

# VIRTUE, CAPABILITY AND CARE: BEYOND THE CONSEQUENTIALIST IMAGINARY

## EXTENDED ABSTRACT

The consequentialist ethical tradition, where the ‘goodness’ of decisions is assessed in relation to their measurable arguments, is often applied to reflections by technologists on the responsibility for creating new technologies such as artificial intelligence (AI), machine learning (ML) or connected systems (i.e. Internet of Things (IoT)). MIT Media Lab’s *Moral Machine* experiment, for example, approached concerns about ethics of connected vehicle systems by listing a series of moral conundrums that these systems are likely to encounter, and propose a ‘universal machine ethics’. This experiment, not only illustrates the challenges in universalising ethics through a consequentialist imaginary where the focus is solely on maximising utility whilst minimising harm; but also serves to stress the rational-individual as the key decision-maker in a solution-oriented mode of technology development. As such, the individual is expected to make structured decisions which would ultimately favour sparing the lives of humans over animals; a higher number of people over fewer; young people over the elderly and fit people over the un-fit and so on. Such thinking however leaves little room if any to attend to the contexts, structures and conditions that enable certain outcomes while removing others from being viable options; and rests on the fundamental assumption that the value of any decision is determined vis-à-vis its outcome.

In this paper, we intend to shift the focus from approaches that assess the morality of decisions made in the context of technology based on their actual or potential outcomes; and instead explore ethical theories that look at individual decision-making within the particularities of contexts, technologies used and relationships individuals are part of. Our approach is a result of our ongoing work with technology developers in the Internet of Things space in Europe where we have been quantitatively and qualitatively following the networks of developers, designers and entrepreneurs in order to understand the values that guide their decision-making during the design, development and marketing of their IoT products. As part of a large consortium of researchers, we have conducted fieldwork in Europe, followed hackathons, accelerator programmes, software and hardware showcases as well as immersed ourselves in meetup groups, and conducted interviews. We have also followed networks of IoT developers through online platforms such as Slack channels, Twitter and meetup. It is based on our ongoing research into the field of IoT that we developed the practical ethical framework that we present in this paper. In this extended abstract, we provide a short description of the ethical approaches that we are drawing upon for a broadly applicable ethical framework for practical ethics in technology design.

Virtue ethics claims that there is a kind of ‘final good’ which represents the desirable aims of someone’s life, and against which these aims can be evaluated. All questions attached to right action are assessed against this final good - known as eudaimonia. This means focusing on excellence, virtue, and eudaimonia, instead of duty, rights, and obligations, which were the typical concerns of popular consequentialist and deontological approaches. More recently Vallor (2016) applied a version of virtue ethics to the problems of technology, calling for a concerted collective effort to develop "technomoral virtues" that can guide the nature and direction of technical innovation in this rapidly changing world to ensure human flourishing. Virtue ethics draws with significant concern on the moral action of the individual and the role of community. Such an approach also offers a methodological opportunity to justify engagement with individuals and their articulations of values and principles as a legitimate pursuit. Yet in terms of identifying values, virtue ethics presents an interesting challenge. We have identified that the social milieu of (especially commercial) IoT development provides many constraints to people’s ability to act in ways that they might think of as ethical (Author’s paper). In particular, the idea of competing in a market or being subject to market pressures provides a particular constraint, which some people talk about transcending through their own personal work or actions or through the creation of alternative organizational structures such as technology trustmarks or manifestos. Part of the difficulty with virtue ethics however, is precisely its tendency to individualize the responsibility for virtuous action even if there is a role for communities in this process. According to MacIntyre (2007), a virtuous agent knows the correct way to act in various contexts while also desiring to act in such a way. This, however, is easier said than done, as several developers we have interviewed told us. They

have also indicated that when pressed with immediate challenges, it is not always straightforward to foresee what is to come and what kind of implications their decisions might have in the long run; or whether their decisions align with the values they hold as individuals.

In trying to understand how ethics manifest as values in action in the contexts of hierarchy and power, we have been increasingly concerned with the questions of what leads some individuals/groups to choose to act in a certain way and what might shape or constrain that choice of action. One important attempt to elaborate on this question has been provided by Amartya Sen in his capabilities approach (Hesmondhalgh 2017). Sen (1999) explains that “a person’s ‘capability’ refers to the alternative combinations of functionings that are feasible for her to achieve. Capability is thus a kind of freedom to achieve alternative functioning combinations.” This means that paying attention to individual’s internal capabilities is insufficient and we must also consider the possibilities created by a combination of internal capabilities and the structural conditions defined by the particular social, economic and political environment within which the individual attempts to act. This recognition that personal principles may need to be compromised to cope with structural constraints point to the importance of understanding what these constraints are and what influence they might exert. Furthermore, technology developers are in a curious position of both having to make decisions within the structural constraints of their context and having to acknowledge that the design decisions they make will result in producing structural constraints and possibilities for their users. Thus for developers to “do good” it is important to not only evaluate how existing constraints affect design but also to consider how these constraints are translated into the design and how these might be mitigated to offer more or different possibilities to the users.

The capabilities framework augments the internally oriented focus of virtue ethics on the moral capacities of the individual, by adding the importance of structural constraints. However, in both of these philosophical approaches decisions are made by individuals (even if within a social milieu) and it is individuals that must take responsibility, accounting for the constraints imposed by the broader social, political and economic contexts. Developers and designers of IoT technologies, just like everyone else, are certainly not alone in making decisions and in facing the consequences. Thus, we bring in *care ethics* to account for the value stemming from relational practice in considering different points of view as well as the possibilities of negotiating conflicts that arise between them. This enables including different points of view than the dominant discourses; such as those made by women or marginalized people who have not been part of the ethical discussion otherwise, and also for considering the ethics of practices such as caring which may have been absent from other readings. Joan Tronto (1993), for example, rejects essentialisms in gender and moral thought and advocates for contingent and historically situated definitions.

Individuals are always entangled in a diversity of relations that hold contradictory values and conflicting demands. In this paper, we bring these differing and at times conflicting demands in focus to illustrate both the complexity of the contexts in which decisions about emerging technologies are made and acted upon; but also how rather than the consequences, the infrastructures, relations and individual and community values shape the way these decisions come to be made. As such, we hope to provide an actionable and practical ethical framework for technology design and development that brings an alternative to the overly-dominant discussions of emerging technologies based on their potential [dystopian] outcomes.

**KEYWORDS:** ethics & technology, virtue, capability, care, consequentialism

## REFERENCES

- Annas, J., 1993. *The morality of happiness*. Oxford University Press. New York; Oxford.
- MacIntyre, A., 2007. *After Virtue: A Study in Moral Theory*, 3rd ed. University of Notre Dame Press, Notre Dame, Indiana.

MIT Media Lab. (2017). Moral Machine. <http://moralmachine.mit.edu/> Accessed 10 October 2019.

Tronto, J.C., 1993. *Moral boundaries: A political argument for an ethic of care*. Routledge, Chapman and Hall Inc., London.

Vallor, S., 2016. *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press, United States of America.