.com/watch?v$=$KcefG-0InLE&ab_channel$=$FineArtsMuseumsofSanFrancisco
[7] Crawford, K. & Joler, V. 'Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources' AI Now Institute and Share Lab, (September 7, 2018) https://anatomyof.ai
[8] Dignum, V., (2022). 'Relational artificial intelligence'. arXiv preprint arXiv:2202.07446.
[9] Fitoussi, D., & Tennenholtz, M. (2000). Choosing social laws for multi-agent systems: Minimality and simplicity. Artificial Intelligence, 119(1-2), 61-101.
[10] De Gyurky, S.M. and Tarbell, M.A., (2013). *The Autonomous System: A foundational synthesis of the sciences of the mind.* John Wiley & Sons.
[11] Kant, I., (2007) [1786] *The Critique of Pure Reason*, London: Palgrave
[12] Kant, Immanuel, Wood, Allen W, and Schneewind, J. B. *Groundwork for the Metaphysics of Morals. Rethinking the Western Tradition.* New Haven: Yale University

Press, 2002.
[13] Latour, B., 2011. 'Network theory| networks, societies, spheres: Reflections of an actor-network theorist'. *International Journal of Communication*, 5, p.15.
[14] Lewis, J. E., Arista, N., Pechawis, A., & Kite, S. (2018). Making Kin with the Machines. Journal of Design and Science. https://doi.org/10.21428/bfafd97b
[15] Luján Escalante, M.A., Moffat, L. and Büscher, M., (2022). 'Ethics through design'. *Design Research Society*. https://doi.org/10.21606/drs.2022.400
[16] Mackenzie, C. and Stoljar, N. eds., (2000). *Relational autonomy: Feminist perspectives on autonomy, agency, and the social self*. Oxford University Press.
[17] Shoham, Y., & Tennenholtz, M. (1995). On social laws for artificial agent societies: Off-line design. Artificial intelligence, 73(1-2), 231-252.
[18] Winner, L., (1986). 'Myth information: Romantic politics in the computer revolution'. in *Philosophy and Technology II: Information Technology and Computers in Theory and Practice*, pp.269-289.
[19] Winner, L. (1980). 'Do Artifacts Have Politics?'. *Daedalus*, [online] 109(1), pp.121–136.