

Applications for self-aware systems. Why would we need conscious shoes?

Robert Pepperell
Metatechnology Research
School of Art Media and Design
University of Wales Newport, UK

Abstract

Many recent developments in interactive design are aimed towards the 'humanisation' of technology, that is, making technology behave in a way that is more 'intuitive', 'friendly' or 'usable'. This assumes that technology is not in itself human but rather some external antagonistic force or object we are in perpetual conflict with. Contrary to this view I will offer the suggestion that technology is but an extension of human activity and therefore part of what constitutes humanity as a whole. One consequence is that machines and devices are no longer regarded as alien agents to be tamed and controlled, but are embodiments of human ingenuity and intelligence. However, although technological devices might be understood as embodying cognition, and hence to some extent consciousness, it could not be claimed that they are self-conscious in the sense we normally attribute to other humans. Looking briefly at some proposed mechanical models of self-consciousness, I consider the question: What functions might self-aware systems perform? How, for example, might they be incorporated into products, and how would such incorporations affect the way we interface with machines? I argue such questions lie at the heart of serious research into applications for artificial consciousness, and despite their inherent complexity need vital consideration.

Technology is commonly characterized as a class of objects or devices that stand apart from and in distinction to humankind. Moreover, as the popular fantasies of futurologists, science-fiction writers and movie makers demonstrate, machines are often portrayed as presenting an imminent threat to human interests. And despite warnings about the dangers of technological determinism, even sophisticated writers on the subject sometimes assume that the devices surrounding us are somehow 'other' than human, almost a self-sufficient living realm with their own laws of evolution and logic of existence.

Given both the ongoing anxieties about advanced technology and its necessity in preserving our standard of living, it is not surprising to find attempts to smooth the often perturbing interface between ourselves and the systems we use, to make them more attractive, accessible or engaging. While it has fallen to artists and poets to elevate raw technology to the aesthetic plane, or to endow it with anthropomorphic values such as 'vitality' or 'heroism' (think of J W M Turner's *Rain, Steam and Speed* of 1844 or F T Marinetti's *Futurist Manifesto* of 1909), the somewhat more practical task of 'interpreting' or 'humanising' technology in order to make it friendlier or more usable has traditionally fallen to designers. Indeed, the widely quoted definition of Industrial Design offered by the International Council of Societies of Industrial Design is precisely

the ‘humanisation’ of technology, design being “. . . the central factor of innovative humanisation of technologies and the crucial factor of cultural and economic exchange.” (ICSID 2001).

But the widespread, underlying assumption of an essential distinction between humans and machines has come under increasing strain in recent decades. From Marshall McLuhan’s *Understanding Media: The Extensions of Man* (1964) to Bruce Mazlish’s *The Fourth Discontinuity* (1993), technology has come to be seen in some quarters less as an autonomous force than as an extension of human attributes, expanding our capacity to see, move, affect at a distance, and so on.

“During the mechanical ages we had extended our bodies in space. Today, after more than a century of electric technology, we have extended our central nervous system itself in a global embrace, abolishing both space and time as far as our planet is concerned. Rapidly, we approach the final phase of the extensions of man - the technological simulation of consciousness, when the creative process of knowing will be collectively and corporately extended to the whole of human society, much as we have already extended our senses and our nerves by the various media.” (McLuhan 1964)

In *The Posthuman Condition* I argued that the perceived separation between humans and machines was, in effect, illusory and that it makes more sense to consider technology not just as a modern-day extension of human attributes but as an integral component of what it is to be human in the first place. Humans cannot be understood in isolation from the technological environment that sustains them. What makes us human is our wider technological domain, just as much as our genetic code or our relation to the natural environment (Pepperell 1995 and 2003a).

Once the perceived distinction between humans and technology is erased — if only theoretically — then the necessity for ‘humanisation’ or anthropomorphisation recedes. As our technological appendages are no less human than our biological ones, it seems erroneous to humanise something that is already, at least by extension, part of the human condition.

In this ‘posthuman’ schema, as I have termed it, technological agency is no different in kind from human agency (like using pliers in place of fingers, for instance) since in either case human will is enacted through the manipulation of the environment. Once the erasure of the human/technology distinction is fully grasped, then the fact that my hand moves to push a button (the button itself a product of human will, ingenuity and labour) has no inherently distinct status from the remote operation of my domestic water heater — each acts as an agent to modify the world according to my will.

The condition of agency is common to both my direct bodily actions and the wider technological domain that services my will. The deeper ontological (indeed, metaphysical) question of who, or where, is the ‘me’ that is being served will not be explored here, pertinent though it is. Instead, for the purposes of this essay, it will be assumed that ‘me’ and my agency are co-extensive, that is, where my agency is enacted I am in some way present. In the same way, then, that physical human attributes such as strength and vision are extended by our technological environment, so mental attributes such as will and desire are extended by the devices that act as agents. To give an example, the motor car massively increases our physical capacity to move about terrain

and transport heavy items, but at the same time embodies a whole set of ideological beliefs, customs, desires and aspirations that are manifest in the material construction and operation of the vehicle.

One of the potential consequences of this line of argument is the claim that insofar as objects embody attributes of human cognition (such as desire, aspiration, beauty, etc.) they themselves acquire cognitive attributes. Or to put it in another way, those technologies that manifest human intelligence also embody it. This means there is a sense in which all technologies can be regarded as conscious or intelligent (eliding for a moment the distinction) insofar as they embody the conscious action of the people who create them and extend the conscious agency of those who employ them. It seems an extreme point of view, but one that is gaining some currency in recent discussions of human-technology interaction. In *How Images Think*, Ron Burnett (2004) makes the case that emerging modes of collaboration between humans and machines, such as remote co-working, peer-to-peer communication, and networked musical composition, mean human intelligence, which was once confined to individual sentient beings, has become a distributed phenomena. With particular reference to the generation and reception of images, he argues the very technology that sustains these new kinds of distributed intelligence itself gains a kind of intelligence status:

“The intersections of human creativity, work, and connectivity are spreading intelligence through the use of mediated devices and images, as well as sounds. Layers upon layers of thought have been ‘plugged’ into these webs of interaction. The outcome of these activities is that humans are now communicating in ways that redefine the meaning of subjectivity. It is not so much the case that images per se are thinking as it is the case that intelligence is no longer solely the domain of sentient beings.” (Burnett 2004)

Although Burnett’s actual claim is somewhat weaker than the title of his book would suggest, he nevertheless supports the idea that technologies in general, and images in particular, can be regarded as having cognitive attributes, such that, “images turn into intelligent arbiters of the relationships humans have with their mechanical creations and with each other.”

Yet, while we might follow Burnett in granting technologies a certain intelligent status, or whether one accepts, as McLuhan claimed, that machines are becoming manifest extensions of human consciousness, it would be far more contentious to make the further claim that such images or devices are ‘self-conscious’ in the sense we normally attribute to each other. In other words, for images to truly think, or for machines to really be conscious, they would have to enjoy some subjective sensibility — some knowledge of their own existence in the world and their relation to other such self-conscious entities.

The philosophical and technical obstacles to implementing self-awareness in mechanical substrates are immense. Yet perhaps the problem of endowing a mechanical system with self-awareness is not insurmountable; certainly numerous research projects are underway and many kinds of approach being adopted. Prominent amongst them is that of neural engineer Igor Aleksander, who expresses some confidence that the project of creating ‘digital sentience’ will ultimately bear fruit. In *How to Build a Mind: Toward Machines with Imagination* (2001) he proposes an “ego centered” model of neural

simulation through which conscious properties, such as imagination, might emerge. Like Aleksander, the philosopher Susan Stuart (2002) argues that the generation of self awareness in artificial agents requires they be embedded in the dynamic structure of the world. My own approach is to stress not only the necessity of embeddedness, but that in any artificially conscious system the mechanical components should interact such that certain parts of the system are able to sense other parts, and moreover, to *sense themselves sensing* (Pepperell 2003b). It is this regressive self-referentially, I contend, that gives rise to the peculiarly self-aware sense of being that humans enjoy, and which when implemented in a mechanical substrate, will arguably generate something similar for machines.

However, even if such a self-aware system were to be successfully built, a significant problem would remain: What would it be used for? What applications require systems that sense their own presence in the world? Is there any need for a self-aware lift, or a pair of conscious shoes, and in what ways would their being conscious affect their functionality? These questions, I would argue, merit attention not because we are on the brink of being overrun by some malign mechanical master race, but because the kinds of answers we give will shape the very purpose and direction of research into artificially sentient agents. Moreover, they impinge directly on issues concerning the usability, accessibility, and engagability of advanced reflexive technologies.

For Igor Aleksander, research into conscious machines serves two purposes: it helps us to understand how natural processes work, and it serves as a model for the development of “useful” products based on the principles derived from the study of such natural processes:

“In addition to being an explanatory device, it is worth asking how such machine consciousness might also empower useful machines. The relationship between the artificial and real versions of consciousness remains just like that between the robot with vision and the night owl. The robot may be quite different, but understanding the properties shared by the two is sufficient to design robot systems inspired by the excellence of owl vision, and to understand owl vision better by knowing how robot vision can be designed.” (Aleksander 2001)

Such naturally inspired systems, he claims, are better fitted to function well in the complex dynamics of the real world by dint of their capacity to build better representations of it:

“...I assume that the world is real and that the more accurately a brain (real or artificial) brings this reality into the consciousness of the individual or the machine, the more successfully will that individual or machine cope with the real world.” (Aleksander 2001)

Artificially conscious systems will then, by implication, be more self-reliant and robust than systems lacking a capacity for the kind of high-level awareness Aleksander imagines machines will some day possess.

At the Nokia Research Center in Finland, principal scientist in cognitive technology Pentti Haikonen conducts investigations into conscious machines (Haikonen 2003a). His work focuses on giving machines a level of understanding comparable to that of humans,

the point being machines that understand in a similar way to us will be able to operate as we do. This means conscious machines will, for example, have a flow of inner speech, an active imagination, visual and narrative recognition, and so on. “Obviously there would be a large number of important applications for machine understanding”, he states, although the commercial sensitivity of his work inevitably restricts wider dissemination of what these might be (Haikonen 2004). However, an entry on the Nokia web site offers some hints:

“A new cognitive technology will arise with unforeseen applications. Will we see artificial personal assistants that are more than digital diaries, ones that are more like trusted friends? Will we see robots that are able to negotiate their way in dangerous locations and save lives? Will we see deep space probes that carry consciousness to infinity and beyond? And finally, will we see gadgets that really help us to use them?” (Haikonen 2003b)

The notion of the mechanical devices as willing collaborator, in the sense of having some degree of free will in making choices to help its human operator, may seem fantastical, but is obviously one that drives much serious research into product development. Stephen Thaler, CEO of Imagine Engines Inc., has developed what he calls the ‘Creativity Machine’, a self-referential neural net architecture that can generate novel “ideas” as well as make associations between patterns, as neural nets normally do. By getting one cluster of nets to monitor the output of another, Thaler claims he has produced a “canonical model of consciousness in which the former net manifests what can only be called a *stream of consciousness* while the second net develops an *attitude* about the cognitive turnover within the first net (i.e., the subjective feel of consciousness).” In his patent application for the Creativity Machine, he writes:

“The present device can be used to tailor machine responses thereby making computers less rigid in communicating with and interpreting the way a human responds to various stimuli. In a more generalized sense, the subject device or machine supplies the equivalence of free-will and a continuous stream of consciousness through which the device may formulate novel concepts or plans of action or other useful information.” (Thaler 1997)

Citing applications from machine vision and robotics to stock market forecasting and virtual entertainment products, the Creativity Machine is presented as a semi-autonomous creative agent endowed with some sentience. It is obviously debatable the degree to which this supposed sentience contributes to the performance of this particular neural net design, and indeed debatable to what degree it is really sentient at all. However, the Creative Machine, Haikonen’s work at Nokia and Aleksander’s neural engineering research each hint at the extent to which the quest for machine intelligence might inform and inspire product design.

Conclusion

There is growing intellectual support for the idea that humans are co-extensive with technology, both in the physical and cognitive sense. This state of co-extension requires that we revise our attitude towards human-machine interaction: if technology is now regarded as an extension of human cognition then the classical model of interaction, whereby two distinct entities are interfaced, one sentient and one insentient, is inaccurate. In its place we must posit an exchange of cognitive activity between the sentient user and the cognition embodied in the device. However, while this might be true in the case of all current and prior technologies, for which no claim of 'self-awareness' or 'self-consciousness' can be substantiated, it would not be so for any system which can be shown to have sentience of its own. In such a scenario, the mode of exchange then shifts to that between two entities, both of which can lay claim to self-consciousness. It seems that in certain research communities the desirability for such conscious architectures is motivated by the belief that 'conscious' products will outperform their insentient counterparts; they will be more effective, friendlier, faster, more rewarding, and so on. Whether this claim is justifiable might itself be the subject for further discussion and research. Moreover, the prospect of sentient machines throws up numerous ethical, social, philosophical, and technical problems that are only in the early stages of being formulated, let alone resolved. But the intriguing question of what conscious technologies might be for remains, with all its implications for the study of usability and engagability in product design.

Bibliography

Aleksander, I. (2001). *How to Build a Mind: Toward Machines with Imagination*. Columbia: Columbia University Press.

Burnett, R. (2004) *How Images Think*. Cambridge: MIT Press.

Haikonen, P. O., (2003a). *The Cognitive Approach to Conscious Machines*, Exeter: Imprint Academic.

Haikonen, P. O., (2004). Personal communication.

Haikonen, P. O., (2003b). *Nokia: Towards thinking Machines*.
<http://www.icsid.org/static.php?sivu=3> (accessed 26th May 2004).

ICSID (International Council of Societies of Industrial Design)(2001). *ICSID Design Definition*. Available at: <http://www.icsid.org> (accessed 26th May 2004).

Mazlish, B. (1993). *The Fourth Discontinuity: The Co-evolution of Humans and Machines*. Yale: Yale University Press.

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. New York: McGraw-Hill.

Pepperell, R. (1995 and 2003a). *The Posthuman Condition: Consciousness Beyond the Brain*. Bristol: Intellect.

Pepperell, R. (2003b). Towards a Conscious Art, *Technoetic Arts*. (Volume 1,2).

Stuart, S. (2002). A Radical Notion of Embeddedness: A Logically Necessary Precondition for Agency and Self-Awareness. *Metaphilosophy* Vol. 33, No. 1, 98-109.

Thaler, S. L. (1997). *Device for the autonomous generation of useful information*. US Patent 5,659,666. See also: <http://www.imagination-engines.com> (accessed 26th May 2004)