EDITORIAL

Drones, robots and perceived autonomy: implications for living human beings

Stephen J. Cowley[1] · Rasmus Gahrn-Andersen[1]

1 Introduction

This Special Issue explores perceived autonomy in drones and robots and its broader implications for human living. Building on Lindemann and colleagues' (2016) Special Issue of *AI and Society*, we turn from considering human–machine relations as "idealized one-on-one interactions" to focusing on the "incorporation of autonomous robots [and other machines] into everyday practices". Specifically, we investigate instances where human subjects attribute *autonomy* to their experience of artificial devices. We thus depart from how autonomy is usually approached in engineering science. Drawing on requirements for robotics, such views treat autonomy as a measure for artificial systems which, very often, is considered as an objective property pertaining to the systems themselves (Haselager 2005). This naturalistic variant of autonomy can be traced to how the new robotics aimed to design "complete intelligent systems" (Brooks 1991) that mimicked how organisms might achieve cognitive outcomes (e.g., McFarland and Boesser (1993)). Later, this work connected with traditions based on Maturana and Varela's (1980) view of 'autopoiesis' and its contribution to artificial life (see, Bedau 2003).[1] While no artificial system possesses autonomy in any naturalistic sense, much progress has been made in designing devices that make decisions without reference to human agents and thus, for naïve subjects, exhibit the appearance of autonomy.

An artificial system can trigger a perception of autonomy when, for example, a drone hovers overhead. In tracing this to an immediate experience (see, for example, Kim et al. 2016), we stress the importance of the idea of autonomy; one that is deeply rooted in Western philosophy and, specifically, the work of Aristotle and Kant. Given the influence of the tradition, it is widely assumed that any perceivable action has a 'source' that centers on an actor/agent. In some undefined sense, humans (and all living beings) are taken to act autonomously. Whatever one's view of such debates, the philosophical idea of autonomy has consequences for living human beings. It is important for sociopolitical reasons, for working with human–machine aggregates, for the future goals of AI and, therefore, for designers of devices. As illustrated by today's predator drones and the killer robots of tomorrow, the AI community is made up by sociocultural actors who have a great ethical and political responsibility.

The Special Issue attempts to shift the focus of discussion away from naturalistic—autopoietic or motivational—autonomy to implications of its phenomenological basis. Accordingly, it looks beyond how the appearance of autonomy affects the immediate perception of human subjects to address some of its many effects. In so doing, the Call for Papers invited the contributors to address one or more of the following questions:

1. In what sense are drones, robots and similar user-controlled devices 'autonomous'?
2. How do human perceptions of autonomy impinge on organizational, social and individual experience and action?
3. In a world where such devices have increasing practical importance, how shall we conceptualise their and

---

✉ Rasmus Gahrn-Andersen
rga@sdu.dk

Stephen J. Cowley
cowley@sdu.dk

1 Department of Language and Communication, University of Southern Denmark, Sdr. Stationsvej 28, 4200 Slagelse, Denmark

---

[1] McFarland and Boesser (1993) term this kind of autonomy 'motivational' and trace it to animals whose decision making attunes to circumstances and propose that the concept be applied to machines. Thirty years later use of *operative autonomy* serves to describe how non-deterministic artificial systems exhibit "a certain level of freedom in the choice of optimal means or given (determined) purposes in accordance with their efficiency and effectivity" (Hubig 2020: 27). For our purposes, there is no need to distinguish between the concepts.

our roles as actors (and entities) and its implications for designers of such machines?

In the papers that follow, perceived autonomy is thus related to not only how autonomy is perceived but also to interactional and situational outcomes, socio-cognitive organization, culture and, thus, the ethical issues that are central to AI. In setting the scene for the Special Issue, Rasmus Gahrn-Andersen's *Seeming autonomy, technology and the uncanny valley* (this volume) presents how the phenomenological category of autonomy appears in pre-predicative and pre-reflexive experience. Drawing on Heidegger (2010), Gahrn-Andersen argues that categorical switches between autonomy and heteronomy give rise to experiential changes similar to but yet more foundational than those of going from the immediate use of a tool (in its readiness-to-hand) to experiencing the tool as obstinate (in its presence-at-hand). Moreover, Gahrn-Andersen links these categorical switches to Mori's (2012) uncanny valley which, he suggests, depends on just such a violation of familiarity. Disturbances linked with the uncanny valley arise in encountering a hand that feels dead or, indeed, with androids or hovering drones. The experience arises, Gahrn-Andersen maintains, even in the case that a reliable supercomputer begins to act as if it possessed a mind of its own. In *2001: A space Odyssey*, what the crew knows of the world's ontologies is challenged. Their experience is disrupted by an epistemological change. While the phenomenological categories of autonomy and heteronomy are usually distinct, they can shift: where such a perturbance occurs a given device falls into the uncanny valley. Both perceived autonomy and the philosophical tradition are thus to be traced to the workings of pre-reflective human experience.

In *Attribution of Autonomy and its Role in Robotic Language Acquisition* (this volume), Frank Förster and Kasper Althoefer present an experiment with a social robot where the phenomenological appearance of the robot impacts on situated events. They use video evidence to show that people relate to the robot's facial expressions as the device manipulates items. As human subjects purport to teach it the names of items, they attempt to harness pre-reflective experience. As in the case of Gahrn-Andersen's work, visceral factors influence human responses: people use tightly coupled enactments of behavior including 'intent manipulations.' The teachers seek out marks of robot comportment that look *as if* the robot exhibited autonomous emotion, volition or understanding. In other words, they look for marks of perceived autonomy. For Förster and Althoefer, this shows the agent-centric or lopsided nature of human intelligence. Interestingly, the case applies in spite of massive differences in participant strategies: while some 'teachers' approach the robot rationally, others rely on robot displays such that the "prosodically most salient words are linked to affect or

motivation" (Förster and Althoefer, this volume). Some even exhibit empathetic responses such as talk about 'hurting' the device. The authors note that even a participant who tried to inhibit joint behavior fails: at a certain moment, they note, he "slipped in terms of self-imposed restrictions" (Förster and Althoefer, this volume). Emotion, in other words, links human in situ responding with a person's changing and immediate sense of robot action.

In considering self-driving cars, Florian Sprenger's *Microdecisions and Autonomy in Self-Driving Cars: Virtual Probabilities* (this volume) traces perceived autonomy to how Advanced Driver Assistance Systems (ADAS) currently implement its naturalistic or operative counterpart. While designed to enable road safety, they link uncertainly in the world to their own models. Hence, they cannot rely on wholly "deterministic processes" but draw on 'micro-decisions' that can give rise to perceived autonomy. Sprenger illustrates a case when a car brakes in response to a situation that is invisible to a driver but which the car successfully anticipates. Given this forward dimension, microdecisions are inseparable from the environment. The cars have a degree of self-management because, as with a robot's expressive movements, they rely on the choice of means. However, they make nontechnical choices of alternatives that use the quantity of information available in a virtual model. The car acts as part of a rapidly changing assemblage that includes the traffic (as well as algorithmic representation of people and the road): it is embedded in a system in which it has a co-constitutive role. Accordingly, the cars too set off impressions of perceived autonomy. Sprenger argues that we need a new grasp of how this is achieved: much of what we do—and what organisations enable/prevent—draws on micro-decisions that arise with humans outside the loop.

Seemingly autonomous technologies are not confined to the individual experience of devices in situated encounters. Quite the contrary. For as Garfield Benjamin shows in *Drone culture: perspectives on autonomy and anonymity* (this volume), the case of drones illustrates how their significance extends beyond experiential switches and embodied interactions. Indeed, in drawing on the appearance—and reputation—of autonomy such devices are important contributors to a socio-technical network of global surveillance and algorithmic decision-making. In Benjamin's terms, they have shaped a 'drone-culture' by their own logic. Powerful actors treat the military use of drones as a 'normal' way of exerting political power, building reputation and masking both the consequences of strikes and who bears responsibility. As 'stand-ins' for a global system, drones pursue imperialist goals and widen inequities. Drone dangers arise in that, first, controllers are often anonymous and, second, they always have an audience. Benjamin draws on Žižek's conception of parallax: as knowing by actor-subjects changes how they rethink

sociopolitical realities and the global order ('ontologies'). The anonymity of drones changes power-relations, shifts attention to artifacts and obscures the fact that the responsibility lies with human actors.

The next three papers focus on how philosophical ideas of autonomy relate to issues of ethics in AI. Jeff White's *Autonomous Reboot: Aristotle, autonomy and the ends of machine ethics* (this volume) and *Autonomous Reboot: Kant, the categorical imperative, and contemporary challenges for machine ethicists* (this volume) take on Tonkens' (2009) view that any project to build artificial moral agents (AMA) would violate the principles of Kant's philosophy. In arguing the contrary, White suggests that AI has the duty to build artificial agents that exhibit Kantian-style autonomy which makes them rational and free. It is striking that the debate falls on well-worn grounds. White argues that we could–indeed should–build machines as moral actors and that this does not lead to a violation of the categorical imperative. In so doing, he challenges how many working in machine ethics treat the autonomy of artificial agents as quite unlike that of natural agents. For the Kantian, ethics and agency depend on the seat of reason or the *mind*: an AMA should be not only rational but also fundamentally subjective. White's Kantian moral robots would therefore pursue, not common interests or those of communities, but outcomes that are consistent with universal, individual and voluntarist reasoning.

In *Could a Robot Flirt? 4E Cognition, Reactive Attitudes, and Robot Autonomy* (this volume), Charles Lassiter makes the case that moral judgements can only be traced to the embodied socialization of a citizen. Drawing on Aristotelian tradition, he doubts the feasibility of designing moral robots and, at once, questions the case for adopting Kantian criteria. Rather, playing up the importance of embodiment—and, by extension, appeal to pre-reflective experience—Lassiter links ecological views with the philosophical tradition. As Aristotle argues, a living human being acts ethically within a social context. On Lassiter's view, autonomy is not intrinsic (by contrast White stresses how moral motivation is not dependent on external factors) but, rather, fundamentally relational. Thus, if one were to have artificial moral agents (something about which he is skeptical), they too would endorse a community's needs: there can be no one moral imperative. In this case, we have different views of autonomy and thus what 'ethics' involves. Indeed, in the thought experiment of designing artificial model agents, one can readily see that agents with Kantian or Aristotelian autonomy would produce contrasting outcomes. The AMAs would differ in evaluating what is good –appealing, on the one hand, to society as a whole and, on the other, to a rational grasp of what is right. Lassiter's argument links the following three strands: (1) how we experience devices, (2) how we see their societal role and, equally, (3) how we regulate and motivate

designers. The analysis shows how much rests on the view that humans, at least, exhibit the autonomy of social beings. However, the sense of the term or, better, the philosophical idea of autonomy remains open to further debate.

In the final paper of the collection, *Autonomous Technologies in Human Ecologies: Enlanguaged Cognition, Practices and Technology* (this volume), Rasmus Gahrn-Andersen and Stephen Cowley present a framework that brings language into account and, by so doing, attempts to integrate themes that have arisen across the Special Issue. Specifically, they thematize the interrelation of (and possible continuity between) subjects' pre-reflective experience of technology, their practical activities and techno-cultures and what can be said. In terms of theory development, the paper establishes the importance of connecting, on the one hand, radical embodied cognitive science and, on the other, performativist approaches to Science and Technology Studies. In so doing, they show that the two strands of research share similar interest and commitments to nonrepresentationalism. At once, neither approach has yet extended these commitments to linguistic phenomena. In offering a positive argument, Gahrn-Andersen and Cowley suggest that, in humans, much cognition relies on language in the sense that human meaning-making activities are fundamentally enlanguaged. Pursuing the case of AUTONOMOUS DRONES, they show how this and similar inscriptions influence not only experience but also practices and culture. Indeed, independently of actual devices, it can exert enormous power. Given the enlanguaged nature of human cognition it can be resemiotized by, for instance, lawmakers, academics or engineers.

References

Bedau MA (2003) Artificial life: organization, adaptation and complexity from the bottom up. Trends Cognit Sci 7(11):505–512

Benjamin, G (this volume) Drone Culture: perspectives on autonomy and anonymity. *AI and Society*

Brooks RA (1991) Intelligence without representation. Artif Intell 47(1–3):139–159

Förster F, Althoefer K (this volume) Attribution of Autonomy and its Role in Robotic Language Acquisition. *AI and Society*

Gahrn-Andersen R (this volume) Seeming autonomy, technology and the uncanny valley. *AI and Society*

Gahrn-Andersen R, Cowley SJ (this volume) Autonomous technologies in human ecologies: enlanguaged cognition, practices and technology. *AI and Society*

Haselager WFG (2005) Robotics, philosophy and the problem of autonomy. Pragmat Cognit 13(3):515–532

Heidegger M (2010) Being and time. SUNY Press, Albany

Hubig C (2020) Benefits and limits of autonomous systems in public security. Eur J Secur Res 5(1):25–37

Kim HY, Kim B, Kim J (2016) The naughty drone: a qualitative research on drone as companion device. In: IMCOM'16: proceedings of the 10th international conference on ubiquitous information management and communication. Article No.: 91 p 16, https://doi.org/10.1145/2857546.2857639

Lassiter, C (this volume) Could a robot flirt? 4E Cognition, reactive attitudes, and robot autonomy. *AI and Society*

Lindemann G, Matsuzaki H, Straub I (2016) Special issue on going beyond the laboratory—reconsidering the ELS implications of autonomous robots. AI Soc 4:441–593

Maturana HR, Varela FJ (1980) Autopoiesis and cognition: the realization of the living. D. Reidel Publishing Company, Dordrecht

McFarland DJ, Boesser T (1993) Intelligent behavior in animals and robots. MIT Press, Cambridge

Mori M (2012) The uncanny valley. IEEE Robot Autom Mag 19(2):98–100

Tonkens R (2009) A challenge for machine ethics. Mind Mach 19(3):421–438

Sprenger, F (this volume) Microdecisions and autonomy in self-driving cars: virtual probabilities. *AI and Society*

White J (this volume) Autonomous Reboot: Aristotle, autonomy and the ends of machine ethics. *AI and Society*

White J (this volume) Autonomous Reboot: Kant, the categorical imperative, and contemporary challenges for machine ethicists. *AI and Society*