Available online at www.sciencedirect.com
ScienceDirect
Interacting with Computers xxx (2008) xxx–xxx
Interacting with Computers
www.elsevier.com/locate/intcom

Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents

Blay Whitby *

Department of Informatics, University of Sussex, East Sussex, Falmer, Brighton, BN19QH, UK

Abstract

This is a call for informed debate on the ethical issues raised by the forthcoming widespread use of robots, particularly in domestic settings. Research shows that humans can sometimes become very abusive towards computers and robots particularly when they are seen as human-like and this raises important ethical issues.

The designers of robotic systems need to take an ethical stance on at least three specific questions. Firstly is it acceptable to treat artefacts – particularly human-like artefacts – in ways that we would consider it morally unacceptable to treat humans? Second, if so, just how much sexual or violent 'abuse' of an artificial agent should we allow before we censure the behaviour of the abuser? Thirdly is it ethical for designers to attempt to 'design out' abusive behaviour by users?

Conclusions on these and related issues should be used to modify professional codes as a matter of urgency.

Keywords: Robot ethics; Abusive interaction; Ethical design

1. Introduction

Just because a prominent computing entrepreneur predicts 'a robot in every home' in the near future (Gates, 2007) or the government of a technologically advanced Asian country makes a similar commitment by 2010 (Lovgren, 2006) does not entail that the long awaited era of the universal domestic robot is imminent. We are usually justified in treating such predictions with extreme caution. However, the technological state of the art strongly suggests that robots and some equivalent technologies will enter domestic and intimate roles in the near future.

The central claim of this paper is that there is a manifest, clear need for wide, dispassionate, and informed discussion of the ethical issues raised by the widespread use of robots and similar technology.[1] It would be desirable to commence this discussion in the immediate future. This should not be taken as any sort of argument against such technology or its future development. Domestic robots, various sorts of household automation, and the use of Artificial Intelligence (AI) in intimate and caring roles have tremendous benefits to offer humanity. The main claim made here is a 'call to arms' for all involved to ensure that these technologies are introduced in as ethical a manner as possible.

The existing codes of conduct of the British Computer Society (BCS) and Association for Computing Machinery (ACM) do not cover the issues considered here. This should be rectified as soon as possible. The central claim of the paper should not be taken naively to entail the view that ethics committees or professional organizations, for example, will always provide the right answer, nor sometimes even any answer at all. This is simply not that sort of area. Similarly it would be difficult, if not impossible, at this point to delineate realistic ethical principles that were uncontroversial. However, this paper is a sincere attempt to initiate useful debate on these issues.

* Tel.: +44 1273 677853.
E-mail address: blayw@sussex.ac.uk

[1] 'Informed' in this context means *both* technically and ethically informed.

0953-5438/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.intcom.2008.02.002

The paper will subsequently talk only of robots for the sake of brevity. This should be taken as including various forms of AI technology including at least smart apartments, computer-based advisors, and decision support systems. Where the precise types of technology have differing ethical consequences this will be indicated.[2]

Some specific ethical issues are considered in detail in subsequent sections. This is by no means an attempt to produce an exhaustive list of the ethical issues raised by robots and artificial agents. Indeed it would be unrealistic to attempt to produce an exhaustive list: many problems will emerge in comparatively surprising ways. However, the fact that we cannot predict all the ethical issues that widespread robot use will bring is no justification for ignoring those that can be reasonably foreseen. Nor is the paper an attempt to cover a representative cross section of ethical issues. The three ethical issues discussed in detail here focus specifically on the abuse of robots and artificial agents. This is an interesting and important area for consideration by both technologists and ethicists.

There are many developments – for example technological solutions for the care of the increasing elderly population of many developed countries – that should draw our attention to this area. However, the lot of a 'robot companion' is not necessarily a happy one. There is already empirical evidence that various forms of mistreatment of robots by humans will be common. This behaviour can be aggravated by the provision of anthropomorphic interfaces (De Angeli et al., 2006) or by placing the robot into an intimate setting (Fogg and Tseng, 1999). Such evidence is all the more worrying because this is precisely the direction in which robotic and computing technology is currently developing.

It is important to be clear that the word 'mistreatment' here does not imply that the robot or intelligent computer system has any capacity to suffer in the ways that we normally assume humans and some animals have. The concern of this paper is only with the sense of mistreatment implied in a statement such as: "Bill mistreats every car he owns".

There has been considerable philosophical interest in the ethics of mistreatment of possible future artefacts which *could be* said to suffer in some morally relevant sense (for example Allen et al., 2000; LaChat, 1986; Torrance, 2000). However philosophically interesting such robots might be, there is little technical prospect of such artefacts being built for the foreseeable future. This paper is concerned only with the ethics of mistreatment of robots that have precisely the same lack of capacity to suffer as does a car. In everyday language we would not talk of the suffering of artefacts such as cars except in a highly metaphorical context.

An immediate reaction to this clarification might be that the mistreatment of cars is not an ethical matter. Cars are usually wholly owned. If, therefore, an owner wishes to treat his or her car particularly maliciously then they will pay for the privilege and it need not concern anyone else. We do not normally form moral judgements about the way in which people treat their own property – at least when it has no further moral consequences. In the language of moral philosophy, the statement: "You ought not to rev the engine so hard or spin the wheels." contains a 'prudential ought', not a 'moral ought'. It means nothing more than that the car will last longer if one refrains from such activities.

Some may see ethical consequences in the mistreatment of cars. For example, poor driving practice or inadequate maintenance may endanger other road users and this would rightly be condemned on moral grounds. If we exclude this possibility by considering only the case where Bill takes his car to a private race track and then proceeds to abuse and mistreat it, there may still be grounds for condemnation. The car contains a significant amount of natural resources and human labour and Bill may therefore be seen as wasteful to a degree that we might wish to condemn on moral grounds.

Now robots, considered from an ethical viewpoint, have many of the properties of cars. Since the discussion is limited to robots that have no capacity in themselves to suffer in a morally significant sense, the words 'mistreat' or 'abuse' applied to robots pick out a superficially similar set of ethical consequences to those picked out by their application to cars. We may also, for present purposes, assume that foreseeable robots will be wholly owned (by individuals or groups) who therefore acquire property rights over them. They also represent significant amounts of natural resources and human labour. If the robot is a large and expensive piece of machinery then mistreating it will entail the same ethical issues raised by Bill's mistreatment of his car in the previous paragraph. If other humans are endangered by either Bill's behaviour or the behaviour of a damaged robot running out of control then this is morally equivalent to poor driving practice and therefore wrong. Also, in spite of Bill's property rights over his robot, we might want to condemn the unnecessary abuse of a large and expensive piece of machinery as wasteful, contributing to environmental damage, or as reprehensible conspicuous consumption.

However, the next section will consider some ethically important differences between robots and cars.

2. Robot companions are not just cars

If robots and cars had all their ethical properties in common then the discussion might end at this point. However, there are also some important ethical differences between cars and robots which should affect our conclusions on the central claim of this paper.

The first, and most important, ethically significant difference concerns the extent to which the robot is seen as human-like. Robots, particularly those used as artificial companions, will be significantly more human-like than cars. Let us clarify what this might mean. A robot may

[2] Following Singer (1979) the terms moral and ethical are used interchangeably.

Please cite this article in press as: Whitby, B., Sometimes it's hard to be a robot: A call for action ..., Interact. Comput. (2008), doi:10.1016/j.intcom.2008.02.002

resemble a human along at least three distinct dimensions. The first is that of physical appearance; the second that of behaviour; and the third that of the role it is designed to fulfil. All three dimensions have ethical consequences. Designers of robots cannot escape the ethical implications discussed in this paper simply by making the robot clearly non-human along one of these dimensions.

For example, deliberately avoiding any anthropomorphism in the appearance of a robot will not enable the designer to escape the ethical issues under discussion, if it is obvious that the robot is very human-like in its behaviour. There is good reason to expect that humans will respond to it as if it were a human, albeit a very different looking one. How this is happening at present in the, perhaps rather unlikely, case of soldiers bonding with military robots in combat situations has already been observed (Garreau, 2007). Even in the case of an automated apartment which does not resemble a human along either the appearance or behaviour dimension we may still validly talk of its being human-like. This is because it is designed to fulfil an intimate caring role with respect to an individual.

How human-like we consider a robot to be is clearly a matter of degree. There may be a popular assumption that robots have to more or less *completely* resemble human beings – usually along the appearance and/or behaviour dimensions to raise significant ethical issues. This is not the case. Partial resemblance is sufficient to cause problems of ethical significance. The so-called 'Turing test' (Turing, 1950) and its modern derivatives may be misleading here. An implied goal of the Turing test is to deceive a human into thinking that they are communicating with a human when, in fact, they are communicating with an artefact.

Now in cases where a human moral agent *believes* for good reasons that they are dealing with another human it seems clear that their behaviour should be governed by the established ethics of human behaviour. If they are subsequently made aware of the deceit in that they were not in fact interacting with another human, that is interesting but not morally significant. It does not imply that the user's behaviour morally should have been different in any way. However, we may well want to scrutinize from an ethical standpoint the designer's motives for deceiving the user, especially if there is evidence of intentional deceit for reasons not in the best interests of the user. This point has been explored at length elsewhere, particularly in Weizenbaum (1984) and Whitby (1988, 1996) so will not be developed here. It would, however, clearly be an important aspect of the sort of debate for which this paper calls.

The specific ethical issues considered in this paper focus on those cases where the human knows, that they are interacting with a robot.[3] The subsequent arguments do not depend on the human being deceived into thinking that they are dealing with anything other than an artefact.

Of course, the broad definition of robot used in this paper and the sliding scale of human-likeness mean that what humans actually know may not be clear-cut in practice. Indeed, the use of robot companions for people with known cognitive deficit (such as the elderly with brain degenerative disorders) may lead to some interesting and difficult cases. However, the existence of such interesting and difficult cases supports the central claim that there is a need for informed debate on the ethical issues involved with the introduction of such technologies.

A second ethically significant difference between cars and robots stems from the fact that cars are an established technology. Since most advanced societies already have widespread car ownership, the moral consequences of their design and use are, for the most part, in clear view. Robots are an extremely novel (though much anticipated) technology. Examination of the moral consequences of their widespread use requires a greater amount of prediction of likely development and consequences.

This second difference should not be superficially read as saying that cars have clear moral consequences, whereas those of robots are speculative. All practical moral judgements involve prediction. It is true that the degree of prediction required is greater in the case of robots but it is reasonable to expect their widespread use in the foreseeable future. With the benefit of hindsight, it is now clear that a more appropriate time to discuss the moral consequences of widespread car ownership would have been in anticipation of it actually happening (perhaps around the time that Henry Ford proposed mass production of cars in 1913). From this observation it follows that the time to discuss the ethics of the widespread availability of robot companions is now.

3. Three immediate ethical issues

As an initiation of the sort of ethical debate called for, the next sections consider three specific ethical questions that seem to be of practical relevance to robot designers. The first is to what extent do we consider it acceptable to deliberately mistreat artefacts – particularly substantially human-like artefacts? To the extent that society declares such behaviour unacceptable, we need to begin discussion on how it is to be deterred, prevented or avoided. Although it might seem appropriate to rely on legislation to do this, robot designers need to be aware of the issues and to play a role in promoting legislation that does not unnecessarily restrict the technology. They should also take an interest in how the ethical issues inform the more practical aspects of design.

The second question depends largely on the responses to the first question. To the degree that we find deliberate mistreatment of robots morally unacceptable, what ethical limits can we justly place on such behaviour? Are these

[3] The argument of this paragraph assumes the familiar ethical principle: ought implies can – normally attributed to Kant (1788) but also found in Hobbes (1651). Like many principles in ethics, it is sometimes contested. However, an examination and defence of this principle would be outside the scope of this paper.

Please cite this article in press as: Whitby, B., Sometimes it's hard to be a robot: A call for action ..., Interact. Comput. (2008), doi:10.1016/j.intcom.2008.02.002

essentially a new set of ethics, or familiar ethics applied to this field?

The third and related ethical question affects the technical design of robots and HCIs (human–computer interfaces). It concerns the ethical consequences of trying to engineer out some of the problems of the first two questions. It might well be possible to design robots which behave in ways that tend not, under usual circumstances, to prompt abusive behaviour in humans. The ethical issues raised by adopting this approach to design are discussed in full in a subsequent section.

These three questions are undoubtedly ethical questions in that they ask what we ought to do in a moral sense and they are not readily resolvable by any empirical methods.

These are ethical problems to which the designers of robots and intelligent systems should pay close attention. However, the fact that these are ethical problems associated with the design of robots does not exempt other groups from moral responsibility. At the very least, governments, their appointed agents, legal opinion formers, and moral opinion formers also have responsibilities in this area.

4. The ethics of mistreating robots

Let us consider the first question: to what extent is it ethically acceptable to mistreat robots or intelligent artefacts? For present purposes we are excluding any case in which the artefact itself is capable of any genuine or morally significant suffering.

We are therefore concerned only with human (and possibly animal) moral consequences. In the previous section two ethically significant differences between the mistreatment of cars and the mistreatment of robots were observed. The first of these was that robots (and certain other types of intelligent technology) were substantially more human-like than cars. Because of this, it might be argued that a person is doing something morally reprehensible to a robot if they mistreat it in ways that we would clearly condemn on moral grounds if they were similarly to mistreat a human.

This argument needs a certain amount of unpacking. The argument that the mistreatment of anything human-like is morally wrong subsumes a number of other claims. The most obvious of these is that those people who abuse human-like artefacts are thereby more likely to abuse humans.

This claim that 'they might do it for real' has received a great deal of attention with respect to other technologies in recent years. The technology most relevant to the present discussion is probably that of computer games. Modern computer games[4] provide a high (and increasing) level of realism, often employing AI technology, to provide the user with a sense of involvement in various simulated activities. Many of these activities are extremely violent. If it were the case that those who participate in such simulated violence were thereby rendered more likely to do it for real, this would be clear moral authority for preventing the use, sale, and production of such games.

Debates about the ethics of computer games are informative for the issues under discussion. Since the definition of robot used in this paper is fairly broad, it would be reasonable to assert that when gamers shoot and kill computer-based avatars containing some degree of AI, then they are abusing robots in the sense under discussion here.

From the example of computer games we can draw some conclusions relevant to the mistreatment of robots. The first is that the empirical claim that such activities make participants more likely to 'do it for real' will be highly contested. The competing claim is usually that the Aristotelian notion of 'catharsis' applies (Aristotle, 1968). This entails that by doing things to robots, or at least in virtual ways, the desire to do them in reality is thereby reduced. The catharsis claim would be that mistreatment of robots reduced the need for people to mistreat humans and was therefore morally good.

I have argued previously that the available empirical evidence is not yet clear enough to bear directly on this debate (Whitby, 1993). However, the overall trend in empirical evidence since that publication strongly suggests that regular involvement with simulated violence does in fact desensitize users to violent activities in real life. Experiment design needs to be ingenious since mere correlation tells us nothing in this area. Both experiment design and interpretation of the results remain controversial but the balance of evidence is worrying (see, for example, Anderson and Bushman, 2001).

In spite of this trend in evidence, controls on the design of computer games are weak and the realism of violence continues to increase. This suggests a second despairing conclusion that most societies will be unable or unwilling to control the widespread abuse of robots. Obviously, this despairing conclusion does not bear on the ethical discussion called for in this paper. Put simply, the fact that we have allowed ethically undesirable things to happen in the past is no good argument in favour of continuing to do it in the future.

5. How far can we justify limits on individuals' behaviour?

A second claim may be subsumed by the argument that mistreating human-like entities is wrong. This is the claim that certain activities are simply not acceptable, even if done in private to an artefact with no capacity to suffer. How much weight we give to this second claim depends on our beliefs about the relative value of individual freedom as opposed to society's right to pass judgements. The classic argument in favour of individual liberty was made by J.S. Mill:

[4] 'Computer games' should be read as including console-based, VR (virtual reality) games, and arcade games.

Please cite this article in press as: Whitby, B., Sometimes it's hard to be a robot: A call for action ..., Interact. Comput. (2008), doi:10.1016/j.intcom.2008.02.002

> The only part of the conduct of anyone, for which he is amenable to society, is that which concerns others. In the part which merely concerns himself, his independence is, of right, absolute. Over himself, over his own body and mind, the individual is sovereign (Mill, 1859).

This view is an important constitutive principle of most Western secular societies. The tolerance of individuals' private activities implied by this view is likely to be influential in debates on the ethics of mistreating robots. It is worth remarking in passing that many, if not most, religious groups are likely to take a view diametrically opposed to that implied by Mill, particularly in the case of robot companions used for sexual purposes.

This paper does not seek to challenge directly this particular principle of individual liberty. On the other hand, a couple of useful clarifications can be made. The first, and most important, is that the principle of individual liberty has hardly ever been extended to children, who are usually seen as needing some level of restriction. It might therefore seem consistent with the restrictions on personal liberty placed on children to adopt a similar attitude with respect to robot abuse. The second is that even the most libertarian societies do occasionally deem certain activities – for example certain types of sexual activity – to be unacceptable even when they are committed in private. This is the position which can be anticipated from the major Christian denominations, for example.

It would therefore be consistent for even the most tolerant societies to place some restrictions on robot abuse. A highly probable, though perhaps not inevitable, candidate for this sort of restriction would be the use of robots for paedophilic activities. When this matter is debated in the popular media it is unlikely that the separation of the claims made in this section will be so clearly made. The argument from sheer unacceptability will inevitably be conflated with the 'they might do it for real' argument.

A further possible conflation is that robot designers will generally be unwilling to be associated with uses of robots, or other intelligent technology, of which there is strong social disapproval. They may portray their unwillingness as ethical, when it is merely commercial.

The important ethical issue here is that of openly declaring some types of activities to be simply unacceptable. The list of unacceptable activities may be fairly short in tolerant Western societies but it would help both robot builders and users to discuss any items potentially on this list now. If the list is empty, that should be clearly stated. If there are intrinsically unacceptable activities then they should be made explicit. Candidate examples might be the use of robots in paraphilic sexual activities and purely as the victims of violence. Although the principle of individual liberty would suggest allowing this, there are clear reasons to believe that it could have morally unacceptable consequences. It may be that there are classes of users – for example children, or those with known psychiatric disorders that morally we should protect from the possibilities offered by the technology. The fact that, as discussed in previous section, we have already allowed a high degree of violent abuse in the context of computer games is not an argument for continuing to do so. Nor does it provide a good moral argument in favour of the list being empty. However, the claims of the principle of individual liberty will count very strongly, one suspects, for the other side in this debate.

As has already been remarked, there is no reason to anticipate widespread agreement on these issues. The central claim of this paper is simply that we need to begin this discussion now. An ideal result would be clear guidance on what, if anything, is on the list of unacceptable activities enshrined in the relevant professional codes and maybe eventually in law.

6. The ethics of 'designing out' problem human behaviour

Given the difficult ethical issues discussed in the previous sections and that a significant amount of the responsibility lies primarily with designers it might therefore seem that they are ethically bound to produce systems which minimise the possibility of abuse. There is current interest in software than can manipulate the affective state of users (du Boulay et al., 1999). Indeed at least one writer (Picard, 1998) has claimed that an ability to respond to human emotion is the next stage in Artificial Intelligence. These strands of research could become united in the deliberate attempt to design systems which take account of the emotional state of the user and behave in ways designed to minimise the possibility of any abuse.

Unfortunately, this is not a complete solution to the ethical problems described above. On the contrary, it simply introduces a further set of ethical problems.

One further ethical problem lies in the possibility that robot companions could become so much more well-suited to their owners' affective tendencies that humans would wish to spend more time with them and less in human society. After all why would one want to engage in the uncertain, risky, and difficult interactions of human society when it is possible to purchase an artificial companion that indulges one's every foible without complaint or even complains only when you want it to?

This may sound like a science fiction possibility but it is already visible in small ways in current technology. The ability of an on-line book shop, for example, to recommend purchases based on stored information about every user vastly exceeds what one could reasonably expect of a human bookseller. A human (ideally) may have considerably more intelligence and genuine care for the customer but has to compete with the on-line system's ability to store and rapidly access vast databases of purchases, ownership and similarities. A human bookseller could do this only for a handful of her best-known customers.

If a bookshop employs part-time workers at relatively low rates of pay, they are unlikely to be able to perform this sort of service at all. They may even become curt when

Please cite this article in press as: Whitby, B., Sometimes it's hard to be a robot: A call for action ..., Interact. Comput. (2008), doi:10.1016/j.intcom.2008.02.002

asked: "Have you got anything similar?" Customers will start to find the level of service from the non-human system superior to that which can be achieved by human systems under normal circumstances.

The technology of robot companions will undoubtedly offer a far higher level of knowledge of and adaptability to their individual users than do on-line retailers. The ethically interesting consequence is the way in which this will make them excessively attractive to their users. There is a distinct possibility that users may find interacting with their artificial companions more attractive than interacting with human companions. This leads to the worry that people will not willingly venture out to meet other humans, since their robot companions offer them a safer and more fulfilling interaction.

The possibility of this leading to a dystopia where human society breaks down irretrievably has been detailed by the social psychologist Frude (1983). We cannot be sure just how likely this is to happen from the perspective of the present but Frude saw the process as inevitable. Boden (2006) gives a comprehensive list of current technological developments which could help bring about convincing 'artificial companions'. This combination of technical potential and human desire is worrying, if nothing more.

In terms of the ethics of design, we do not need to accept the full dystopian vision of the future. Smaller effects in the reduction of social skills by those who spend too much of their time with robot companions are also worrying. Individuals who are more socially awkward will have less incentive to improve their social skills if robot designers are prepared to meet their needs for companionship by artificial means. If this matter is left solely to the market then there is the distinct possibility that some users will be disadvantaged by their use (or over-use) of the technology.

There are a number of responses that might be made by robot designers to these ethical worries. The first would be to claim that they are only giving users what they want. If people want to spend their time with artificial rather than real companions then society has no right to attempt to prevent them.

It is clear that society has every right to prevent this sort of development if it threatens the continued existence of social practices which are considered important. It is a primary function of government and law to ensure that people live in reasonable harmony. Since this technology has the potential to be antisocial, in the most literal sense of the word, societies have the clear duty to protect themselves against its worst effects. A second, related response might be to observe that we have, as a matter of history, largely accepted the similar dissocializing effects of previous technologies such as television. Of course the fact that this has happened in past (in itself debatable) does not provide any argument in favour of our allowing it to happen in the future.

A third response might be to propose to use robot companions only in a limited and ethically responsible manner. In the case of the care of the elderly, it might be argued that the robot companions should only be made available to those people whose family and friends are not available to help in their care. This response is certainly ethically preferable to that of 'just giving people what they want'. However, it leads to some further difficult ethical dilemmas. For example, should we indiscriminately allow old people to purchase (where they can afford it) robot carers? Or should we first determine whether it might be better to persuade their younger family members to care for them? What criteria might we use to decide? Should we allow socially awkward young people to freely purchase robot companions? Or is this too hastily condemning them to social isolation which we would be better advised to try to remedy – perhaps by training in social skills?

These dilemmas are more of a problem for society as a whole than for robot designers and governments and legislators may have to play key roles. However, everybody engaged in the design of robots needs to give at least some attention to these issues. This is firstly because they affect the general reaction to and enthusiasm for the technology and secondly because they bear on the design of the technology. It is usually a design goal to make the interaction with the user as attractive as possible. This may well not be always ethically correct.

A fourth, primarily technological, response by robot designers might be to use human–human interactions as a template. It might be argued that if the robot companion behaves more or less as a human would do under the same circumstances then these ethical dilemmas are much reduced. In terms of the attractiveness of the interaction, it need be no more and no less attractive than interacting with a human. In as much as this turns out to be technically feasible, it should be ethically desirable.

Unfortunately this fourth response leads us straight back into the ethical problem of the first section. Robots that behave more or less as humans do are precisely the sort of robots that are most likely to be abused. If we design systems that appear to express anger or loss of temper when treated unreasonably or 'insulted' by users then that will tend to prompt abusive, maybe even violent, behaviour in the user. The need for the sort of ethical debate called for in this paper should be clear. Not only will it sometimes be hard to be a robot, it will sometimes be hard to be a robot interaction designer too.

7. Can we apply existing ethical principles to this area?

Since the technologies discussed in this paper are under development, rather than in widespread use, there is at least a possibility that they might be seen as requiring new approaches to ethics. The growing field of roboethics is certainly prompting moral philosophers to re-examine some fundamental assumptions. It is a key claim of this paper that we already have adequate established ethical and metaethical principles with which to deal with the problems under consideration. It is possible, therefore, to

Please cite this article in press as: Whitby, B., Sometimes it's hard to be a robot: A call for action ..., Interact. Comput. (2008), doi:10.1016/j.intcom.2008.02.002

resolve the issues under discussion with our existing ethical concepts.

One of the most powerful arguments in favour of the view that conventional ethics apply to this area might be called 'technological transparency'. This stresses the unimportance of technological developments to fundamental moral arguments. There is no *moral* difference between the wrongness of injuring someone with a horse-drawn vehicle and injuring them in a similar fashion with a modern car. The wrongness of this act stems from its human consequences, not from the technological means by which it is performed. The technology in this case may be relevant to moral discussion in that it may provide a greater number of opportunities for immoral acts to a greater number of people, but the wrongness of the act is unaffected by the nature of the technology used to perform it.

A counter-argument to this is the claim that the problems emerge with new technology demand new ethical responses. This is often framed in terms of the technology actually altering the society which adopts it. In short, this counter-argument would entail that the ethics of a society with widespread usage of robot companions is different from the present society and we should not force our values upon it. The premise of this argument which claims that societies with widespread usage of robot companions will be different from present societies seems highly plausible. The conclusion that we cannot therefore adopt an ethical stance towards it is less secure, however.

There are a number of ways of dealing with this type of counter-argument. Firstly, it must be pointed out that it contains a degree of social relativism that is excessive, if not bizarre. If we cannot ever legitimately apply existing moral principles to the (perhaps much changed) society of the future then, taken to its logical conclusion, such total neutrality entails that we must be ethically neutral towards *any* potentially society-changing events.

A more convincing dismissal of the counter-argument is to show that the transition from the present state to the future society consists of a number of small steps. Every decision in this process contains an ethical element because it is causative in bringing about the change. Decisions about what sorts of robotic research to fund; decisions about what sort of user to design for; decisions about what is an appropriate interface; and decisions about who should have access to the technology all play a part in bringing about the future we are considering. This paper calls for ethical discussion and, wherever appropriate, ethical input in all of these stages.

8. Conclusions

The evidence that has already emerged on human behaviour suggests that there is a need to take ethical issues into account in the design of robot companions. There would seem to be an urgent need for professional codes of conduct to be extended to the areas under discussion. It would also be better to have an awareness of the ethical implications in the general public before market forces drive this technology into everybody's lives.

The best time to consider the ethical implications of widespread car ownership would have been before we had built an infrastructure around car usage that makes it very difficult for people to give up, or even reduce, their usage of this technology. The same almost certainly applies to robot companions. We need to discuss the ethical implications of their widespread ownership before we reach a stage where people feel that cannot live without them.

The overall picture is worrying, particularly when we consider what has happened with some previous technologies. The penultimate section argues that it may often be unethical to build robots that are inappropriately pleasant to their users. Unfortunately, this will probably be precisely what the market demands. Even if robot manufacturers are prevented by law from producing intelligent artefacts that tolerate morally incorrect behaviour by their users, one can safely predict hacking and illegal modifications to allow this. There is also an international dimension to this problem. Individual governments may be largely powerless to steer the development of the technology in the ways called for.

Returning to the analogy with cars gives more grounds for pessimism. Car manufacturers have tended to produce cars that will sell well, rather than perform ethically. The continued and deliberate lack of attention to crash safety by US manufacturers through the 50s and 60s would be a good example. This is the sort of ethical issue that cannot be readily solved by markets and may be extremely resistant to legislation. Modern businesses often express the desire to behave in an ethical manner. However, in the absence of the sort of debate called for in this paper they will be motivated primarily by market demand, rather than these ethical considerations. Those who are concerned about ethics may have to fight to be heard. It is worth stating again that these practical difficulties do not amount to an argument for neglecting the ethics of this area.

Despite all this concentration on negative impacts, it is important to close by stating once again that there are many positive benefits to this technology, even in the areas under discussion. There seems little doubt that, despite the problems considered in the paper, robot companions are a worthwhile and useful technology. We will be changed by the technology and perhaps not always for the better but it will significantly improve life for many people. Even if the abuse of robots becomes widespread this may give us important scientific insights into the roots of human abusive behaviour. After all, we are just humans.

Acknowledgements

I am much indebted to Professor Margaret Boden and Dr. Sharon Wood for their detailed comments on an early draft of this paper.

Please cite this article in press as: Whitby, B., Sometimes it's hard to be a robot: A call for action ..., Interact. Comput. (2008), doi:10.1016/j.intcom.2008.02.002

References

Allen, C., Varner, G., Zinser, J., 2000. Prolegomena to any future artificial moral agent. Journal of Experimental and Theoretical Artificial Intelligence 12, 251–261.

Anderson, C.A., Bushman, B.J., 2001. Effects of violent video games on aggressive behaviour, aggressive cognition, aggressive affect, psychological arousal, and prosocial behaviour: a meta-analytic review of the scientific literature. Psychological Science 12 (5), 353–359.

Aristotle, 1968. In: The Poetics. Aristotle, OUP.

Boden, M.A., 2006. Mind as Machine. A History of Cognitive Science. Oxford University Press, pp. 1094–1095.

De Angeli, A., Brahnam, S., Wallis, P., 2006. Misuse and abuse of interactive technologies, In: Proceedings of CHI'06, pp. 1647–1650.

du Boulay, B., Luckin, R., del Soldato, T., 1999. The plausibility problem: human teaching tactics in the 'hands' of a machine. In: Lajoie, Susanne P., Vivet, Martial (Eds.), Artificial Intelligence in Education: Open Learning Environments: New Computational Technologies to Support Learning, Exploration and Collaboration. Proceedings of the International Conference of the AI-ED Society on Artificial Intelligence and Education. IOS Press, Le Mans France, pp. 225–232.

Fogg, B.J., Tseng, H., 1999. The Elements of Computer Credibility, Proceedings of CHI'99, pp. 80–87.

Frude, N., 1983. The Intimate Machine: Close Encounters with New Computers. Century, London.

Garreau, J., 2007. Bots on the ground: in the field of battle (or even above it), Robots are a soldier's best friend, Washington Post, May 6th, 2007.

Gates, B., 2007. Scientific American Magazine, January, pp. 58–65.

Hobbes, T., 1651. Leviathan, XIV, pp. 25.

Kant, I., 1788. Critique of Practical Reason. 8, pp. 287.

LaChat, M.R., 1986. Artificial intelligence and ethics: an exercise in the moral imagination. AI Magazine 7 (2), 70–79.

Lovgren, S., 2006. A Robot in Every Home by 2010, South Korea Says, National Geographic News, 6th September, 2006.

Mill, J.S., 1859. On Liberty, reprinted in John Stuart Mill, A Selection of his Works. Robson, J.M. (Ed.), (1966) Macmillan, Toronto, pp. 14.

Picard, R., 1998. Affective Computing. MIT Press, Cambridge, MA.

Singer, P., 1979. Practical Ethics. Cambridge University Press, pp. 1.

Torrance, S., 2000. Towards an ethics for epersons. In: Barnden, J. (Ed.), Proceedings of the AISB 2000 Symposium on Artificial Intelligence, Ethics and (Quasi-) Human Rights. University of Birmingham, pp. 47–52.

Turing, A.M., 1950. Computing machinery and intelligence, Mind, vol. LIX, No. 236.

Weizenbaum, J., 1984. Computer Power and Human Reason. Penguin, Harmondsworth.

Whitby, B.R., 1988. Artificial Intelligence: A Handbook of Professionalism. Ellis Horwood, Chichester, pp. 152–158.

Whitby, B.R., 1993. The virtual sky is not the limit – the ethical implications of virtual reality. Intelligent Tutoring Media vol. 3, No. 2.

Whitby, B.R., 1996. Reflections on Artificial Intelligence: The Legal, Moral, and Ethical Dimensions. Intellect, Exeter, pp. 93–105.

Please cite this article in press as: Whitby, B., Sometimes it's hard to be a robot: A call for action ..., Interact. Comput. (2008), doi:10.1016/j.intcom.2008.02.002