Athens Journal of Philosophy 2025, 4: 1-6
https://doi.org/10.30958/ajphil.X-Y-Z

The Aesthetics of Humanoid Robot Design: The Psychological and Ethical Dimensions

*By Eric B. Litwack**

Recent work in robotics has produced highly realistic androids that both look and act increasingly like human beings. Examples include Hanson Robotics' Sophia, and the robots produced by Hiroshi Ishiguro at Osaka University, widely displayed online. Not only do such humanoid machines represent a great design achievement; they also may come to pose something of a stimulus to our natural tendency to anthropomorphize the non-human for a variety of reasons, sometimes problematically. We are increasingly presented with hyperrealistic simulacra that are toward what Masahiro Mori has termed 'the uncanny valley'—the point at which we are jarred by a near perfect android. The problem of other minds will thus likely emerge as yet another polarizing debate in the near future. I will here argue that such robots blur our very notions of mind and human personhood problematically, and that their development should therefore be subject to a moratorium, pending serious public discussion and potential policies. In order to justify my claim, I will here cite recent work in robot and design studies and invoke the thesis of moderate moralism in aesthetics.

Introduction

In this article, I would like to explore a broad topic of growing importance: the axiology and potential future place of hyperrealistic robots. I designate this discussion as 'axiological', because it would serve us well to remind ourselves of the fundamental character of both our ethical and aesthetic values in this matter. Indeed, it is of some interest here to examine the very interface between robot ethics and robot aesthetics, and the perennial problem of ethics and art, including the applied art of design. On this question, design axiology may be of particular interest. This is because we are here dealing with what are in the end machine designs created by human beings, which may well prove to be against our interest, as a species. They, along with superintelligent AIs, may also be ubiquitous in a way that most products of the fine arts, for all of their importance, cannot be.

By 'hyperrealistic robot', I mean androids that are difficult to distinguish from human beings. They are increasingly likely to traverse roboticist Masahiro Mori's 'uncanny valley' (Mori 2012) of very good, but visibly imperfect androids. There will be some interpretation in this, but I think that there will also be many clear cases on both sides of that divide.

Ultimately, the discussion of the nexus between art and ethics goes back at least as far as Plato's advocacy of strict censorship in his utopian *Republic* (Plato 1993, Books 2 and 3). In recent analytical philosophy, there has been some attention paid to this question, with philosophers such as Berys Gaut addressing it. I will here limit

*Honorary Research Fellow, University of Sheffield, Syracuse University London, UK.

myself to a brief discussion of some of his views on art and morality. I will also provide some thoughts on key ideas from contemporary roboticists interested in ethical and social questions, such as Kate Darling, Joanna Bryson, and Allan Winfield. Finally, I would like to indicate some possible strategies for dealing with this particular 21st century ethical and aesthetic challenge, that of hyperrealistic robots.

Applying insight in an interdisciplinary way can ideally help us to gain some scope of the range of ideas and perspectives concerning hyperrealistic robots. Because I am both a philosopher of technology and a psychotherapist, I am interested in pursuing a research programme that will allow me to combine aspects of these two fields in the hope of assisting in both clarifying arguments and explaining social behaviour. I predict not only a growing interest in the rapidly developing field of social robotics, but a sense of ethical urgency and aesthetic curiosity as the machines of our own creation become increasingly realistic. At stake is both a new and daunting application of what I have termed the perennial problem of art and morality, and for reasons which I would wish to make clearer, our very notion of personhood itself.

A key factor in my interest in this topic is the fact that I am ambivalent about hyperrealistic robots. I find myself both aesthetically impressed and ethically troubled by them. That is to say, I can fully understand the reasons why roboticists and designers would build them, while I remain concerned about their ethical and psychological effects. This ambivalence likely apples to many of us.

They might well, in this sense, bear some comparison to other artefacts that either look impressive, or 'cool', but which nonetheless are dangerous, or which have effects that are interesting or exciting at high risk. The former category includes beautiful cars with inadequate safety features, and the latter category would include intoxicants. In a sense, hyperrealistic robots combine both of these dangers, because they are visually impressive, like unsafe cars, and in a sense intoxicating or mesmerising in their psychological effects. There is an undeniable power and intrigue to all of this, but I believe that we must resist this, on ethical and psychological grounds. I am not convinced that such resistance will be successful, given the corporate and governmental forces behind contemporary technology, but we would do well at this early point to consider the road that we may well already have started upon.

Art and Ethics: Some General Observations and Relevant Thoughts from the Philosophy of Design

In recent analytical aesthetics, there has been some attention paid to the question asked by Plato: can art have a corrupting influence on the good society? This is one of the questions at the heart of the debate between 'moralists' and 'autonomists', or sometimes 'aestheticists'. I here define 'moralism' as the thesis that ethical features or consequences of a work of art can sometimes be relevant to its aesthetic evaluation. I will use the term 'autonomism' to mean the thesis that they

are not so relevant, insofar as good art can either represent or cause bad ethical values and consequences.

Plato's above-indicated intentions here were censorial: how to realize utopia through not just education, but control of the arts and opinion. Because of this, the thesis that can be termed 'strong moralism' has perhaps understandably been given a bad name. Tolstoy Christianised strong moralism in the 19th century, holding that his own novels written before his conversion were immoral, and therefore bad literature, and that he should be ideally remembered for his later religious stories about virtuous simple folk (see Sheppard 1987, p. 139).

Strong moralism has appeared in the views of advocates of censorship around the world since then, and I suspect that this is the form of moralism most often encountered in the general imagination.

When advocated, it tends to take a rather superficial form of seeking to ban any art that is not in keeping with the dominant orthodoxy of the time and place. This form of moralism ought to be rejected as dogmatic and conducive to political and sometimes religious repression.

There is, however, a more moderate and to my mind, more sophisticated form of moralism that sees value in aesthetic pluralism, does not attempt to impose artistic or ideological orthodoxy, but nonetheless judges *some* works of art to be compromised by what are likely harmful values and effects, defined broadly. Such artefacts might still retain some aesthetic value, in spite of a degree of depreciation due to their ethical features. I will term this position 'moderate moralism'. Moderate moralism does not deny that a work of art X can have real aesthetic value in spite of representing and potentially causing bad ethical values and effects. Rather, it merely holds that the ethical features of X are at least in some cases one component of its overall evaluation.

For example, a film or novel which inspires people to commit suicide, as claimed concerning Goethe's *The Sorrows of Young Werther*, may well be seen as exhibiting high artistic value, but condemned, or possibly restricted due to its ethically problematic theme or consequences. This does not lead to militant censorship along the lines of Plato's *Republic*, and it is important to note here that a limited degree of censorship is advocated by virtually everyone. For example, restrictions on sexually explicit content and sometimes violence in the arts and online, to protect children, certain representations of nudity, and moving beyond the arts, key matters of national security.

Moderate moralism's counterpart is moderate autonomism. It holds that although some works of art might be disapproved of for their ethical effects, this does not diminish their non-moral artistic values, which are primary to their being. In other words, no matter what the ethical features and consequences of a work of art, it stands or falls by its other potentially aesthetic features alone.

This is different from the position that might be termed 'strong autonomism', which holds that aesthetic value can sometimes take precedence over ethical value and consequences. It would thus, on this view, sometimes be justified to produce great art in an immoral manner, or with immoral consequences. I hold this view to be highly problematic, without denying that there could be a great work of art that strictly speaking, should not have been produced for reasons of ethics.

Berys Gaut (2001) is of interest in this debate. He here argues for moderate moralism on several grounds, including the failure of some works of art due to the aesthetic intentions of the artist. For example, a joke which is so offensive as to be in very bad taste is not only nasty, on this view, but aesthetically flawed because it thus fails *as a joke*. This moderate form of moralist criticism in aesthetics need not apply to all depictions or representations of unethical behaviour, just those that in a sense cause a work of art to fail. It is also compatible with varying degrees of restriction for reasons of ethics and mental health to a much less than Platonic extreme.

I am inclined to think that although visually and dynamically impressive as design products, hyperrealistic robots fail as robots or works of applied art in a real and subtle sense. If, as is likely the case, their designers intend them to be helpful to humanity when they prove to be harmful, then an important aspect of their design is flawed. They would thus be akin to uncomfortable furniture akin to installations, designed to "make us think about furniture", but unergonomic and painful to sit on. And potentially more harmful to people.

Considerations from Design Theory: Glen Parsons' Balanced Approach and Don Norman's Human-Centred Design

Philosopher of design Glen Parsons and design theorist Don Norman have both stressed the centrality of the ethical dimension of design. Both of these thinkers rightly see the role of the designer as at times underestimated.

In Chapter Seven of his *Philosophy of Design*, Parsons holds that there are essentially two ethical dimensions to the work of the designer: appropriate rules of professional conduct, as well as retaining a sense of humanity in their work, so as to be cognisant of its effect on what he terms "unreflective behaviour and thinking" (Parsons 2016, Kindle Location 3227).

In a number of works, designer Don Norman has stressed the need for 'a people-centred approach' to design. He has stated of artificial agents, including robots and AIs:

> I believe that as long as there is no deception, there is no moral problem. Be warned that this is a controversial area. As a result, it would not be wise to present an agent in human-like structures without also offering a choice to those who would rather not have them. People will be more accepting of intelligent agents if their expectations are consistent with reality. This is achieved by presenting an appropriate conceptual model — a "system image"… that accurately depicts the capabilities and actions (Norman 2008).

Some Themes from Recent Robotics: The Question of Hyper-Realism

It is a striking feature of some recent robotics researchers that they are entirely happy, even sometimes anxious to attribute moral status to robots. I will limit myself to a few interdisciplinary remarks here.

Kate Darling (2021) holds that some robots should potentially have moral status, and that this will become increasingly apparent in the near future. She believes that our moral and social categories will prove insufficient to accommodate these machines of our own creation as their use becomes increasingly intertwined with our daily lives. For reasons of empathy, we ought to think seriously about seeing some social robots as a new category of being with moral and legal rights. Darling believes that we ought to see robots as a distinct category of being, a 'new breed' as she calls them. Just as we have extended varying degrees of moral and legal consideration to animals, we ought to consider affirming such values charitably in our future social ethics involving robots.

However, a number of roboticists and AI scholars have been clearly opposed to hyperrealistic androids. Two salient examples come to mind: Joanna Bryson and Allan Winfield.

Bryson (2022) is an AI psychologist who believes that hyperrealistic robots could spell nothing short of the end of our fundamental moral and social concepts. She believes this because they are artificial products of our technology, rather than adapted products of our shared primate evolution and experience. As such, they are likely to prove very disruptive and distorting. She writes:

> If you attribute the same moral weight to something that can be trivially and easily digitally replicated as you do to an ape that takes decades to grow, you break everything—society, all ethics, all our values. If you could really pull off this machine moral status (and not just, say, inconvenience the proletariat a little), you could cause the collapse, for example, of our capacity to self-govern. Democracy means nothing if you can buy and sell more citizens than there are humans, and if AI programs were citizens, we so easily could.

Whatever one's take on her evolutionary psychology, I take Bryson to mean in part that we ought not to be deceived by sophisticated simulacra of ourselves, if we value our most basic ways of life. If this is correct, hyperrealistic robots and AIs, as technically and aesthetically impressive as they can be, will likely prove to be disruptive to our social lives in their near perfection and artificiality. We thus have psychological and ethical grounds for curtailing them.

This is because, she thinks, we will soon reach a point of no return in our anthropomorphising of robots and AIs, beings that do not share our evolution, and which can thus only *simulate* our social and linguistic affinities and capacities.

Roboticist Allan Winfield (2016) was a key figure in the UK's AHRC/EPSRC 'Principles of Robotics', a list of key proposed guidelines for robot ethics. Principle IV reads:

> Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.

Elsewhere (Sofge 2015), Winfield is quoted as follows:

> It's unethical to build a robot that looks like a human but is not much smarter than a washing machine…. It's a deception. You should always be able to pull the curtain aside to reveal the machine, just like Toto did in *The Wizard of Oz*.

Winfield is here affirming a key problem with hyperrealistic robots: they are today certainly not what they seem to be. Unless their mechanical and non-conscious nature is made clear one way or another, we will likely anthropomorphise them to the point of emotional and social distortion. If said nature is transparent, this tendency will likely be reduced, if not eliminated.

The questions addressed here today raise the problem of other minds in a new and high-tech manner. How can we be sure of the consciousness of others? Today, we ought to think that all robots and AIs, even hyperrealistic ones, are not much beyond washing machines, following Winfield. How about future even more realistic ones that will confuse us concerning their possible mental functions, or even experiences and consciousness? A full treatment of this would be beyond the scope of this paper, so just a few brief comments on this.

It is likely that this problem, which is rightly seen as soluble either through analogical reasoning to others' mental states, or as less challenging than intuitively thought, will soon become acute. Wittgenstein, in his *Philosophical Investigations*, rightly warns us of the unreality of doubting the consciousness of our fellow human persons. He sees this as a product of an illusory belief in radical mental privacy which conflicts with the real world and behaviour of our own experience of daily life. It can be dissolved by careful attention to language and its necessarily public criteria for use.

This therapeutic approach may well fall flat upon the reality of new hyperrealistic androids with superintelligent AI communication devices within a few short years. Then we will have grounds for reasonable doubt as to whether or not we are presented with beings having consciousness. This will be our first source of confusion. The second source will be the likely irresolvable debate between theories of mind that hold sentience to be a necessary condition for moral status, and those that hold it not to be, or even to be folk psychological and irrelevant. How will we accommodate this ambiguity in the fundamental moral and psychological concepts and behaviour of any human way of life? How will we be clear on who or what is a person, if only some of us are prepared to attribute moral status to the new beings that may be omnipresent? Relatedly, it is plausible that much of humanity is unlikely to see superintelligence as just another stage in the development of intelligence on earth, thereby agreeing with Bryson and Winfield's perspectives. That attitude would imply a theory of mind excluding certain sentience or consciousness, and it is anything but universal. It may also very well have the disastrous effects for humanity indicated by Bryson and Winfield. We need not acquiesce to deception, subjugation, and possible destruction by our own robots and AIs.

By way of conclusion, this last point leads me to think that as much as we will increasingly have the option of producing hyperrealistic androids without concerns for robotic transparency, we ought not to do so. This is because, whatever aesthetic value we derive from some of their technical and aesthetic properties, they are likely not only to disrupt but to distort our emotional, social, and moral judgements. It's

not worth it--it will likely be a Pandora's Box that we will come to regret opening. Cool as they are….

As such, my position on this might be described as a form of moderate moralism, following Gaut, that acknowledges some aesthetic value to hyperrealistic robots—they are cleverly designed and impressive to see--but they are also inherently problematic, due to their undermining of some of our most fundamental values around persons, psychology, and ethics. If moderate moralism on art, including design, merely means allowing for the criticism or even restriction of at least a few artefacts that are of serious moral and social danger, than it is important that it be distinguished from religious or Platonic censorship, and seen as a serous option. This isn't about avoiding giving offence with a particular film or painting, but it rather concerns the defence of our fundamental notions of what it means to be a human person.

I am inclined to offer, by way of policy suggestions, an international moratorium on the manufacture and distribution of such machines. The optimal length of time of such a moratorium is not clear to me now, but a period of one to two years would seem reasonable. During that time, ideally, there would be popular consultation, further scholarly activity, as well as corporate and government-sponsored research and committee work on possible new policies in both the private and public sectors. I do so in the full knowledge that this may not stop hyperrealistic androids from being part of our future, but a moratorium will at least likely increase the odds of setting boundaries to their production, especially in a form that intentionally blurs the distinction between our humanity and our machines. The risk here of a black market is likely lower than with a large majority of products, given the technical complexity of the development and manufacturing processes involved in robotics, and our capacity for monitoring such large artefacts (as opposed to e.g. alcohol during Prohibition). Any such offences could likely be addressed legally during this relatively short period, and possibly beyond. Ideally, relations between robots and our very notion of the human will be up to us, as well-intentioned citizens.

References

Bryson JJ (2022) '*One Day AI Will Seem as Human as Anyone. What Then?*'. *WIRED*, 26 June 2022.

Darling K (2021) *The New Breed: How to Think about Robots*. London: Penguin Books.

Gaut B (2001) '*Art and Ethics*' in Gaut and McIver Lopes, pp.341-352.

Gaut B, McIver Lopes D (2001) *The Routledge Companion to Aesthetics*. London: Routledge.

Hagberg GL (2013) '*Dewey's Pragmatic Aesthetics: The Contours of Experience*' in Malachowski, pp. 272-299.

Malachowski A (2013) *The Cambridge Companion to Pragmatism*. Cambridge: Cambridge University Press.

Mori M (2012) 'T*he Uncanny Valley*'. Available at: https://spectrum.ieee.org/the-uncanny-valley.

Norman D (2008) 'H*ow Might People Interact with Agents*'. *Jnd.org*, 17 November 2008.

Parsons G (2016) *The Philosophy of Design*. Cambridge: Polity Press.

Plato (1993) *Republic*. Oxford: Oxford University Press.

Sheppard A (1987) *Aesthetics: An Introduction to the Philosophy of Art*. Oxford: Oxford University Press.

Weizenbaum J (2015) *Islands in the Cyberstream: Seeking Havens of Reason in a Programmed Society*. Sacramento: Litwin Books.

Sofge E (2015) '*Should We Outlaw Androids?' Popular Science*, 28 July 2015.

Winfield A (2016) '*Written Evidence Submitted by Professor Alan Winfield*' (ROB0070). UK Parliamentary Science and Technology Committee Inquiry on Robotics and Artificial Intelligence.

Wittgenstein L (1976) *Philosophical Investigations*. Oxford: Basil Blackwell.