REVIEW ESSAY

Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond David J. Gunkel (2023), 246pp., $US45 paperback, MIT Press, Cambridge MA, ISBN: 9780262546157

Introduction

David Gunkel is a prominent scholar in the philosophy of social robotics and artificial intelligence. *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond* (*PTR*) is the thirteenth addition to a highly influential collection of monographs, including *The Machine Question* (2012), *Of Remixology: Ethics and Aesthetics After Remix* (2016) and *Robot Rights* (2018).

    PTR is a focused attempt to analyze (beyond the earlier *The Machine Question*) the strict philosophical and historical divide between persons and things by way of an entity which resists either categorization – robot. Gunkel takes the distinction between person and thing to reflect a harmful tendency in the dominant philosophical and cultural framework of Western, modern thought – a tendency to see the world through binary glasses, thereby inflicting a violent and false duality upon the world around us. Although he hints early in the book at the need for a third category to accommodate such entities as robots, by the end of the book, we get the sense that Gunkel is after something even deeper. He wants us to resist the urge toward ontological categorization altogether and to take up a wholly new way of seeing and interacting with others. In the seventh and final chapter of the book, Gunkel points to the source for this new way, the 'ethics first' philosophy of Emmanuel Levinas. An alternative, though admittedly less catchy, title for *PTR* might have been *From Ontology and Logic to the Politics and Ethics of Otherness*.

    While your reviewers are avowed analytic philosophers, lacking scholarly familiarity with some of the philosophical traditions which inspire *PTR*, we find ourselves highly sympathetic to the concerns which motivate its thesis, concerns which we will identify and respond to in the course of this review. It has piqued, though as a mostly deconstructive project (not yet satiated) our interest in how an ontology-free approach to ethics could assuage these concerns. We are eager to learn more about it in future work. Gunkel's writing is passionate and penetrating, yet accessible, and covers a substantial amount of the current terrain of the robot debate as well as its historical (social and scholarly) origins. In this light, it promises its reader (whether philosopher, legal scholar, roboticist or curious member of the public) an impressive and wide-ranging introduction to the relevant literature, as well as a window into future avenues for discussions and policies regarding the moral and legal status of robots.

    PTR is organized as follows. Chapters 2–5 present and assess the philosophical and historical (social, legal and moral) dichotomy between persons and things as well as the debate about robot status as one or the other. The aim of this discussion is to show that this debate is at a stalemate and requiring a third category. Chapter 6 addresses a candidate for this third category, that of robots-as-slaves, which Gunkel dismisses because of its negative baggage and the inherently problematic hierarchy of the master–slave relationship. Chapter 7 presents the new way forward by inverting the relationship between person and thing and offering a 'deconstructive alternative' which 'exceeds the grasp of the entire conceptual order' and is thereby meant to serve, we believe, as an alternative to categorization.

    We provide a detailed overview of the book, after which we identify some potential lines of criticism and develop these. We then focus on Gunkel's argument against social constructivist proposals in Chapter 6, and on the positive proposal of Chapter 7. These more focused criticisms take

up (a) our concern that existing proposals have more to offer than Gunkel thinks, and (b) our worry about how any new proposals can avoid the problems he attributes to existing ones.

Overview

Gunkel sets up the robot status debate in each of the first five chapters in terms of critics versus advocates, with a different target for each, the first being robots-as-persons. Chapter 2 opens with the claim that both camps agree on the status of things – as objects, not subjects; they diverge only about whether robots are things. Gunkel invokes the instrumental definition of technology given by Heidegger (1962), that it is a means to an end and a human activity (p.25). In the activity of human labor, our concern is the work rather than the tools which become part of us. This strikes Gunkel as particularly and problematically ethnocentric. The critics, he argues, simply reiterate the instrumentalist position. Robots lack the necessary intrinsic features which would merit individual social, legal or moral status. Rather than disputing the instrumentalist position, the advocates challenge the categorization of robots as things. They make robot status conditional upon the acquisition of certain necessary features. This, in turn, renders their approach vulnerable to the problems of the 'properties approach' (Coeckelbergh's term (2012) for the 'ontology first' approach to moral/legal status; see also p.37).

Both camps share the problem of determining what these necessary conditions (properties) are (rationality, sentience, subject-of-a-life, consciousness, etc.). What is worse, these conditions are context-dependent. The critics claim that robots do not or cannot meet these conditions: the advocates say that, even if they do not at the moment, in principle, they can. There is the further problem (even if we can agree on what they are) of how to detect whether these conditions are present. This, it would seem, is an insurmountable (epistemic) problem. To get around it, we must resort to our social practices which are rife with mindless sociality attributions and anthropomorphizing. In effect, both camps operate from within the framework of the properties approach. They agree that things are objects without rights or obligations, but they disagree about who counts as more than things. Gunkel argues that since robots resist reification, we should seek out alternative approaches.

Chapter 3 is devoted to the concept of a person. Traditionally, a person is defined by reference to a human being, though some human beings do not meet the full conditions of personhood. It is generally accepted that beings other than humans could meet these conditions. Nonetheless, there is no consensus about the necessary and sufficient conditions for metaphysical or moral persons. Legal personhood, on the other hand, is a strictly formal and neutral legal device which allows an entity or legal personality to have rights and duties. It is a device determined by real-world situations and relationships. Rights and, by extension, duties, are fundamentally social and relational (Hohfeld, 1920) and legal persons are subjects of these rights and duties. There are natural rights and legal rights. The former are taken to belong, by nature, to any natural person. The latter, however, are conferred or withheld at will via an exertion of power.

Gunkel goes on to show that there are no natural persons strictly speaking and that the task of the advocate is to show that robots are capable of achieving the benchmarks of a natural person. But this task is mission impossible, so to speak, because it would require surmounting the previously described insurmountable problems of determining and detecting these benchmarks. However, proceeding on the basis of artificial personhood gives rise to normative problems concerning interest, authority and power in the conferment of status and its associated rights and duties. Provided that human judges are making these determinations, there will be an inherent ethnocentric, anthropocentric drive behind them. When these determinations are made about non-humans, there is an inherent conflict of interest. This is the starting point of Chapter 4, which is devoted to considering whether robots ought to be treated as natural persons.

Critics of robots-as-natural persons argue that moral status attaches to biological history; only sentient or phenomenally aware beings can be subjects of moral concern or appraisal. This

gives rise to a deceptively clear-cut line which machines cannot cross. However, this line is challenged by such entities as monsters, chimeras and cyborgs, as well as, for that matter, technologically augmented humans (Haraway, 2008). Robots can also be experienced as living things, which is the case in Japan (Robertson, 2014). The binary distinctions between natural and unnatural, organic and non-organic, biological and machine are artificial constructs 'which are the product of and that patronize particular cultural traditions' and are patently hegemonic (pp.78–9). This affirms Gunkel's earlier claim that there are no natural persons, strictly speaking.

The advocates in this debate embrace the properties approach, according to which a person is that which possesses the properties of personhood. Rights and duties are, in turn, conditional upon demonstrably possessing these properties. At this point in the text, Gunkel probes the problem of detecting these properties more deeply and reviews the various proposals to avoid this problem (p.90). These 'work-arounds' include the 'theoretically minimal argumentative strategy' (Schwitzgebel and Garza, 2015, 2020), Coeckelbergh's (2010) appearance basis for moral status, Danaher's (2020) ethical behaviorism, Neely's (2014) better-safe-than-sorry approach (akin to Bess's (2018) presumed persons approach), Darling's (2021) indirect duties approach and Nyholm's (2020) owe-it-to-ourselves approach (to uphold the patient in humanity).

Chapter 5 returns to the theme of artificial or legal personhood. This notion of person, Gunkel points out, has less to do with what something actually is and more to do with how it is situated within social reality. The critics fall into two groups – those who dismiss and deny the very idea of robot status on the grounds of undesirability or inconceivability, and the rest. Gunkel dismisses the first group and focuses on the second. The second group comprises overlapping pragmatic and conservative arguments. The first kind targets the *telos* of law – whether conferring status to robots serves the objectives of the law. The law, he notes, is human-centered; accordingly, this kind of argument falls prey to European-Christian juridical humanism (p.109) and assumes an ideal of human supremacy (following Estrada, 2020).

The pragmatist arguments make a related objection to conferring robot status because of the risks of liability shields and accountability gaps (cf. Bryson *et al.*, 2017), but these issues, Gunkel claims, have been solved in the context of other artificial agents (e.g., corporations). The conservatives seek to maintain the *status quo* view of robots-as-things in order to prevent what they take to be an undermining of the legal system and to hedge against the potential negative outcomes of granting robots legal/moral status.

The advocates of legal/artificial personhood offer what we take to be a hybrid approach to robot status and some creative and potentially viable proposals. Like the pragmatists, however, these proposals tend to hinge on cost/benefit analyses. While for the critics, the costs outweigh the benefits, for the advocates, the reverse is true, and their proposals are meant to enable these benefits. These proposals (e.g., robots as trusts or LLCs) are, in this light, expedient and possibly *ad hoc* solutions. Gunkel suggests that they are all based on a view of robot-as-tool and thus as a means to other ends, where robots have only contingent value. Even more damning, per Gunkel, is that these proposals are just as speculative and conditional as the natural personhood debate. He concludes that there is nothing, at this point, to tip the scales in favor of either robot-as-person or robot-as-thing; hence the stalemate and the need for a third category.

Gunkel begins Chapter 6 by noting how the fictional monster, Frankenstein, challenges the person/thing dichotomy. He goes on to discuss various potential new categories, determining that the dichotomy is in fact a false dichotomy, and reviewing the legal literature in search of a third term. There are a number of authors who have proposed 'slave' as one such third term. They refer to the Roman legal concept of *peculium* – a set of money and assets granted by the head of a household to his slave for the purposes of conducting business on his behalf. This legal mechanism granted slaves a certain instrumental power within society without according them legal personhood (in contrast to contemporary corporations).

Something similar has been proposed for artificially intelligent artifacts: a 'digital peculium'. Gunkel criticizes these approaches for uncritically endorsing such a deeply, historically

troubled concept. Furthermore, he notes that the solution is at best only partial since the concept of digital peculium concerns only the civil law, not the criminal law, and that the extension of the master–slave relationship to robots and AI systems is culturally specific and distinctly Western. Gunkel also presents several other proposals to synthesize hybrid solutions without using the category of slave, but ultimately dismisses them on the grounds that they too are, in one way or another, rooted in that problematic category.

Chapter 7 presents the much anticipated way forward, past the stalemate between the critics and advocates of robot status, and altogether away from the false person/thing dichotomy. Gunkel warns the reader that his aim, true to the Socratic mission, has been primarily to show the need for a new way rather than to flesh one out in its entirety. Chapter 7 is thus the culmination of the deconstructive project, crowned with a glimpse of a wholly new way of approaching the robot question.

Gunkel opens the chapter with an interesting concession: robots *are* things. But they are not the kind of things which can be wholly distinguished from persons. Specifically, they are not 'merely objects for a subject, instrumental means as opposed to an end, or *res* in distinction to *persona'* (p.162). Robots are things, in effect, which deconstruct the existing logical order. Most importantly, from an ethical point of view, they prompt us to take responsibility for things that are situated otherwise.

The problem with robots-as-things, then, is not that they *are* things, it's the *kind* of things we take them to be. If we embrace a view of things which is in opposition to persons (the false dichotomy), then we inevitably end up with the wrong conception of thing. This problematic concept is thing-as-object, the counterpart of which is thing-as-subject. The solution, then, is to let go of this dichotomy and to reconceptualize thing as thing-in-itself: Thing. Understood this way, Thing is out of opposition to Subject (qua person) and is liberated from this unequal relationship.

Non-Western world views (e.g., Ubuntu and Confucian) do not impose inherently unequal relationships upon the physical world, with superior subjects and inferior objects. Nor do they assume that these categories are preordained; the Ubuntu tradition, for instance, sees personhood as something which must be achieved. These worldviews shed some insight on how we might do otherwise. As for how we can take responsibility for things, we need an ethics of things (p.170). Heidegger takes the first step with his idea of letting things be (or abandoning oneself to things), while Levinas takes the second by putting ethics ahead of ontology. Enter thus the ethics first approach and relational view. Relations are prior, Gunkel says, to the things related. Consequently, the moral and legal status of things does not depend on their properties, but on how they stand in relation to us, and how we decide, in the face of the other (following Levinas), to respond. It is here, in the face (or 'faceplate', as Gunkel says) of the other, where from our privileged place of power we can and should take responsibility.

Critical Remarks

General observations

Chapters 2–5 provide a thorough and accessible overview of extant positions in the debate about the moral and legal status of robots. The organizational strategy Gunkel employs, dividing these positions into two camps (advocates and critics), all of which embrace the properties view, is part of this accessibility. However, it is also part of what we take to be somewhat of an oversimplification and, in some cases, a distortion of these positions. Gunkel attributes the stalemate between these camps to their inability to agree on either the necessary and sufficient conditions of personhood or on whether robots do or could ever satisfy them. In any case, the epistemic problem of detecting these conditions looms ever large. While we take this to be a mostly fair characterization of the mainstream positions in the debate, we also think that more can be said on their behalf.

Gunkel often uses rhetoric that suggests that the person/thing dichotomy might amount to more than it actually does. For instance, he sometimes seems to suggest that the problems related to

the person/thing dichotomy might be indicative of something being askew with the law of contradiction; e.g., it might not be 'in touch with the complexity and exigency of facts on the ground' (pp.13–14). However, the law of contradiction allows that there are true dichotomies, such as person/non-person and thing/non-thing. The law of contradiction does not say anything about the relation between persons and things, but leaves open whether these categories are distinct or whether they overlap, and whether there might be something in the world that goes beyond both categories. To the untrained eye, it might, at points, seem as if Gunkel finds the exclusive–inclusive feature of our concepts, the feature that some things fall into their domain and some do not, to be problematic. One might mistakenly read Gunkel as if he were saying that the law of contradiction leads to 'the excluded middle' and thus makes our conceptual machinery incapable of dealing with such entities as robots, neither persons nor things.

In other places, Gunkel appears to suggest that everything that exists is either a person or a thing, and if something is not a person, it is a thing. A thing can be someone's personal property and thus treated however its owner likes. Even if we bypass problematic categories that are difficult to place on either side (such as those of events and attributes), this claim is obviously incorrect since there are things that cannot be owned, such as oceans, nation states, and integers, and there are things that, even though they can be owned, cannot be treated in arbitrary ways (such as animals, forests and historical buildings). Of course, Gunkel is aware of all this, as his previous books and other passages in this book show, so we wonder if there are some rhetorical devices at work here.

As a big-picture observation, *PTR* employs, theoretically and conceptually, a rather sophisticated and versatile apparatus. This sets the bar of deep understanding of the moral of *PTR* somewhat high. The discussions of *PTR* cover fundamental features of human language and thinking, constraints of human perspective, such as 'humanity first' (p.109), as well as serious ethical and political problems. Gunkel points at a deep conceptual problem which has to do with the fact that the person/thing dichotomy is deeply rooted in our language and thinking, from which it seems to follow that we are ill-positioned to accommodate robots in our unquestioned ontological order. It is not always clear whether the worry concerns two-valued logic, the fact that our concepts function in a 'violent' way in that they include some entities in their domain and exclude others, or that we humans cannot escape our conceptual apparatus and are thus imprisoned in our conceptual perspective, or that there are ethical issues and problems out in the world. It is not always clear, therefore, how Gunkel conceives of the explanatory relations among these general philosophical themes. It is important to know whether the features of our conceptual apparatus are being used to argue for the morally problematic features out there in the real world.

In Chapter 5, Gunkel appears to broach the more constructive part of the book. He discusses legal personhood and points out that 'person' or the concept thereof is actually a social-*cum*-legal construct. This makes the reader anticipate that Gunkel aims to find a solution to the moral and legal status of robots via the constructive line of reasoning. Indeed, such a status is a result of social and collective deliberation concerning a great variety of issues from our values and the aims of our institutions and societies, and what serves them the best. That is, it seems that Gunkel makes the transition from an ontological and metaphysical debate to a political, legal and ethical discussion via a constructivist move.

Robots-as-slaves

Although Gunkel does spend time in this chapter exploring socially constructive alternatives to the proposals discussed in Chapters 2–5, we find that he dismisses them somewhat unfairly, utilizing a family resemblance or guilt by association approach, characterizing them, in essence, as slavery proposals. We do not think that the proponents of the slavery concept (e.g., Bryson, 2017) are in any way blind or insensitive to the moral wrongness of its implementation or the horrific historical baggage to which it attaches. We do think that if these alternatives fail, it is rather because they fall prey

to the troubles with the properties approach. The heart of the problem with the concept of slave is the asymmetry in status between the slave and the master. This is something that Gunkel emphasizes, pointing to the Kantian concern regarding how assuming the role of master would affect our humanity (p.148).

Yet Gunkel presses much more heavily on the negative history of the real-world application of the concept to human beings. If the problem is lack of equality, then we are faced with the question of how the class of equals is determined. This is bound to lead to a debate about the relevant properties of the members of the class of equals. Historical development has been towards more inclusion, not necessarily towards big change, in the qualifying properties. In this light, the problems with these proposals are actually the problems of the properties approach.

Gunkel does not explore the social constructivist political, legal and ethical alternatives much further, though it is here especially where we think that there is more nuance and flexibility than Gunkel grants. Indeed, to our surprise, he returns instead to further metaphysical discussion about the features of our conceptual apparatus and metaethics from the point of view of developments in Continental philosophy. *PTR* ends up questioning and emphasizing the importance of questioning our conceptualization of our social living world. Some more analytically inclined philosophers may not be entirely convinced. Gunkel anticipates this.

The ontological reversal

In spite of Gunkel's criticism of the properties approach (in its various forms), it seems to us that an argument can be made to the effect that properties approaches are compatible with the relational view and practice-driven approaches more generally. It appears that the revolutionary import from Levinas's philosophy against the properties approach is the turn of the ontological order. We generally decide how to treat an entity based on our view of what the entity is. First, we decide what the properties of the entity are and then we decide how we should deal with it. In Levinas, we have relations without relata, and the properties of the relata are determined by how they appear in the relation. This is indeed a change of order and direction. However, from the point of view of analytic philosophy, one might still ask, how radical is the difference between this idea and the basic structure of philosophical social constructivism? According to the latter, we possess beliefs which become true by virtue of our believing them, and that there are entities which have (social and institutional) properties, not by virtue of their physical properties or character, but by virtue of our collective imposition of status functions upon them. Social constructivism in this sense can easily accommodate relational properties – something is a person only if that something is recognized by others as being a person.

Although Gunkel gives us good reasons to be dubious about the properties approach (or the ontology-first approach) to moral/legal status (some of which we, in turn, have challenged), the case against it is not complete without a robust conceptual alternative. And although Gunkel rightly points out that a philosopher's job is first and foremost to call the existing order of things into question, and explicitly concedes at the end of the book that the answer-seeker will be disappointed by its largely negative (deconstructive) project, this project itself nonetheless depends on there being – in principle – a way out of the problems attributed to the properties approach. Unfortunately, the positive account is not communicated sufficiently clearly for us to understand how it avoids the problems of the other approaches. Mere reversal of the relationship between whether what something is determines how it is treated, or vice versa, does not seem to eliminate either the problem of anthropocentrism or the importance of properties to our decisions about how others are to be treated or classified.

Conclusion

PTR is enticingly written and covers a broad range of philosophical themes. We are confident that it will be a staple among robophilosophers and, because of its accessible presentation and topical focus, among the wider public as well. For anyone wondering about the nature and moral or legal status of robots or why we should care about this, *PTR* provides an excellent overview of the state of the art. Our remaining worries concern whether (1) extant positions really lack the flexibility to accommodate robots, even in all their puzzling complexity, (2) there is a viable alternative to a human-centered way of doing ethics and (3) *PTR* successfully bridges the Continental philosophical paradigm in which *PTR*'s proposal is rooted and the analytic realm in which we reviewers reside. Such a bridge is possible, we believe, in existing analytically rooted social constructivist approaches.

Finally, while granting that Gunkel is right about the properties approach, we remain uncertain whether any alternative can escape the problems (e.g., human-centeredness, subjectivity, authority, etc.) he attributes to it. Until we have a fleshed out alternative, we can only, at this point, feel a sense of hope. But this in itself is no small achievement and one which *PTR* has succeeded in inspiring.

References

Bess, M. (2018) 'Eight kinds of critters: a moral taxonomy for the twenty-second century', *Journal of Medicine and Philosophy,* 43, 5, pp.585–612.

Bryson, J., Diamantis, M. and Grant, T. (2017) 'Of, for, and by the people: the legal lacuna of synthetic persons', *Artificial Intelligence and Law,* 25, pp.273–91.

Danaher, J. (2020) 'Welcoming robots into the moral circle: a defence of ethical behaviourism', *Science and Engineering Ethics*, 26, pp.2023–49.

Darling, K. (2021) *The New Breed: What our History with Animals Reveals about our Future with Robots,* Henry Holt, New York.

Coeckelbergh, M. (2010) 'Moral appearances: emotions, robots, and human morality', *Ethics and Information Technology*, 12, 3, pp.235–41.

Coeckelbergh, M. (2012) *Growing Moral Relations: Critique of Moral Status Ascription*, Palgrave Macmillan, New York.

Estrada, D. (2020) 'Human supremacy as posthuman risk', *Journal of Sociotechnical Critique*, 1, 1, pp.1–40.

Haraway, D. (2008) *When Species Meet*, University of Minnesota Press, Minneapolis.

Heidegger, M. (1927/1962) *Being and Time* (tr. Macquarrie, J. and Robinson, E.), Harper & Row, New York.

Hohfeld, W. (1920) *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, Yale University Press, New Haven CT.

Neely, E. (2014) 'Machines and the moral community', *Philosophy & Technology,* 27, 1, pp.97–111.

Nyholm, S. (2020) *Humans and Robots: Ethics, Agency and Anthropomorphism,* Rowman Littlefield, New York.

Robertson, J. (2014) 'Human rights *vs.* robot rights: forecasts from Japan', *Critical Asian Studies,* 46, 4, pp.571–98.

Schwitzgebel, E. and Garza. M. (2015) 'A defense of the rights of artificial intelligences', *Midwest Studies in Philosophy*, 39, 1, pp.89–119.

Schwitzgebel, E. and Garza, M. (2020) 'Designing AI with rights, consciousness, self-respect, and freedom' in Liao, M. (ed.) *Ethics of Artificial Intelligence*, Oxford University Press, Oxford, pp.459–79.

Dane Leigh Gogoshin, Pekka Mäkelä and Raul Hakli
RADAR Research Group,
Department of Practical Philosophy,
University of Helsinki
dane.gogoshin@helsinki.fi