This is a repository copy of *Robots are not just tools*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/121304/

Version: Accepted Version

Article:
Prescott, T.J. orcid.org/0000-0003-4927-5390 (2017) Robots are not just tools. Connection Science, 29 (2). pp. 142-149. ISSN 0954-0091

https://doi.org/10.1080/09540091.2017.1279125

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Robots are not just tools*

Tony J. Prescott

Department of Psychology & Sheffield Robotics, University of Sheffield, Sheffield, United Kingdom

Department of Psychology, University of Sheffield, Western Bank, Sheffield, S10 2TN.
t.j.prescott@sheffield.ac.uk

Abstract

The EPSRC principles of robotics make a number of commitments about the ontological status of robots such as that robots are "just tools" or can give only "an impression or real intelligence". This commentary proposes that this assumes, all too easily, that we know the boundary conditions of future robotics development, and argues that progress towards a more useful set of principles could begin by thinking carefully about the ontological status of robots. Whilst most robots are currently little more than tools, we are entering an era where there will be new kinds of entities that combine some of the properties of tools with psychological capacities that we had previously thought were reserved for complex biological organisms such as humans. The ontological status of robots might be best described as liminal—neither living nor simply mechanical. There is also evidence that people will treat robots as more than just tools regardless of the extent to which their machine nature is transparent. Ethical principles need to be developed that recognize these ontological and psychological issues around the nature of robots and how they are perceived.

Keywords: robot ethics; principles of robotics; ontological status of robots; perceptions of robots; machine intelligence.

---

* Prescott, T. J. (2017). Robots are not just tools. Connection Science. 29(2), pp. 142-149. http://dx.doi.org/10.1080/09540091.2017.1279125

At the heart of the EPSRC principles of robotics (henceforth 'the principles') are a number of ontological claims[1] about the nature of robots that serve as axioms to frame the subsequent development of ethical challenges and rules. These include claims about what robots are, and also about what they are not. The claims about what robots are include that "robots are multi-use tools" (principle 1) or "just tools" (commentary on principle 2), that "robots are products" (principle 3) and "pieces of technology" (commentary on principle 3), and that "robots are manufactured artefacts" (principle 4). The claims about what robots are not include that "humans, not robots, are responsible agents" (principle 2), that robots are "simply not people" (commentary on principle 3), and that robot intelligence can give only an "impression of real intelligence" (commentary on principle 4).

On first reading these statements seem straightforward assertions of obvious truths. I will argue that this is not the case. Instead, I will propose that these ontological commitments lack nuance, they assume all too easily that we know the boundary conditions of future robotics development, and they obscure or ignore some of the important ethical debates. If this is at all true, then progress towards a more useful set of principles could begin by thinking carefully about the ontological status of robots.

If we look at how the principles are presented, there seems to be an implicit process of induction at work that allows statements about what most current robots are, to be re-interpreted as statements about what robots must essentially be. For example the statement that robots as multi-use tools in principle 1, slips into the claim that robots are "just tools designed to achieve goals and desires that humans specify" in the commentary on principle 2 and to the statement that "robots are simply tools of various kinds, albeit very special tools" in the preamble. Whilst it is easy to agree with a general statement that robots are multi-use tools, especially in the context of a discussion about dual use (principle 1), the much stronger claim that robots are just tools, or simply tools, denies that they could sensibly belong to other disjoint categories.

Take the category of 'companion' for instance. There is a major effort around

---

[1] I treat these as ontological claims since the language used is descriptive ("are", "are not") rather than prescriptive ("should be", "should not be"). There are a good number of prescriptive statements in the principles (particularly about what robot designers *should* do) whose rationale builds on these ontological statements about what robots are.

developing robot companions that can provide social and emotional support to people as partially acknowledged in the discussion of principle 4. The category of tools describes physical/mechanical objects that serve a function, whereas the category of companions describes significant others, usually people or animals, with whom you might have a reciprocal relationship marked by an emotional bond. The possibility that robots could belong to both these categories raises important and interesting issues that are obscured by insisting that robots are just tools.

Indeed, consistent with the view of robots as tools, the discussion of robot companionship in the principles is quite dismissive, describing them as toys that could afford some pleasure to people who are unable to, or cannot afford to, keep animal pets. Robots are faux companions on this account that create an "illusion of emotions" and their intelligence is artificial and not "real". The faux nature of robot companions, it is argued, creates a real ethical problem in that robot companions are potentially deceptive[2] and so should be designed so that their "machine nature is transparent".

The ontological problem here particularly concerns the claim that robots could never possess psychological capacities such as "real" emotions or intelligence. What these are, in human terms, is hotly debated in the cognitive and brain sciences. There is therefore no *a priori* reason to suppose that these capacities must be unique to humans and could not be shared by machines. Indeed, there are counter-claims that robots, suitably configured, can have emotions (Fellous, 2004), whilst the future of artificial intelligence, as intelligence, has no obvious ceiling at below-human level. In other words, the distinction that the principles seeks to make between real and artificial psychological capacities may prove to be baseless over time.

A further problem concerns the assumption about how people will *see* robots—specifically, that robots will be seen as tools if they are shown in a transparent way.

---

[2] Principle 4 states "[robots] should not be designed in a deceptive way to exploit vulnerable users". But what counts as deceptive design? According to Sparrow and Sparrow (2006), for instance, any use of robots in a social setting (i.e. as a "companion") is deceptive, since, by their view, robots can never be genuine social actors. Principle 4 could therefore be used to argue that robot companions are by their very nature unethical (at least, in the context of vulnerable users). The commentary on principle 4 suggests a weaker interpretation, nevertheless, what this illustrates is that ontological claims about the extent to which artificial capabilities are real or not could have substantive ethical implications.

This could easily be wrong, for instance, people may anthropomorphise robots regardless of how obviously they are manufactured products. One reason to think this could be the case is the strongly social nature of our brains, and how easily our empathy is triggered by something that appears life-like. The Heider and Simmel (1944) animations of simple geometric figures (see fig. 1), show just how crude this information can be and yet we will still see intentionality, motivation, even emotion. Similarly, the invention of the Tamagotchi digital pet demonstrated that a simple 2-d animation of an animal-like creature can create a compelling urge to care (Levy, 2007).

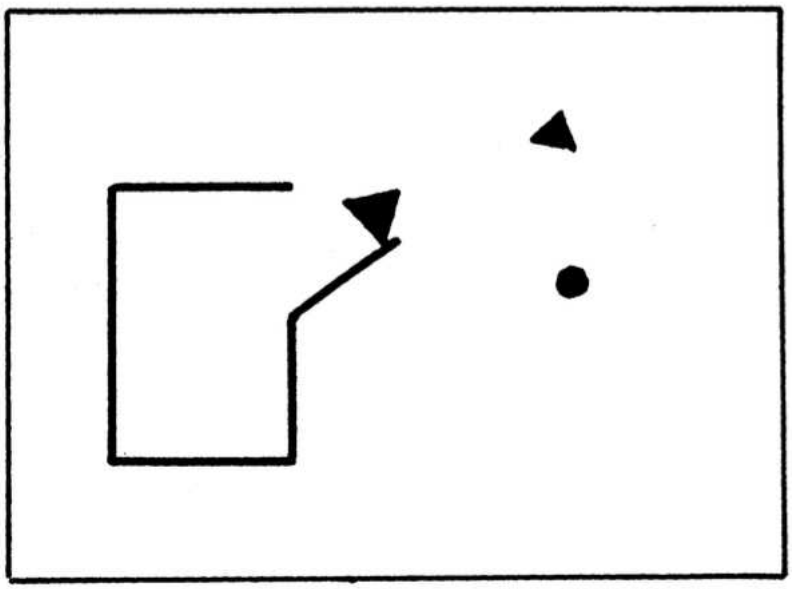

Figure 1. Geometric shapes moving around in a simple animation were interpreted has "animated beings, chiefly persons", in this famous 1944 study by Heider and Simmel.

The situation is further complicated by the fact that people can simultaneously hold multiple attitudes towards an entity such as a robot as characterised by Dennett's three "stances" (Dennett, 1987): the physical stance which views an entity as subject to physical laws such as gravity, the *design stance* which views a manufactured entity as acting according to purposes for which it was made, and the *intentional stance* which views a behaving entity as acting according to an internal set of beliefs and goals. Robbins and Jack (2006) added to these three levels a *phenomenal stance* that attributes consciousness and inner experience. There is some evidence that people will more readily see robots as having intentional states compared to phenomenal states (Huebner,

2010), suggesting a sophistication in our attitudes to robots that recognises them as more than machine but also less than human. This richness and complexity in how people will perceive robots needs to be given greater consideration in discussions of robots as deceptive devices and in proposals for greater machine transparency in robot design. It is clearly possible to view Heider and Simmel's animation from an intentional stance, or even a phenomenal stance, and not be deceived as to its real cartoon nature. Indeed, watching their animation invokes a "suspension of disbelief", just as when you are watching a movie or a play, that allows you to emotionally engage with the animation as an unfolding social narrative whilst also seeing it for what it is as a sequence of moving geometric shapes. Similarly, we do not need to believe (or be deceived) that the psychological states, intentional or phenomenological, that we read into an artefact, such as robot, are akin to our own in order to experience an authentic and meaningful emotional response[3].

A systematic analysis of ontological and psychological issues in human-robot interaction has previously been made by Kahn and colleagues (2007). Following a similar approach, we can describe four general ways in which ontological perspectives on *what robots are*, and psychological perspectives on *how robots are seen*, could combine. These are illustrated in the following table along with some of the ethical issues they entail.

[3] For further discussion of philosophical and ethical issues around the experience of emotions towards companions robots, and their similarity to emotions felt towards fictional entities, see Rodogno (2016).

I. Robots are just tools (o), and people will see robots as just tools unless misled by deceptive robot design (p). *Ethical issues: We should address human responsibilities as robot makers/users and the risk of deception in making robots that appear to be something they are not.* ***This is the position of 'the principles'.***	*II.* Robots are just tools (o), but people may see them as having significant psychological capacities irrespective of the transparency of their machine-nature (p). *Ethical issues: We should take into account how people see robots, for instance, that they may feel themselves as having meaningful and valuable relationships with robots, or they may see robots as having important internal states, such as the capacity to suffer, despite them not having such capacities.*
III. Robots can have some significant psychological capacities (o) but people will still see them as just tools (p). *Ethical issues: We should analyse the risks of treating entities that may have significant psychological capacities, such as the ability to suffer, as though they are just tools, and the dangers inherent in creating a new class of entities with significant psychological capacities, such as human-like intelligence, without recognising that we are doing so.*	*IV.* Robots can have some significant human-like psychological capacities (o), and people will see them as having such capacities (p). *Ethical issues: We should consider scenarios in which people will need to co-exist alongside new kinds of psychologically significant entities in the form of future robots/AIs.*

Table 1. How ontological *(o)* and psychological *(p)* perspectives on robots can combine (after Kahn et al., 2007). Note that only one quadrant of this table (I) is addressed in the EPSRC principles, but that II, III and IV are all possible, at least theoretically.

To conclude this essay I want to briefly consider some of the ethical issues that arise in quadrants II–IV.

In quadrant II, interesting questions arise about how robots should be treated—not because they are sentient agents but because people will choose to treat them as such. For instance, the idea that it should be unlawful to wilfully damage robots, proposed as part of the South Korean "Robot Ethics Charter" (Lovgren, 2007), or that we might mourn the loss of a favourite robot, as has been reported for some Japanese owners of *Sony Aibo* robot dogs (Brown, 2015), does not seem so strange when viewed from the perspective of how robots are seen by people rather than in terms of what they are. Of course, appearance and function do matter, but transparency of "machine nature" will only be one factor of many influencing how people see and behave towards robots, and it may be naïve to assume that it will be a decisive one. The bonds people will form with some robots may be similar to those we develop with other valued possessions, such as cars and mobile phones. On the other hand, for some robots, they may be more like the relationships we have pet animals, including for instance, wishing to support and nurture them (something that we may ourselves find rewarding). Finally, some human-robot relationships may share similarities to human-human relationships. For instance, I may develop a bond with my companion robot not because it looks human but because it has the capacity to remember and communicate with me about some of our shared experiences. More generally, what may be needed, in order to develop suitable ethical principles, is to develop a taxonomy of the different forms of emotional bonds that could exist between robots and people and analyse the factors that could underpin the development and maintenance of such relationships (Collins, Millings, & Prescott, 2013).

Quadrant III concerns the possibility of robots having significant psychological capacities that are in danger of being over-looked by people. This raises the potential for ethical risks[4] that are not discussed in the principles, but that have been highlighted

[4] To have any specific psychological capacity does not necessary imply any specific moral consequence. On the other hand, psychological capacities are at the core of many views of morality. For instance, having the ability to suffer has very significant moral implications for most people and comes up repeatedly in debates about, for instance, animal rights. Similarly the presence of consciousness is widely seen as an important consideration for ethical debates about the treatment of patients in coma or with locked-in syndrome.

by others. For instance, Metzinger (2009) has argued that we could build robots that are capable of experiencing suffering without realising that we are doing so, and therefore create a new kind of sentient entity that suffers unnecessarily due to our actions; many people would see this as ethically problematic if it were to happen. Although this may seem unlikely in the near-term, there are grounds to consider that this could be a risk in the medium- to long-term as cognitive architectures for robots become more sophisticated. Several trends in on-going research on human consciousness also support this possibility. First, one of the major contemporary theories of consciousness (Tononi, 2008) asserts a critical role for integration of information that does not necessarily require a biological substrate. Neurologists are also re-appraising whether islands of integrated activity in the brains of 'locked-in' patients might constitute a form of minimal consciousness (Qiu, 2007). Finally, there is an active debate as to whether animals with smaller brains than ours, such as fish, might be sentient in a significant way, for instance, that they may experience pain (see, e.g. Seth (2016)). These developments suggest that consciousness could be possible in an artificial agent without having to match the size or complexity of an intact human brain. Dennett (1994) has argued that "crude, cheesy, second-rate, artificial consciousness" (p. 137) could be possible in robots[5], and Bryson (2009) has proposed that today's robots might already have some simple forms of consciousness that meet some commonly proposed criteria[6]. None of this is to claim that we are in quadrant III yet, but given the risks, ethicists should be pressing us as to how we would know if we were.

One of the consequences of the view of robots as "just tools" is the implicit dismissal of the possibility of strong AI—that future robots could have human-level, or beyond human-level general intelligence. A quadrant III/IV issue, recently discussed by noted scientists and innovators such as Stephen Hawking, Elon Musk, and Bill Gates, to name a few, and analysed in-depth by Bostrom (2014), is that an AI 'singularity' could reverse the master-slave relationship between humans and robots. The conviction that

---

[5] Indeed, Dennett (1994) specifically considers, and dismisses, the proposal that robots cannot have significant psychological properties such as consciousness simply because they are artefacts (contradicting Principle 4).

[6] In subsequent papers Bryson has argued that consciousness per se may not qualify robots as moral patients and that we should avoid building robots that have psychological capacities, or other properties, that would qualify them as such (Bryson, 2010, 2012). This view, and the active philosophical and scientific debates it relates to, speaks to some of the ethical questions raised in quadrants III and IV.

robots/AIs are "just tools" may keep us from recognising the early signs of such a self-bootstrapping super-AI. An ethical approach would surely encourage more vigilance. A more positive quadrant IV stance on the AI singularity debate is the perspective of the 'global brain', proposed by Heylighen (2002) and others, that humans and advanced AIs could co-exist to our mutual benefit. This reminds us that that ethics must be about analysing the potential benefits as well as the risks.

Although quadrant III/IV scenarios may seem far-fetched or at least distant, such concerns have captured the public imagination and have prompted significant calls for debate (e.g. Future of Life Institute (2015)). In my own experience of talking to members of the public, and of the media, these are often the topics about which there is the greatest interest and concern. The attempt to create a rhetorical barricade against these issues by insisting that robots are just tools may do little to calm the voices and could come across as hegemonic and condescending. Whilst approaches to these longer-term ethical challenges are necessarily speculative, a starting point is to acknowledge that there are concerns here that are worthy of further attention.

A more candid approach may be to recognise that, whilst most robots are currently little more than tools, we are entering an era where there *will* be new kinds of entities that combine some of the properties of machines and tools with psychological capacities that we had previously thought were reserved for complex biological organisms such as humans. Following Kang (2011), the ontological status of robots might be best described as *liminal*—neither living in quite in the same way as biological organisms, nor simply mechanical as with a traditional machine. The liminality of robots makes them both fascinating and inherently frightening, and a lightning rod for our wider fears about the dehumanising effects of technology (Szollosy, 2016).

The Association of Manhattan Scientists wrote in 1945 of their feeling of collective responsibility for their role in developing a technology with "potential for great harm or great good" (atomic energy) and of their "special awareness" that it might lead to the "advance of our civilization or its utter destruction". In promoting the capability of robotics and AI towards a largely unknown end, our generation of researchers also have a special responsibility to understand and be outspoken about what the future of robotics might bring and its potential benefits and threats.

Acknowledgements: Preparation of this commentary was part-supported by the FP7 co-ordination action the *Convergent Science Network for Biomimetics and Neurotechnology* (http://www.csnetwork.eu), som of the ideas explored here were developed during the UK Arts and Humanities Research Council (AHRC) project *Cyberselves in Immersive Technologies* (AH/M002950/1).

Disclosure: The author is a director of the company *Consequential Robotics* that develops companion and assistive robots.

References

Association of Manhattan Scientists. (1945). Preliminary Statement. Retrieved from https://www.gilderlehrman.org/history-by-era/postwar-politics-and-origins-cold-war/resources/physicists-predict-nuclear-arms-race-

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Brown, A. (2015, 12th March 2015). To mourn a robotic dog is to be truly human *Guardian*. Retrieved from http://www.theguardian.com/commentisfree/2015/mar/12/mourn-robotic-dog-human-sony

Bryson, J. J. (2009, April, 2009). *Crude, Cheesy, Second-Rate Consciousness.* Paper presented at the The Second AISB Symposium Computing and Philosophy, Edinburgh.

Bryson, J. J. (2010). Robots should be slaves. In Y. Wilks (Ed.), *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issue* (pp. 63-74).

Bryson, J. J. (2012). A role for consciousness in action selection. *International Journal of Machine Consciousness, 04*(02), 471-482. doi:10.1142/s1793843012400276

Collins, E. C., Millings, A., & Prescott, T. J. (2013). *Attachment to Assistive Technology: A New Conceptualisation.* Paper presented at the Assistive Technology: From Research to Practice: AAATE 2013.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge: The MIT Press.

Dennett, D. C. (1994). The practical requirements for making a conscious robot. *Philosophical Transactions of the Royal Society of London A, 349*, 133-146.

Fellous, J.-M. (2004). *From human emotions to robot emotions*. Paper presented at the AAAI Spring Symposium on Architectures for modeling emotions: Cross-disciplinary foundations Menlo Park, CA.

Future of Life Institute. (2015). An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence. Retrieved from http://futureoflife.org/ai-open-letter/

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology, 57*(2), 243-259. doi:10.2307/1416950

Heylighen, F. (2002). The Global Brain as a New Utopia. In R. Maresch & F. Rötzer (Eds.), *Zukunftsfiguren*. Frankfurt: Suhrkamp.

Huebner, B. (2010). Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and the Cognitive Sciences, 9*(1), 133-155. doi:10.1007/s11097-009-9126-6

Kahn, J., Peter H., Ishiguro, H., Friedman, B., Kanda, T., Freier, N. G., Severson, R. L., & Miller, J. (2007). What is a Human?: Toward psychological benchmarks in the field of human–robot interaction. *Interaction Studies, 8*(3), 363-390. doi:doi:10.1075/is.8.3.04kah

Kang, M. (2011). *Sublime Dreams of Living Machines: The Automaton in the European Imagination*. Cambridge, MA: Harvard University Press.

Levy, D. (2007). *Love and Sex with Robots*. London: Harper Collins.

Lovgren, S. (2007, March 16th, 2007). Robot Code of Ethics to Prevent Android Abuse, Protect Humans. *National Geographic News*. Retrieved from http://news.nationalgeographic.com/news/2007/03/070316-robot-ethics.html

Metzinger, T. (2009). *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. New York: Basic Books.

Qiu, J. (2007). Probing islands of consciousness in the damaged brain. *The Lancet Neurology, 6*(11), 946-947. doi:10.1016/S1474-4422(07)70255-7

Robbins, P., & Jack, A. I. (2006). The Phenomenal Stance. *Philosophical Studies, 127*(1), 59-85. doi:10.1007/s11098-005-1730-x

Rodogno, R. (2016). Social robots, fiction, and sentimentality. *Ethics and Information Technology, 18*(4), 257-268. doi:10.1007/s10676-015-9371-z

Seth, A. K. (2016). Why fish pain cannot and should not be ruled out *Animal Sentience, 2016.020*.

Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines, 16*(2), 141-161. doi:10.1007/s11023-006-9030-6

Szollosy, M. (2016). Freud, Frankenstein and our fear of robots: projection in our cultural perception of technology. *AI & SOCIETY*, 1-7. doi:10.1007/s00146-016-0654-7

Tononi, G. (2008). Consciousness as Integrated Information: a Provisional Manifesto. *The Biological Bulletin, 215*(3), 216-242.