

Applications for conscious systems

Robert Pepperell

Received: 6 February 2006 / Accepted: 28 October 2006 / Published online: 24 November 2006
© Springer-Verlag London Limited 2006

Abstract Many recent developments in technological 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, however, that technology is not in itself human but rather some external antagonistic force or object. Contrary to this, I will defend the suggestion that technology is part of what constitutes humanity as a whole, to the extent of embodying some degree of cognition and consciousness. Looking briefly at some proposed mechanical models of self-consciousness, I consider the question: What functions might self-conscious systems perform?

1 Introduction

HAL, the artificially consciousness computer in Stanley Kubrick’s *2001: A Space Odyssey* (Kubrick 1968), is a chilling hybrid of the human and the mechanical. He is a clever, friendly and protective companion for the spaceship’s crew, gradually becoming suspicious of their motives and finally driven by ruthless self-interest to acts of violation. At the same time, he is so much immobile hardware—circuits, lenses, cables, etc.—of the kind we might rou-

R. Pepperell (✉)
School of Art and Design, University of Wales Institute Cardiff,
Cardiff, UK
e-mail: pepperell@ntlworld.com

R. Pepperell
School of Art & Performance, University of Plymouth, Plymouth, UK

tinely use, misuse, reconfigure or trash. Despite HAL's deadly behaviour one feels something close to pity as the logic boards are ripped from his heart towards the end of the film, as though someone rather than something were being destroyed.

The case of HAL exposes deep anxieties about our relationship to technology, not because HAL is inhuman but because he is so very human. We recognise in him aspects of our own companionable, suspicious, ruthless natures and wonder whether in his shoes we also might put the interests of the mission before those of the crew, as he does when he turns on the two main protagonists, Dave and Frank. Yet we are reminded HAL is a only machine, albeit a machine that implicitly extends the motives and skills of those officials and engineers back on Earth who sponsored, designed and commissioned him. Lest the computer's decision to eradicate Dave and Frank seem 'inhuman', 'cold' and 'logical' we can assume it was one already preordained by his human makers, especially in light of HAL's claim that, "No 9000 series computer has ever made a mistake or distorted information". HAL's ultimate purpose is to flawlessly extend the will of his sponsors, and presumably the designers at H.A.L. felt this purpose would be best met by a machine with a capacity for conscious thought rather than one without. With his conscious faculties HAL becomes not an autonomous mechanical entity but a super-efficient proxy for distant human minds.

This paper considers the relationship between humans and machines, and in particular machines that might have some sense of consciousness of the kind we recognise in ourselves. No such machines exist now. Yet it is possible—even likely I would claim—that they will, and the question raised here is "what will their purpose be?" Why would we need conscious machines, and how might sentient technologies be incorporated into products and artefacts?

2 Humans and technology

Anxiety about technology, of the kind represented by HAL, often rests on an implicit assumption that the human and the technological realms are essentially distinct. For example, the definition of design offered by the International Council of Societies of Industrial Design is precisely the "... central factor of innovative humanisation of technologies..." (ICSID 2005), implying that technology in its raw state is somehow 'inhuman'.

However, 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, extending our capacity to see, move, affect action 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)

Elsewhere I have argued that the apparent separation between humans and machines is, 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 always has been to be human. Consequently, human beings cannot be understood in isolation from the technological environment that sustains them. What defines humanity is our wider technological domain, just as much as our genetic code or our relation to the natural environment (Pepperell 1995/2003a).

Once the apparent distinction between humans and technology is erased then the necessity for ‘humanisation’ recedes. Since 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 essence 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.

3 Technological embodiment of cognition

I want to suggest that insofar as technological objects manifest—by extension—attributes of human cognition (such as intelligence, thought, ingenuity, etc.) they themselves acquire such attributes. Or to put it in another way, those technologies that are the products of human intelligence thus 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 manifest the conscious attributes 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, once seen as 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 intelligent 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.”

4 Self-conscious technologies

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-consciousness 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.

5 The function of self-conscious technology

Even if, and by whatever means, such a conscious 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.

At the moment they are questions that seem not to have been addressed specifically but rather generally, often on the working assumption that sentient technologies will somehow be better or more useful than insentient ones. In the case of Igor Aleksander, for example, 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. But exactly what functions they might then be able to perform are not specified.

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 that machines that understand in a similar way to us will be able to operate as we do. 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 (P. Haikonen, 2004, personal communication). 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)

Again, these are rather generic aspirations, with little indication of what specifically conscious products or devices might result.

Stephen Thaler, CEO of Imagine Engines Inc., has developed what he calls the ‘Creativity Machine’, a self-referential neural net architecture that he claims 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 declares 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).” (Thaler 1997). 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 generalised 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. But it is 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. The latest blurb on the product web site

mentions ‘consciousness’ and its role in the functionality of the system in only the slightest of terms (Imagination Engines 2005).

The Creative Machine, Haikonen’s work at Nokia, and Aleksander’s neural engineering research each tantalisingly hint at the extent to which the quest for machine consciousness might inform and inspire product design. Yet none supplies concrete examples of what a specifically *self*-conscious machine (as opposed to just a very *clever* machine) might be for. What would my self-conscious shoes be able to do that my merely intelligent shoes couldn’t?

6 Closing remarks

According to the fiction of *2001*, the sense of self-awareness implanted into HAL was crucial to his ability to complete his mission, from the pastoral empathy he establishes with his human charges to the ruthless self-interest and sense of mistrust that drive his murderous actions—in short, it’s what made him so good at his job. And it is very likely that future machines will be implanted with sentience for precisely the same reason—that it makes them better at their jobs. Even if such devices are in fact best considered humans by proxy, i.e., remote implementations of the cognitive attributes of their makers, they are no less human for that. As HAL politely tells an interviewer from the BBC:

“I am putting myself to the fullest possible use, which is all I think that any conscious entity can ever hope to do.”

Despite the terror of human redundancy that HAL represents to some, there is growing intellectual support for the idea that humans are not antagonistic but 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 relations: if technology is now regarded as an extension of human cognition then the classical model, whereby two distinct entities interact, one sentient and one insentient, is inaccurate. In its place we must posit a distribution of cognitive activity between the sentient user and the device.

But 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 and a sense of selfhood. It seems that in certain research communities the desirability of 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, more innovative, and so on.

I would argue that that whether this claim is justifiable should be the subject of further discussion and research, and the intention of this paper has been to direct attention towards this. Clearly 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 specific question of what conscious technologies might be for, with all its implications for the study of soft and wearable technology in product design, remains to be fully addressed.

Acknowledgments Some ideas in this paper were first presented at the 1st International Design and Engagability Conference, University of Central England, July 2004. I am grateful to Stephen Thompson for drawing my attention to the ICSID definition of design.

References

- Aleksander I (2001) How to build a mind: toward machines with imagination. Columbia University Press, Columbia
- Burnett R (2004) How images think. MIT Press, Cambridge
- Haikonen P (2003a) The cognitive approach to conscious machines. Imprint Academic, Exeter
- Haikonen P (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) (2005) http://www.icsid.org/about/Definition_of_Design/ (accessed 29th December 2005)
- Imagination Engines (2005) <http://www.imagination-engines.com/cm.htm> (accessed 29th December 2005)
- Kubrick S (1968) 2001: a space odyssey (Dialogue transcribed from film)
- Mazlish B (1993) The fourth discontinuity: the co-evolution of humans and machines. Yale University Press, Yale
- McLuhan M (1964) Understanding media: the extensions of man. McGraw-Hill, New York
- Pepperell R (1995/2003a) The posthuman condition: consciousness beyond the brain. Intellect, Bristol
- Pepperell R (2003b) Towards a conscious art. *Technoetic Arts* 1:117–134
- Stuart S (2002) A radical notion of embeddedness: a logically necessary precondition for agency and self-awareness. *Metaphilosophy* 33:98–109
- Thaler S (1997) <http://www.imagination-engines.com>, including Device for the autonomous generation of useful information, US Patent 5,659,666. (accessed 26th May 2004)