AISB/IACAP World Congress 2012

Birmingham, UK, 2-6 July 2012

Framework for Responsible Research and Innovation in AI

B. C Stahl, M. Jiroka and G. Eden (Editors)

Part of

Published by
The Society for the Study of
Artificial Intelligence and
Simulation of Behaviour

http://www.aisb.org.uk

ISBN 978-1-908187-20-8

Foreword from the Congress Chairs

For the Turing year 2012, AISB (The Society for the Study of Artificial Intelligence and Simulation of Behaviour) and IACAP (The International Association for Computing and Philosophy) merged their annual symposia/conferences to form the AISB/IACAP World Congress. The congress took place 2–6 July 2012 at the University of Birmingham, UK.

The Congress was inspired by a desire to honour Alan Turing, and by the broad and deep significance of Turing's work to AI, the philosophical ramifications of computing, and philosophy and computing more generally. The Congress was one of the events forming the Alan Turing Year.

The Congress consisted mainly of a number of collocated Symposia on specific research areas, together with six invited Plenary Talks. All papers other than the Plenaries were given within Symposia. This format is perfect for encouraging new dialogue and collaboration both within and between research areas.

This volume forms the proceedings of one of the component symposia. We are most grateful to the organizers of the Symposium for their hard work in creating it, attracting papers, doing the necessary reviewing, defining an exciting programme for the symposium, and compiling this volume. We also thank them for their flexibility and patience concerning the complex matter of fitting all the symposia and other events into the Congress week.

John Barnden (Computer Science, University of Birmingham)
Programme Co-Chair and AISB Vice-Chair
Anthony Beavers (University of Evansville, Indiana, USA)
Programme Co-Chair and IACAP President
Manfred Kerber (Computer Science, University of Birmingham)
Local Arrangements Chair

Ethical Implications for Quality of Life in Robot Assisted Care of the Elderly

Denis Roche[1]**,**

Abstract. As researchers think about Machine Ethics and how ethical decision-making might be implemented in a machine, philosophers such as Torrance[1] and Coecklebergh [2]argue that in order to do so, we must reconsider the boundaries of, and broaden, our moral community. According to Torrance[3] the quest for an ethical 'producer', as a practical research programme involving the engineering of artificial moral agents quickly 'shades into a broader, more theoretical inquiry in to the nature of ethical agency, moral value...and the extent to which autonomous A.I agents can have moral status of different kinds'. That is, those worthy of ethical treatment in order to include them in our ethical or moral community.

Taking these ethical debates as a backdrop, I carried out a qualitative survey of the intuitions of nursing staff and care workers regarding their ethical concerns about the use of robots in two care-of-the-elderly facilities in the Republic of Ireland. Using methodology grounded in Experimental Philosophy[4], a semi-structured interview using a- priori themes derived from the literature was used to collect data, which was transcribed and analysed using Template Analysis [5]. Participants were asked to respond to a series of questions in the form of a structured interview which investigated themes such as participant's knowledge of robots and their feelings about the use of robots in care of the elderly. Participants were asked to consider any ethical issues relating to the use of robots, attitudes to robots being solely responsible for clinical care and their attitudes to a humanistic relationship developing between the older person and a robot.

Overwhelmingly the concept of patient autonomy was to the fore in all of their considerations and responses and was frequently used as the benchmark against which they weighed their responses.The responses of these naive participants, highlighted and matched a significant number of the deliberations and narratives of philosopher experts. A novel finding from this small-sample- size research was the discovery that if the field is to advance, the methods of Experimental Philosophy will need to be relied on more as a method for deriving the necessary information on the successful implementation of ethical comportment in the design of robots. It was clear from respondents that the contract of care, that they recognized as existing between them and their older charges, extended beyond a mere provision of service. Therefore, the danger in designing robots as service providers lies in the 'not fully grasping' of this concept.[1]

[1] Vivartes,National College of Art & Design, 40 The Weir ,Kilkenny, Co.Kilkenny, Ireland Email: {denis.roche@vivartes.ie}

1 INTRODUCTION

As the AI project advances, becoming more sophisticated and integral to contemporary life, the question of addressing the development and engineering of an 'artificial morality' comes increasingly to the fore. Indeed from the discipline of agent-based computing we see a growing need for a set of rules or conduct to govern the behaviour of software agents as they interact [6]. Genuine concerns [7] for the predicted increase and ambitions of artificial computational intelligence is prompting researchers to take seriously the project of developing an ethical or moral dimension to the behaviour of robots, particularly those designed for use in the military and medical arena.

Philosophers such as Torrance[8]and Coecklebergh [9] argue that to comprehensively address the modelling of artificial morality we must reconsider the boundaries of, and broaden, our moral community. According to Torrance [10] the quest for an ethical 'producer', as a practical research programme involving the engineering of artificial moral agents quickly 'shades into a broader, more theoretical inquiry in to the nature of ethical agency, moral value...and the extent to which autonomous A.I agents can have moral status of different kinds'. That is, those worthy of ethical treatment in order to include them in our ethical or moral community .One of the main reasons to expect that robots will be used in the care of the elderly is that the number of elderly people in the population is beginning to overtake the numbers of young people able to do such caring[11]. In 2009, it was estimated that 16.2% of the population in the UK was aged 65 or older [12]. Spain and Italy are the oldest in Europe with 18.1, and 20.2% over 65 respectively[13]. These figures are increasing sharply. In the UK, the fastest growing age group is made up of those aged 80 years and over who in 2009 constituted 4.5% of the population. In the US, 12.8% are over the age of 65, expected to rise to around 20% by 2030. It seems likely that Europe and the US may want to follow the Japanese lead into robot care.

The empirical research reported here covers a number of important practical and theoretical issues concerning the introduction of robots into a particular morally sensitive area - namely assisting in the care of elderly people. Underlying this are some deeper philosophical issues concerning the moral status of such robots - in particular, how far an artificial agent of this kind could be a moral 'producer' unless it was also a moral 'agent'."

This research set out to investigate the types of moral 'expectations' that care workers might have of a robot carer. It also investigated the areas of ethical concern that they themselves identified in providing care for an older person and the attitudes of nursing staff and carers to the use of robots in caring for the elderly. The research was conducted at two different nursing homes in the Republic of Ireland from the 30th of August to the 1st September 2010. The research involved ten participants; four staff nurses and two activity coordinators. These staff were chosen as they have the most contact time with elderly residents and have an active role in developing the daily care regime for residents. They are also the members of staff that would be most likely to be replaced by or work along side robot carers.

2 TAKING A PLACE IN THE MORAL COMMUNITY

Moral agency depends on at least two conditions and one common precondition[14]. First, one has to have the capacity to freely choose one's acts; the agent's behaviour is not compelled by something external to it. Furthermore, this requires that the person deliberates, or has at least the capacity to do so. Free choice also presupposes that the agent be rational. A second condition is that one has to know the difference between right and wrong. This requirement is often understood as knowing how to apply moral concepts and principles.

In his paper, Moral Appearances: Emotions, Robots, and Human Morality, Mark Coeckelbergh[15] begins by acknowledging that if morality depends on emotions then it seems unlikely that we will be able to build genuinely 'moral robots' any time soon .

Stating that robots currently do not "meet the standard necessary conditions for having emotions: they lack consciousness, mental states and feelings" [16]. He adds that the best that could be hoped for is that robots could be programmed to follow rules which could dangerously result in a 'psychopathic robot' that lacked full moral agency. Setting aside the bigger question of whether a robot can be genuinely moral in its behaviour, he argues that human "social and moral life depends on appearance" [17]. In normal social interaction, we do not demand proof of mental states in another person, instead we engage in an act of interpretation of the others behaviour and give it the status of emotion. We also assume that the other is doing the same. Referring to what he calls 'virtual intentionality', he states that we "tend to interact with them as if our appearance and behaviour appeared in their consciousness", giving them virtual subjectivity or quasi-subjectivity [18].

Coeckelbergh [19] states that if robots manage to imitate subjectivity and consciousness in a sufficiently convincing way - "they too could become the quasi-others that matter to us in virtue of their appearances". That is, that we, as emotional and social beings would come to care about how we appeared to robots - about what robots would 'feel' and 'think' about us. Robots would become virtual subjects or quasi-subjects with virtual emotions or quasi-emotions. As we will come to see later in the research findings, some of these ideas will be in some way borne out by the response of the participants. Coeckelbergh [20] proposes that we must shift philosophical attention in moral anthropology from what we really are to anthropomorphology, the human form, what we appear to be, and how other beings appear to us given(our projections and recreations of) the human form. It is his intention to make plausible the idea that "it is not their intentional state, but their performance that counts morally and that we can gain from moving a discussion about artificial intelligence to artificial performance".

3 ROBOTS IN THE WORLD OF CARING

Amanda and Noel Sharkey consider the ethical implications in relation to the upsurge of assistive robots being developed for use in care of the elderly. In their paper, Granny and the Robots [21], they outline six main ethical concerns; the potential reduction in the amount of human contact; an increase in the feelings of objectification and loss of control; a loss of privacy; a loss of personal liberty; deception and infantilisation(this has a correlate in the work of Sherry Turkle et al)[22] ; the circumstances in which elderly people should be allowed to control robots

As we will see later in the research report, these issues prove to be saliently observed and are the subject of practical feedback from participants. Sharkey and Sharkey point out, that in upholding the human rights of the elderly, it is essential to ensure that any robot introduced into a care setting is in fact improving the life of the older person and not just reducing the burden on the rest of society.

Taking each of the points listed above, they begin by considering the problem of the potential for loss of social contact through the introduction of robots to the care environment. Citing Sparrow & Sparrow [23], they worry that by replacing humans with robots in the performance of many menial jobs such as cleaning or feeding, that valuable opportunities for social contact will be lost. The Sharkeys' ethical concerns continue within the context of the problem of objectification of the elderly through the use of robots.

> Robots designed as replacement nurses or carers that carry out some of the same tasks of feeding, lifting etc., may make their charges feel like objects. Such robots could make elderly people feel that they had even less control over their lives than when they are dependent on human nursing care. [24]

Sharkey and Sharkey are more positive when it comes to ethics concerning robots and the dignity of older people. They see more potential for the use of robots as tools for the elderly, which could empower the elderly, increasing their autonomy thereby improving psychological and physical wellbeing [25]. Turning to the question of lack of privacy, the Sharkeys examine the ethics of monitoring robots such as the CareBot . The use of such robots could, in their opinion, reduce actual human contact and companionship and infringe on the right to privacy. Privacy can be expressed both as a right, but also as a generally recognized human value which has been discussed before in terms of computers, the internet and surveillance in general. Another ethical concern is the loss of personal liberty

which could occur if a robot was operating in a pre-emptive fashion all the time while monitoring an older person.

When it comes to the ethics concerning the issues of deception and infantilisation of the elderly, the Sharkeys cite Sparrow and Sparrow [26] who argue that "any beneficial effects of robot pets or companions are a consequence of deceiving the elderly person into thinking that the robot pet is something with which they could have a relationship.
Turkle et al [27] expressed a similar concern stating that

> the fact that our parents, grandparents and our children might say 'I love you' to a robot who will say 'I love you' in return, does not feel completely comfortable; itraises questions about the kind of authenticity we require of our technology.

Sparrow & Sparrow [28] argued that the relationships of seniors with robot pets,

> are predicated on mistaking, at a conscious or unconscious level, the for a real animal. For an individual to benefit significantly from ownership of a robot pet they must systematically delude themselves regarding the real nature of their relation with the animal. It requires sentimentality of a morally deplorable sort.Indulging in such sentimentality violates a (weak) duty that we have to ourselves to apprehend the world accurately. The design and manufacture of these robots is unethical in so far as it presupposes or encourages this.

Sparrow& Sparrow [29] point to some of the likely outcomes of the introduction of robots in caring for the elderly, stating that there are far fewer prospects for the ethical use of robots in care of the elderly settings than would initially appear to be the case. They argue that economic pressures would likely result in a reduction in the amount of human contact that an older person being cared for by a robot would experience.

One of their more controversial points is that "it is not only misguided, but actually unethical, to attempt to substitute robot simulacra for genuine social interaction"[30]. When writing about care and the human touch in elder care contexts, they identify two distinct areas of application: Firstly there is the 'physical services' which supplement the activities of residents or staff. Lifting and turning bed-bound persons, monitoring residents who are frail, or fetching or carrying heavy object. Secondly there is the caring and emotional dimension such as conversation, social interaction, sympathy and emotional support.They point out that under current models of care, the two often go hand in hand. For many older people, the only regular human contact they have is with the people who provide the physical care for them, with much of the human contact being provided by the cleaning staff. They point to studies that examined interactions between staff and residents in care-facilities and highlighted the fact that good communication is essential to high quality care and quality of life.They continue to expand on their third point, stating that;

> the intuition here is that what it is to be a real friend, or to really love someone, or to possess genuine rather than ersatz intelligence, is not something which can be exhaustively specified or captured by any algorithm or set of algorithms. Instead, what is required is that the candidate for these descriptions behaves in an (only loosely specified) appropriate fashion in a wide range of circumstances. Crucially, the forms of behaviour that are appropriate for someone, or something, who possesses the qualities necessary to be able to take on a caring role include some that only have their sense because humans (and to some extent, other creatures) are biological corporeal entities with particular limitations and frailties. Thus, for instance, if we care for someone, we reach out to take their hand, stroke their brow, wipe away their tears, or shed tears ourselves for them, when appropriate [31]. For robots to be capable even of imitating these responses successfully they would need to possess physical bodies capable of the same level of expressiveness and individuality as human bodies. Moreover, entities which do not understand the facts about human experience and mortality that make tears appropriate will be unable to fulfil this caring role.Sometimes the only appropriate response to another's suffering is the acknowledgement that we too share these frailties, as for instance, when our friend's suffering moves us to tears. Entities which do not share these frailties are therefore incapable of responding appropriately to them. Robots would therefore have to have a similar set of capacities and frailties as human beings in order to be capable of genuine emotional responses.[32]

When it comes to the question of supporting a delusional relationship between an older person and a robot, the Sparrows see two problems. Failing to apprehend the world accurately constitutes a minor moral failure stating that "it is a sad thing to be deceived about the world; it is a bad thing to perpetuate and prolong such deception ourselves"[33]. Secondly, they state that

> such deception is a bad thing because... preferences are unlikely to be met, our interests advanced, or our well-being served, by illusions. What most of us want out of life is to be loved and cared for, and to have friends and companions, not merely to believe that we are loved and cared for, and to believe that we have friends and companions, when in fact these beliefs are false. That is, we desire the real world to be a certain way and not just our beliefs about, or experience of, the world to be a certain way.[34]

4 RESEARCH METHODOLOGY

The research took the form of a semi-structured interview which was conducted at the nursing homes. Participants, who were drawn from a group of professional carers of staff nurses and activitity co-ordinators, were asked to view three separate videos of the CareBot, RIBA Robot and PARO Robot robots in operation. These robots were chosen as examples as they are primarily intended for use in key areas of elder care: home help or monitoring, lifting assistance in a care-home setting and therapy/companionship. Participants were then asked a

series of questions which were based on a-priori themes that formed the structure of the interview. Participants were free to talk around the topics as much as they wished. Audio recordings of the interviews were made which were later transcribed by a professional transcribing service. Participants were asked to respond to five questions that formed the basis of the interview;

1.What do you know about the use of robots with older people

This question was intended to elicit general knowledge from the participant on their awareness of the use of robots in care of the elderly.

2.How do you feel about the use of robots with older people?

This question was intended to give the participant the opportunity to express general opinions about the use of robots in a care setting; what were the potential applications of the technology in their opinion and to respond directly to the videos that they viewed.

3. What are your views on any ethical issues with robots being used to care for older people?If it is possible to achieve, do you think a robot needs a sense of ethical understanding in providing care for older people?

This question was intended to give the participant an opportunity to identify any ethical issues from their own professional perspective that they thought would arise from the introduction of a robot into a care setting.The second part of the question aimed to explore the perceived issues associated which might lead to a need for an ethical reasoning system to be embedded in a robot carer.

4. What are your views on robots being given full responsibility (decision making) for overseeing an older persons care?What are your views on the use of robots primarily to assist carers.

This question was intended to explore issues around robots being given clinical decision making responsibilities in caring for an older person. An example of a automatic defibrillator was used as a current instance of a purely technological delivery of a critical treatment procedure during which human intervention is not possible.

5.What are your views on the possibility of the development of a humanistic type of relationship between the older person and robot.

This question was intended to elicit opinions on the ethical issues of allowing an older person to participate in a fantastical relationship with a robot.

Responses to the questions were analysed using Template Analysis, which was considered to be the most appropriate framework for data analysis in this study. This framework includes a number of techniques for organizing and analysing textual data thematically and, it can be used within many epistemological positions[35]. Template Analysis requires the researcher to identify themes from the data in advance of analysis. These are also known as 'a priori' themes and indicate that the researcher assumes that particular relevant issues relating to the topic being studied are contained within the data.

Computer assisted qualitative data analysis software known as NVivo 8 [36]was used to manage and support the qualitative data.

5. MAIN FINDINGS

Theme 1: Knowledge about Robots

Participants overall had little to say in terms of their knowledge of the use of robots in caring for the elderly. They asked some incidental questions about the capability of the robots in question but generally their only experience was of the robots that were presented in the videos just prior to the interview. Although Participants were asked about their 'feelings' about the introduction of robots into the care environment, many began this part of the interview by offering responses that outlined what they thought was the most suitable or the most useful application of the robot in their work environment.

'surgical use' - *'Yeah I heard, I don't have much idea about this* and I have heard like you know in some countries they are using robots for the patients, robot*ic surgery something like that'* - respondent 3

Theme 2: Attitude to Companion Robot (CareBot)

*'*its probably good for basic things like for reminding people to take medication or different things like that but if a person is just being solely minded by *a robot,...I don't think that's right, it's not really showing dignity or respect'* - respondent 7

Theme 3: Attitude to Lifting Robot (RIBA Robot)

'Yeah because you still have the carer there like and they're able to fill another need for the patient or the resident if needed, if the robot can do 80% of things, then the carer can step in and do that 20%, you know the emotional needs or whatever. So it would be great to have an assistant like that, that can do all the heavy lifting for us (laugh*)'* - respondent 7

Many of the female participants responded in this way whereas the male nurse pointed out a flaw in the robot design in terms of the way it was lifting residents. There was little or no hesitation amongst the participants in accepting the RIBA Robot.

Theme 4: Attitude to Therapy Robot (PARO Robot)

'Yeah well I suppose it comes down to the individual, what their needs are and if that seal is providing a safeguard for them or a friendship or someone that they are caring for, well then if that makes their life good and their family are happy with it, well then I *don't have a problem with it. Its like they talk about doll therapy* today, some people would have a doll that they look after and you know some would say no, its childish, but like *I wouldn't have a problem with it because at the end of the day its, like we're here, we're trained, everything I suppose to be about the individual,* client-*centered approach, if they're happy to have it well then I wouldn't have a problem with it.'* - respondent 8

'I suppose there would be some ethical issues, I suppose in the sense of grieving if something happened to the robot or it wore out or *something, you can understand that its not a human that they'd be dealing with'* - respondent 9

Theme 5: Can't Replace Human Contact

This theme emerged out of the question of 'feelings' about the use of robots in caring for older people.

'*I wouldn't see at the moment or in the near future a robot* replacing a human carer, because I think again with older people *it's the human contact that is very, very important* because an awful lot of them have been isolated by bereavement *or by family, by virtue of the fact that they're in a home such as this, means that they don't have family support or they have an illness where the family can't support them any longer and that's a huge life changing experience anyway'.* - respondent 6

'contact maintains cognition'

'to maintain cognition, to maintain the ability that people have it has to be stimulating and I me*an if that's only a reminder that's* actually, the robot is doing their job but its not actually stimulating the person, its not maintaining the cognition or *whatever they have.'* - respondent 1

Theme 6: Explicit Ethical Issues

'if the robot is going to take over without the consent of the *Individual, looking after the person without that person's* consent or understanding of what the robot is doing, then it *wouldn't be acceptable'* - respondent 5

'if you think about ... do robots meet the psychological needs of the client or the resident and the emotional(needs), you know obviously the physical needs would probably be... looked after *because they're programmed to do X, Y and Z but whereas, you know they won't probably pick up if a person is upset* or, you know *what if a person can't communicate, we'd have residents here that don't speak and they're not able to communicate and you know we* have to interpret their body language, their gestures, what they *need'* - respondent 7

'Well I suppose if you thi*nk about it, you don't want older people* just to be objects, that you know to preserve their dignity and have respect for them, that you have people looking after them, *that they're not just objects that you can send in a machine to* look after them a*nd to interact with them'* - respondent 9

Theme 7: Should Robots be given Responsibility?

'You know and that's the concern I have with technology, you know that we then become too dependent on technology and that we would give over all responsibility to technology which is not an optimum situation either.You know technology is *there to assist us and I mean all of our technology that we're* going to be looking at, I think should be assistive technology and its another tool that we have in our tool box rather than something that takes over completely what we are supposed to *be doing, do you know what I mean'* - respondent 8

'I don't ever envisage that...not even ourselves have full responsibility or can take full responsibility for an older p*erson's* care as such, I mean our objective is to, you know involve the person as much as they possibly can be involved themselves in *making care decisions about their own care'* - respondent 8

6 DISCUSSION

In analysing the responses of the participants who were relatively naive with respect to the use of robots in elder care it was both surprising and heartening to discover that such naive respondents were 'finely tuned' to the ethical issues concerning the possible future relationship between robots and their patients at their places of work.

Respondents were generally naive in terms of their knowledge about the use of robots in elder care with most of their experience being restricted to the videos they were shown. Many respondents spoke of the use of robotics in other areas of medicine, however the source was generally media reports. In effect their knowledge base was clearly limited.

When respondents were asked about their 'feelings' with regard to the introduction of robots into the care environment, many responded with suggestions for the practical use that they saw for the robot in the work environment, suggesting that robots who fulfilled this purpose would be accepted into the community of workers. Although respondents agreed in general with manufacturers' intended use they raised a number of ethical concerns that could arise out of the use of these robots. In particular they were very protective of the potential loss of human contact that could arise from the assistive robots. This would correlate well with key issues raised in the work of Sparrow and Sparrow. Interestingly these concerns seemed to dissipate when the robot was primarily designed to assist carers, in the case of the RIBA robot, even though the potential for physical contact with the elder client would be reduced. Respondents were generally positive towards the prospect of their manual work load being reduced agreeing with a key point in the work of Coeckelbergh, that would suggest that humans are willing to admit robots to their moral community, if the robot assumed the position of 'slave'. In some cases respondents thought that this would create more time for carers to engage in social interaction with their elderly clients. Respondents' attitudes to Companion Robots as in the case of the CareBot were also primarily governed by a concern for a reduction in human social contact with the elder person. When asked about their attitude to the ethics of allowing a humanistic type of relationship to develop between the elder person and a robot respondents saw no overwhelming ethical concern if the elder person expressed satisfaction with the arrangement. This would be in direct conflict with the Sparrows who deemed this "infantilisation" of older people and in their opinion, is, without qualification, ethically incorrect. Citing the use of 'Doll Therapy' as an analogue of a practice that already exists within the health care systems for older people the respondents expressed some initial ambivalence towards this but again referred back to the concept of client autonomy and satisfaction as their benchmark. Abstract concepts of patient dignity, as outlined by the Sparrows in relation to the ethical concern of patient deception in the use of the PARO robot did not seem to correlate with the responses of care workers surveyed. Again turning to the notion of patient autonomy they generally saw

no issue or conflict if the patient was satisfied. However they did voice concerns about how the development of a humanistic type relationship between the older person and the robot might be viewed by other members of society i.e. family and other residents. Respondents drew attention to the high degree of non-verbal communication that takes place between carer and the elderly in institutional settings, voicing concerns about the capability of a robot to 'read' the body-language of another human. The implication of this was a lack of confidence in the current ability of artificial intelligence agents, however should this capability improve sufficiently as to provide human equivalent performance then their ethical concerns greatly reduced.

In another important point respondents stated that in dealing with patients with dementia that sometimes all that was needed was the touch of another's human hand. This has far reaching implications for the types of demands facing engineers in the development of assistive robots.

Ethical concerns were stratified by the respondents according to the functionality of the robots. Respondents generally expressed ethical concerns when considering robots as being used 'indiscriminately' implying that 'one size fits all' was not a viable concept in the production and use of robots in this highly sensitive context of human-robot relationships.

It was evident from their responses that the ethics of care, present in these settings, is intrinsically embedded in the work practices of carers. Overwhelmingly the concept of patient autonomy was to the fore in all of their considerations and responses and was frequently used as the benchmark against which they weighed their responses.

The responses of these naive respondents, highlighted and matched a significant number of the deliberations and narratives of experts in the field of Machine Ethics . A novel finding from this small-sample-size research was the discovery that if the field is to advance, the methods of Experimental Philosophy will need to be relied on more as a method for deriving the necessary information on the successful implementation of ethical comportment in the design of robots. It was clear from respondents that the contract of care, that they recognized as existing between them and their older charges, extended beyond a mere provision of service. Therefore, the danger in designing robots as service providers lies in the 'not fully grasping' of this concept.

When asked it robots should be given full responsibility, including decision making, in the care of the elderly, respondents overwhelmingly responded in the negative. In qualifying their response, carers stated their discomfort with critical decisions being the domain of non-human agents.

7 CONCLUSION

The methodology outlined by Template Analysis was fit for purpose in that it allowed for relaxed reflective exchanges between participant and enquirer. In a close reading of existing literature, particularly the work of Sparrow and Sparrow and Sharkey and Sharkey, the successful identification of range of a priori themes was facilitated by this methodology.

The importance of machine ethics is without question. Further research on modelling ethics in multi-agent systems could use an Ethics of Care as it's philosophical basis. Robots are likely to be members of our Moral Community long before they have any genuine ethical understanding.Multiple challenges face engineers in the task of engineering ethics particularly in gaining a full understanding of the nature of the care relationship that exists between a carer and their charge .A broader 'reading' of Ethics than is currently the case will be required by engineers attempting to produce artificial moral agents.

REFERENCES

[1][3][8][10] S .Torrance. Machine Ethics and the Idea of a More-Than-Human Moral World. , 1-24. To be published in M. and S. Anderson, eds Machine Ethics, Cambridge University Press.

[2][9][15][16][17][18][19][20]M. Coeckelbergh. Moral appearances : Emotions, Robots and Human Morality.Ethics and Information Technology, 235-241, DOI: 10.1007/s10676-010-9221-y (2010)

[4] J. Knobe & S Nichols. Experimental Philosophy. New York: Oxford University Press. (2008).

[5][35]N.King In Essential Guide to Qualitative Methods in Organisational Research(Eds, Cassell, C. and Symon, G.) Sage Publications Ltd, London, pp. 256-270. (2006).

[6]M.Luck.Computer Weekly.(2008)

[7]J.Moor The Nature, Importance, and Difficulty of Machine Ethics. IEEE Intelligent Systems, 21(4), 18-21. (2006)

[11] R. Sparrow & L.Sparrow. In the hands of machines? The future of aged care. Minds and Machines, 16(2), 141-161. (2006)

[12] [13]CIA World FactBook 2006 (2005): ISBN 1-57488-997-4

[14]KE. Himma Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent? Available via Social Science Research Network. http://ssrn.com/abstract=983503 (2007)

[21][24]A Sharkey & N Sharkey. Granny and the robots: ethical issues in robot care for the elderly. Ethics and Information Technology. (2010)

[22][27]S.Turkle, W. Taggart,C.D Kidd, &O. Daste . Relational artifacts with children and elders: The complexities of cybercompanionship. Connection Science, 18, 4, 347–362.(2006)

[23][26][28][29][30][32][33][34]R.Sparrow & L.Sparrow In the hands of machines? The future of aged care. Minds and Machines, 16(2), 141-161.(2006)

[25] E.J.Langer & J.Rodin. The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional setting. Journal of Personality and Social Psychology, 34(2), 191–198.(1976)

[31]R.Gaita. A common humanity: Thinking about love & truth & justice. Melbourne: Text Publishing, pp. 263–268.(1999)

[36]P.Bazely. Qualitative Data Analysis with NVivo, Sage, London.(2007)

A Robot Ethics: The EPSRC Principles and the Ethical Gap

Neil McBride[1]

Abstract: This paper posits a practice-based approach to robot ethics in which the ethics is understood with the boundary of the practice and the community of practitioners within which the robot acts. The approach involves creating a taxonomy of robot practice, identifying boundaries of practice, and creating a robot ethics ontology. The approach is set in the context of current robot ethics principles, a discussion of the limits of robot ethics and the limits that the uncodability of moral decision making.

1 INTRODUCTION

As robots in their widest sense take on roles in society, in businesses and home, the question of managing and defining their ethical behaviour in environments where they interact with humans or work on behalf of humans becomes increasingly a practical issue. Interactions in the home, with elderly, in society in education and childcare, and in war with autonomous weapons require that boundaries, rules of engagement, and ethical behaviour within relationships with humans need to be defined.

Discussion of robot ethics and ethical frameworks suffers from three problems. Firstly, the search for universal rules, often with reference to Asimov's rules for robot behaviour, results in such a generality that the rules are of little use beyond the support of good stories. The variety of application of robots, the range of environments, and the different human endeavours involved all render a simple set of rules of only superficial value. A recent EPSRC Robotics Retreat considered new rules for robots using Asimov's rules as a starting point. The first part of this paper will critically appraise those rules. Such rules will be seen to be of limited value since they are do not address the context and the practice in which the robot operates. It will be suggested that effective ethics requires consideration of the practices and how robot behaviour is developed and constrained within that practice.

Secondly, the properties of robots, and the apparent autonomy lead to a false perception that the robot does or can make its own ethical decisions, and that there is some concept of machine ethics which develops independently of humans. The ability of robots to move around in an environment and to take actions in response to varied stimuli can lead to an unhealthy tendency to anthropromorphise the robot. The output from the EPSRC robotics retreat is clear that the requirement for ethical behaviour resides with the creator, the robotist not the artefact [2]. As will be discussed, the limits of the Von Neumann architecture mean that the robot is only following a pre-programmed set of instructions. The illusion of humanness is created by the immense number of instruction that can be executed each second and the extent to which changes in the robot's environment can be anticipated and responses coded. The ethics of the robot must be the ethics of the maker. Two conditions could perhaps move ethical responsibility to the robot. Either the complexity of the coded architecture becomes too difficult for the robotist to understand and the behaviour effectively is emergent, or the architecture of the robot is based on an entirely different paradigm.

Thirdly, there is a limit to the extent to which ethics can be coded as rules, and it may be argued that ethical behaviour is uncodifiable. Decision procedures which determine the right action in a particular case may be adequate for the obvious cases. Ethical decisions require judgment and wisdom applied in the context of a community of practice. Such aspects of learning and judgement cannot adequately be coded. However, such uncodifiability leads to a robot ethics gap because any ethical action or avoidance by the robot must be coded because the architecture of the robot operates on coded instructions only. Currently, any ethical decision making must be translated into coded instructions where, like it or not, translation strategies must be developed which bridge the gap between human ethical decision making, grounded in practice, history and tradition, and machine ethics which requires precision and certainly, even if that can be somewhat disguised by the application of fuzzy logic.

Starting with a critical review of the EPSRC principles for designers, builders and users of robot, this paper discusses each of these three issues as a basis for proposing an alternative approach to robot ethics which starts with an understanding of the practice within which the robot operates, the boundaries of that practice and hence the characteristics a robot practicing in a particular environment should display.

2 THE EPSRC PRINCIPLES

The EPSRC principles for designers, builder and users of robots used Asimov's laws as a starting point and were seen as a revision of those laws. Asimov's laws state:

- a robot may not injure a human being or, through inaction, allow a human being to come to harm;

[1] Centre for Computing and Social Responsibility, De Montfort University, The Gateway, Leicester, LE1 9BH, UK

- a robot must obey any orders given to it by human beings, except where such orders would conflict with the first Law,
- a robot must protect its own existence as long as such protection does not conflict with the first or second Law.

The problem will these 'laws' is that they seem not designed to resolve ethical dilemmas and create foundations for rules concerning ethical decisions, but rather that they create dilemmas, the fodder of good stories, but not a framework for good ethical reflection. Most medical interventions involve some initial harm with the end result of benefiting the patient. Hence development of new robot ethics should really steer away from such literary devices, which the new EPSRC principles do not. However,. The EPSRC principles do raise a number of points which can be critically assessed and lead towards a more reflective framework for robot ethics.

2.1 Principle 1. Robots as killers

Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm
humans, except in the interests of national security.

The drive to develop military uses of robots constitutes one of the main forces in robot development. Clearly once such robots are available, their use may not be limited to just wars. The mere availability carries the risk of immoral use in crime and terrorism. Furthermore, robots designed to, say, eliminate vermin on a farm may be turned on humans. Control of such robots becomes an issue.

In most cases robots will be built by institutions whose goals are directed towards external goods, including profit and power and hence may not be ethically motivated. Ethical motivation should be driven by practitioners who address internal goods and practice ethical behaviour as part of a quest for the good life [5].

The fact that robots are multi-use tools requires a focus on the actual tasks undertaken at any point in time and the practice within which those tasks occur. Military robots operate within a specific practice. Within that practice particular virtues such as courage, loyalty and restraint will be appropriate and should be learnt within that practice.

2.2 Principle 2. The Ethical Robotist

Humans, not robots, are responsible agents. Robots should be designed & operated as far as is practicable to comply with existing laws & fundamental rights & freedoms, including privacy.

This principle clearly associates any ethical behaviour of the robot with the designer. The robot can only do what the designer wishes it to do. Hence reflection on the practice of the robot and the ethical effects is in the domain of the robotist. The robot being sensing, embodied and able to respond physically to its environment raises problems for the robotist who must anticipate the environment the robot will operate in. However, there are limits to the designer's ability to predict environmental changes. Robot can operate in volatile environments. The limits of prediction lead to a greater risk of morally inappropriate behaviour. It will be possible to control conditions or program a sufficient variety of responses.

The robot is a reflection of the designer's intent and hence it is the designer's ethics that are expressed through the robot rather than some autonomous machine ethics. The prime moral responsibility remains with the robotist; so a prime task of robot ethics is to develop ethical training for robot engineers which promotes reflective thinking and helps the designer gain sufficient empathy to understand the point of view of the human who will interact with the robot.

2.3 Principle 3 Robots as artefacts

Robots are products. They should be designed using processes which assure their safety and
security.

This principle underlines the fact that robots are manufactured product the same as cars or washing machines. Hence manufacturing processes and standards can be equally applied, underpinned by regulation and the law. Safety standards, testing, recall and repair, legal liability and so on should be the same as for any other major technology. Recycling and environmental considerations should feature.

Arguments concerning rights of robots and respect for robots [6] are amusing but of limited value. A child may drag around a robot dog instead of a teddy bear, but there should be no more risk of mistaking a teddy bear for a human than a robot dog for a real dog [1,8].

Stewardship of the robot will involve good maintenance, efficient use to get value out of the robot, and sustainable manufacturing and disposal the same as any other product.

Hence there is a moral responsibility for the same rigorous engineering leading to robust, reliable and sustainable robotics as would be expected from medical systems. Engineering standards and practice could well be drawn from the medical sphere.

2.4 Principle 4 Avoiding illusions and sleight of hand

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

Robots are only machines, executing billions of very basis calculations per second to take decisions within the boundaries of the options programmed into it. It is this speed of execution and response that can give rise to an illusion of free will, of humanness, which is only emphasised by its embodiment. A smiling robot responding to a human smile has no understanding of that smile or actual empathy with the human. Face and gesture detection, fuzzy logic systems, detailed algorithms can create the illusion of humanness. The robot is not sentient and can only imitate this.

The illusions is similar to that of movement in a film resulting from 24 frames a second being shown, There is a certain acceptance that what is seen on the film is a representation of reality but is not actually reality. There is a whole language of film making, In movies, meaning is expressed through various shots, signs and semantics. These are culturally accepted and understood. Such an understanding needs to be developed in robotics.

However, a naïve user may not understand how the robot is working and what the limitations are and may believe the robot is genuinely responding to him. This is clearly a deception and is dishonest. Transparency requires that the user can understand the limits and workings of the robot.

This requires well-designed user education through technical manual, videos, courses, that will help the user to understand what the robot is doing and what the limits are. Such education concerning the nature of computer architecture and the computational features of it need to be fed back into school. 'How it Works' session will help students to appreciate robot technology and use it effectively.

Failure to identify robot limits will be unethical, leaving the user open to harm and disappointment. Hence documentation and user training (like any engineering project) are an integral part of design and development.

2.5 Principle 5 Attributing legal responsibility

The person with legal responsibility for a robot should be attributed.

The legal responsibilities of the robot manufacturer should be no different from any other manufacturing. This will require the manufacture to understand the nature of the tasks undertaken by the robot and the boundaries of practice.

However such institutional goals, codes of conduct etc, must be underpinned by communities of practice within which ethical behaviour is situated and applied. The development of virtuous robotists, gaining in insight and wisdom, acts to inhibit corruption within the institute. A community of robotists cannot operate without the institutional structure providing the money and infrastructure for development. These can be confined by rules but require practicing communities within them to practice virtues and deliver ethical behaviour.

Hence our target in robot ethics should concern the development of reflective practitioners who are prepared to look beyond the confines of the engineering to consider the ethics of the tasks and the environment within which the manufactured artefacts will operate.

The principle of legal responsibility requires also the development of mechanisms of provenance and traceability so that manufacturing sources can be traced and there are audit trails of manufacturing and testing. Provenance should also apply social computing mechanisms to develop continuous dialogues between the robotists and the end users to enable continuous learning to occur as the robot act within dynamic and unpredictable environments.

3 LIMITS OF ROBOT ETHICS

The EPSRC principles focus on the nature of the robot as an artefact, and push the ethical responsibility away from the robot to the robotist. However, when the robot is out there, in the environment, acting 'autonomously', carrying out programmed actions, ethical behaviour should be in the frame.

As the robotist designs the embodiment and the algorithms for detecting multiple stimuli and responding to them, moral issues and limits should be addressed. This is precisely why reflective ethical practice needs to be developed and ethical principles encoded in the robot design.

Most robots, unless including neural networks, operate by running programs which act conditionally according to inputs. The robotist must decide what the inputs are and what the alternative actions are. The robot can only do what it is told by the robotist. This requires an understanding of a number of issues, and a number of questions to be asked:

- What assumptions am I making about unknowns in the environment?
- How well do I understand the environment to robot will operate in?
- What is my cultural and worldview, and the worldview of my design community?
- How can I understand or empathise with the worldview of the user community?
- What is the nature of the practice in which the robot will collaborate?
- Do I understand the underlying purpose, the telos of the robot?
- How valid are my predictions about the environment the robot will act in?
- What are the boundaries of the environment, the practice and the tasks?
- How do I make the limits of the robot clear to the user and avoid deception?
- Can I match the variety of moral decision making to the variety of the environment?

The results of considering these questions will be a set of programs which control the robot in its practice environment and allow it to select from a number of options according to particular stimuli.

The range of options generated by the robotist will inevitably be inadequate because the environment into which the robot will be placed will have unknowns and uncertainties, unless the environment itself is contained and artificial. Additionally there may be unexpected or emergent effects resulting from feedback loops between the environmental stimulus and the robot's response. There may also be problems with interpretation and understanding the meaning of environmental stimuli. Indeed, in interacting in a social environment semantic failures are inevitable.

To manage the ethical situations which the robot encounters, the variety of decision-making generated by the rules coded in the robot must match the variety of ethical situations that may be encountered in the environment. This is the basis of Ashby's ;aw of requisite variety. But the variety of ethical situations in any human society will be very large, if not infinite. Hence a robot expected to behave ethically in any human situation is bound to fail at some point because it is unlikely to match the variety of situations with the variety it can generate through a complex ethical rule set,

Robots using Von Neumann architecture are limited to following rules which in essence break down to the 0 or 1, yes or no answers which computer architecture requires. Von Neumann [7] noted that certain aspects of biological responses are qualitative, and do not require, or cannot be expressed in the quantitative, numerical way that the computers in robots require. Von Neumann concluded in 1957 that the language of the brain is not the language of mathematics. But the language of mathematics is the only language that the computer understands, and anything that cannot be reduced to that language, cannot be understood by the computer.

The expression of our language systems in computer code confers no semantic understanding autonomously on the computer system. The computer system only acts as a tool for transferring symbols and communicating meaning between humans. Therefore, morally all I can do is tell the robot what I would do in a given situation so that it acts as a dumb intermediary of the robotist's moral intentions and views. No moral understanding is conferred on the robot.

'In this situation do not shoot!". Because that is what the rules says. The robot has no understanding of why, or how that action relates to any moral frameworks, to culture and community, or even to basic human rights.

A moral code may be executed by a robot automaton, without ownership or understanding, just as a child can learn not to spit. Because Daddy said so. But if Daddy is not looking and the wish to spit is irresistible then he will spit. To move beyond not spitting because I've been told not to… but I may of no one's looking requires that I move beyond the rule to a moral engagement.

This moral engagement requires that I feel for the other person, the person who is spat on or who encounters my spit. It requires that I empathise with their point of view. It also requires an engagement with community, through the development of relationships so that I understand that spitting in unacceptable because it is frowned upon. I have seen what happens when others spit, I have experienced the disapproval. I also need to have understood the character of grown ups in that community. And I develop a qualitative understanding of virtues such as respect, self-control and empathy. It also requires a development of knowledge about hygiene, bacteria in saliva, disease, a knowledge that has a historical ancestry and has been developed possibly over generations.

Hence there is a difference between a behaviourist learning of a rule which may be developed in the young child who is smacked or put on the stairs for spitting and the social and moral wisdom and character which develops in maturity and requires community engagement and practice.

For a robot as a machine, none of this is available. It can only enact instructions issued by a moral agent, expressed in a social situation to another moral agent. It is a vehicle of communication of a moral framework between the robotist and the user expressed in actions or inactions.

So any apparent autonomy is an illusion and hence a deception, produced by sleight of hand, as the robot takes in a pattern of sensory information, scans that pattern against an array of possible environmental patterns and then executes a pattern of responsive behaviour dictated by the code associated with the match, the near match or the fuzzy match.

A robot meeting a moral Turing test is simply giving an illusion of moral behaviour. That illusion may be sufficient for the purpose of avoiding harm and fulfilling its defined telos.
However it does not involve understanding, maturity, wisdom, and insight.

In terms of expert systems, our problem is that we can only express the explicit, codifiable knowledge – moral or otherwise – in a system. The tacit, intuitive knowledge is excluded by the very nature of the machine.

Only if the technical paradigm changed and consciousness, self-reflection, awareness .. whatever you like to call it .. emerged would moral autonomy emerge. That paradigm will be very different from the machine implementations of robots we have today.

So, in the current paradigm, the robot can only transmit the ethics of its maker and make decision based on rules it is given, which can be expressed in binary code and where actions can be selected as 'Yes do it' or 'No don't'

4 UNCODABILITY

"The effort of trying to imagine someone reaching a correct moral decision by cranking through a decision procedure without exercising judgement brings us to another insight that moral knowledge, unlike mathematical knowledge, cannot be obtained merely through attending lectures and is not characteristically to be found in people too young to have experience of life" [3]

There is a growing awareness that efforts by enterprises to come up with a set of rules has failed. And the idea that moral decision making computer programs can be developed which will output a decision that can then be executed is difficult to support. Such an idea has been found wanting in expert systems which deal with empirical data (see for example, Ivandic et al, 2000), let alone decisions involving practices, history and tradition. Expert systems replacing humans set a dangerous precedent as we see in the Quant systems active in financial markets. All we should expect from computers is that they will help organise

information, raise issues and structure information to compensate for limited human memory.

The human, using the information, then forms a judgement which is rooted in the education provided in families, in communities and derived from social situations and tradition. And yet, when it comes to robot ethics, we are requiring the robot to carry out the ethical situation in an uncertain and volatile environment in the absence of human intervention.

Any codification of ethics, even in an expert system (see below) requires a numerical ranking of alternatives, a ranking of rules which will inevitably come up against cases which require a changing of priorities (Hursthouse, 1999). Codification requires a stripping out of the tacit knowledge, the moral knowledge derived from socialisation in culture, community and tradition and the experience which drives motivation and intent.

The problem then is that even in resolvable dilemmas the answer may not be in black and white and in unresolvable dilemmas all options may carry deleterious effects. And the moral consideration may involve the expression of a remainder, a regret that a virtuous person may carry concerning the decision that eventually had to be made.

There is then a gap between the rich, intuitive moral decision making which humans engage in and the impoverished mechanical moral decision making we are asking of robots. This is the codification gap in which the uncodifiable must be reduced to the codable in the robot. In reducing a complex moral decision (tacit, intuitive, deriving knowledge from maturity) to the execution of a set of coded instructions, we are throwing away vast stretches of knowledge, socialisation and learning not only built up in the individual, but also in the community and the history of that community, and replacing it with some naïve 'yes' or 'no' decisions which we might recognise as a substantial defect in someone with Asperger's or on the severe end of the autistic spectrum. We are asking the pure autism of a robotic machine to support the empathy, compassion, wisdom, patience, self-control and other virtues on which human moral practice is built.

The gap between the rich domain of moral knowledge, seeded by culture, learning, tradition, history and religion contrast with the impoverished domain of universal rules and codes of practice is reflected in the gap between complex human moral decision making and the mechanical obeying of rules. The stripping out of tacit knowledge, the socialisation, experience and wisdom leaves only the dry bones of axioms to code in the robot's instructions. Additionally, universal rules will either be uncodable or unusable, too general and abstract to provide any guidance to how the robot should react.

5 BRIDGING THE CODABILITY GAP

I would argue that, while uncodability restricts the potency of any autonomic machine ethics, some compromise must be found so that ethical consideration are part of the robot's programming and ethical element should be expressed in the robot's behaviour.

Characterisation and understanding of practice

An understanding of the practical ethics associated with robot action should start with an understanding of the practice (as defined by MacIntyre – not the skills) within which the robot will operate. A definition and understanding of the practice will confine the range of ethical situations which might be encountered and hence reduce variety to a level which may be responded to in the robot's coding. Steps towards an understanding of robot practice may include:

- Creating a taxonomy of robot practice. Defining and classifying the range of purposes for the use of robots in the service of human activities and identifying of relating characteristics which might lead to the identification of ethical issues.

- Identification of boundaries. Characterising practice will lead to an understanding of the boundaries within which the robot will work and the range of valid situations the robot may encounter. Defining and constructing boundaries is important in managing ethical variety.

- Creating a robot ethics ontology. Using the taxonomy to identify robot ethics terms which can then be given definitions. This may focus on the virtues and vices associated with robots active in practices. The language and methodologies of the semantic web could be applied in developing a robot ethics ontology.

- Identification of situations. Developing a range of situations within the practice which require ethical engagement. This will involve defining the environment, the sensory inputs and may be presented in a format similar to a risk catalogue.

6 DEVELOPMENT OF AN ETHICAL BRIDGE

A fundamental understanding of the relationship between a virtue, that is a character trait considered worthy of expression, the situation within the environment of practice and the action will need to be developed. If an ethical response is required, how do we jump from an ethical response to physical actions such as stop, sound an alarm, go backwards, hold on to patient or whatever.

This will require the design and implementation of an expert system which connects situations, virtues and actions. Such an expert system would support learning through the addition of examples from experience which will enable the variety in the expert system to match the variety in the environment of practice.

Robot learning will also occur by referring to the robotist and the user, looking at examples of good virtuous behaviour which can be expressed as situation/virtue/action triplets. This element of learning from the wise is a key part of virtue ethics, where the observation of example of good character are part of moral development.

However, the prime aspect of the ethical gap is that between the situation and the action. In bridging the ethical gap we recognise that the behaviour of the robot is never really autonomous but is made with reference to the robotist and the user.

A system will need to be developed which links the output of the expert system with programmable robot behaviour. This will involve interpreting sensory inputs and finding corresponding situations in the expert system; and the testing of that situation.

6 CONCLUSIONS

This paper attempts to bring together a number of aspects of robot ethics. Firstly I address the EPSRC robot ethics principles which while constrained by their pedigree as developments of Asimov's rules, clearly identify the relationship between the robot and the robotist and the derivation of any ethical actions from the robotist. Although the robot may be apparently engaging in autonomous ethical action, that action is limited by the programs defined by the robotist and based on the robotist's ethics and understanding of the environment.

The paper also discussed the limits of robot ethics and posits a gap between human ethical behaviour which in completely codable and the hard logic coding required by the robot's Von Neumann architecture. The identification of the uncodifiability of human ethics means there is a gap between the ethics-in-practice and what can be coded as sets of instructions and paths within the logic of the robot.

Finally, I address the need for a new approach to robot ethics, grounded in the understanding of practice, a characterisation of practice and a linking of situations, virtues and actions resulting from learning which is encased in an expert system.

References

[1] Blackford, R. Robots and Reality: A reply to Robert Sparrow Ethics and Information Technology 14, 41-51 (2012).
[2] EPSRC Principles of Robotics http://www.epsrc.ac.uk/ourportfolio/themes/engineering/activities/Pages/principlesofrobotics.aspx. (2010).
[3] Hursthouse, R. On Virtue Ethics Oxford University Press. 1999.
[4] Ivandic,M, Hoffman, W. and Guder,W.G. The use of knowledge-based systems to improve medical knowledge about urine analysis. Clinica Chemica Acta 297, 251-260. (2000).
[5] MacIntyre,A. After Virtue. Duckworth, 1981.
[6] Peterson, S. The Ethics of Robot Servitude Journal of Experimental and Theoretical Artificial Intelligence 19(1) 43-54 (2010).
[7] Von Neumann, J. The Computer and the Brain. Yale University Press, 1957.
[8] Young, J.E., Hawkins, R, Sharlin, E and Ugarashi, T. (2009) Toward acceptable domestic robots: applying insights from social psychology. International Journal of Social Robotics 1, 95-108 (2009).

Good reasons for making bad bots

Danny Weston[1], Catherine Flick[2] and Sam Waters[3]

Abstract. Web based 'bots' are currently ubiquitous online. They are used both at individual and industrial scales to gather, process and disseminate information and also to carry out human-like interactions on online discussion forums and other internet fora. This paper considers both the appropriate philosophical conceptualisation for such 'bots' and ethical issues surrounding their creation and use, in particular for data gathering 'bots'.

To understand what it means to even discuss or identify a 'bot' meaningfully, Actor Network Theory (ANT) is deployed to provide a range of descriptions of 'bots' as 'bot-like assemblages'. It is argued that a 'bot' cannot be conceived correctly in isolated terms of simply the code used and moreover that the 'bot' as both 'actor' and 'actant' only means something in the context of the wider network within which it is placed.

Following on from this, an ethical account of 'bots' so conceived is argued for on the basis of Floridi's 'Information Ethics' (IE), where epistemology and virtue ethics are combined to make a case for the preservation of data – such an account also provides a normative direction for 'bot' development and deployment more generally. A practical example is used to highlight the usefulness of a conjunction of ANT and IE in the description and analysis of 'bots'.

1 Introduction

The perennial concern in robot ethics for philosophers goes "beyond concern for what people do with their computers to questions about what machines do by themselves" [1] For the purposes of this discussion, 'bots' here refers to programs – generally active online rather than restricted to a particular network or single computer – that act without continual human supervision. Much of the literature in computer ethics focuses primarily on physical robots and their actions in the world, according much less attention for their disembodied cousins in the realm of software that acts within networks.

In the 'embodied realm' of the robot it is generally quite straightforward to delineate a robot from its environment and its behaviours from its effects on the external world and other agents even if it is still controversial as to how to similarly delineate programmers' responsibilities and accountability from their creations. This is not the case for programs with something of a life of their own when 'in the wild'. And it is for this reason that a program cannot be considered a 'bot' without also considering its relationship with its environment. Almost any program's actual behaviour can be altered dramatically by a change in environment. A straightforward analysis of source code, which is itself a kind of abstraction, can only lead to a similarly abstract conclusion with regards to what the program's effects will be. This means that, with regard to the multitude of environments any particular piece of code may find itself in, it could quite easily be classified as 'bot' like, acting in an independent fashion without human supervision. Defining a 'bot' is then less about the source code and more about its ultimate behaviour – something any programmer will have difficulty predicting in advance with no knowledge of the eventual end environment(s) in which it will be deployed. We will drive further into this point in the following section when attempting to characterise the issue in Actor-Network terms.

A criticism often raised against Actor-Network interpretations is that they are essentially amoral. To some, 'amoral' always reads as synonymous with 'immoral'; a hidden premise that without any guidelines a chosen behaviour is inevitably more likely to be immoral than moral by outside reckoning. Latour attempts to address this in more recent work [2],. Underlying the practice of Actor-Network Theory (ANT) is the attempt to be descriptive and not - specifically not - prescriptive. It is on these grounds that we suggest Floridi's Information Ethics is a suitable (though by no means not the only) addition to subsequently give any such analysis of 'bot' behaviour a normative direction.

2 The 'bot' and the ANT

Harmless code can quickly be rendered dangerous with even a minor accidental alteration or when placed in an unanticipated environment. For example code designed to benevolently copy a file a limited number of times may be accidentally called via an infinite loop in another program leading to a rapidly filled hard drive and possibly a slower computer too. To consider this in Actor-Network terms means extending the meaning of 'network' beyond the other hardware and software that we have thus far been referring to as the 'bot's potential environment. Human (and to some extent non-human) actors outside the virtual space also form a key part of the total assemblage. In the ANT lexicon, the assemblage is often (though - importantly - not always[4]) 'socio-technical'. And it is this latter aspect that comes to the fore when 'bot manners' are considered briefly later in this paper.

An important deviation in terminology should be noted at this point – ANT's history begins sometime before computer networks became ubiquitous. As such – and as Latour notes with some evident frustration [2] – the term 'network' has since become synonymous with a computer network in the popular

[1] Department of Communication and Creative Arts, University of Greenwich, England. Email: danny@censoring.me

[2] Centre for Computing and Social Responsibility, De Montfort University, England. Email: cflick@dmu.ac.uk

[3] Department of Philosophy, University of Sheffield, England. Email: pia08sw@sheffield.ac.uk

[4] Many 'bot' and virtual activities take place without direct human intervention (aside from those who start the processes in motion) such as for example load balancing servers that monitor network activity and adjust how many resources are deployed to cope with variable demand.

consciousness and it is inappropriate to confuse this with the meaning of 'network' in 'Actor-Network'. He argues that:
"With Actor-Network you may describe something that doesn't at all look like a network—an individual state of mind, a piece of machinery, a fictional character; conversely, you may describe a network—subways, sewages, telephones—which is not all drawn in an 'Actor-Networky' way. You are simply confusing the object with the method. ANT is a method, and mostly a negative one at that; it says nothing about the shape of what is being described with it."

Latour has proposed substituting the term 'worknet' to designate the original ANT meaning – a network of actors (and actants) that have to continually do work (change) to remain a meaningful network at all. Sometimes a particular instance of computer network infrastructure will fit this definition perfectly, other times it will not (such as when the network is passive and no signals are being exchanged, or nothing changes at all). We also, Latour argues, have a specific shape in mind when thinking in current terms of 'network' – usually something resembling a spider web with two or three dimensions. He refers to this as a 'networky shape' – and again, it is something that may be designated an Actor Network but is not synonymous with it. An Actor-Network could just as well have a single bi-directional line between actors with no additional redundant links. On that basis then we will be using the term 'worknet', as per Latour's suggestion, to refer to ANT networks and 'network' to mean all varieties of computer network.

To reiterate – primarily, a 'bot's behaviour is the 'bot'. One could write the most malevolent code (for example something that is designed to replicate itself across a network and at a set time in the future wipe the hard drive of all infected computers), yet the code itself could not reasonably be referred to as a virus or worm unless it is actually launched and – moreover – unless it actually successfully carries out its hostile task. There is also the fact that these functions of the program taken singly have legitimate uses – for example one may wish to wipe the hard drive prior to installing a new operating system or some anti-virus programs may adopt a similarly viral strategy of replication in order to catch up with their target virus(es). Up until the actual point of deployment it is just so much more code and network traffic idling as electrical potential in memory somewhere.

The worknets of ANT are always in flux and – it is asserted – it is only through such changes that they can be observed at all as they leave 'traces'. Any and all objects can be considered 'agents' and part of a 'worknet' in ANT as well as humans even though Latour et al. will acknowledge that the former do not have any intentionality as such, they do however have effects[5]. Additionally, they all form associations with other actors to form worknets which in turn link to other worknets also consisting of both humans and non-humans. It is important to note that no methodologically significant distinction is made between these two types of agent. Instead ANT theorists identify two types of observable agents – mediators and intermediaries. The former effect some change on a signal or behaviour whilst the latter simply repeat or transmit it. 'bots' can be both – both by dint of their original programming and their interactions 'in the wild'.

Actors build networks (worknets) but are not necessarily constrained or predetermined by them as they may be in a computer network. The 'machinery' of the network however may well act on particular actors. This is another important subtle difference between the two meanings of 'network'. 'Machinery' means to operate in a 'machinic' way in the subtle Deleuzian sense that a collection of actors can act as a Machine[6] in a very abstract way and possibly without any 'hard' objects at all, though certainly many moving parts (n.b. 'moving parts' here includes humans and their (changing) associations with one another). The worknets of ANT allow their components to act together to produce a consistent, (and usually reproducible in similar worknets), effect. This can be something as concrete as a hardcopy document going through multiple revisions via members of a team in an office or something more obscure such as a worknet that transforms beliefs into facts that are taken for granted when its (this 'machine' worknet's) components are forced to act as if in agreement.

The 'actors' in ANT are almost always 'actants' - things made to act by other actors that themselves may well also be actants in most other respects. To be made to work together, actors (actants) may have the way they act changed in this 'bringing together'. A corporate structure for example constrains and changes the behaviour of actors within it in order to operate a consistently replicable worknet. Agency is often regarded in ANT as something that is an effect of worknets, not necessarily prior to them.

One of the most obvious 'traces' (representations) of the behaviour of actors and worknets is recorded data. It is the direct result of interactions of some form or another and so allows the researcher to begin tracing them. This particular aspect provides something of a paradox in attempting to apply ANT to 'bots': there is no more richly comprehensive and complete set of traces than those to be found digitally and online. However, most traces of digital activity are easily modified or even deleted permanently. There is a severe epistemological problem here, especially as so many aspects of digital activity are self-referential. The 'traces' of activity on a Microsoft Word document for example, are contained within the document object – and modifiable by anyone with the requisite knowledge. There is no external arbitrator except in the uncommon cases where strict version control is enforced – for example, on a wiki, or using revision control software to keep track of every change to a piece of software. Such a limitation means a typically foundationalist or correspondence based epistemology is useless outside of carefully audited electronic domains such as the examples just given. 'Bots' are most often used online to gather and disseminate information – information which is itself often of doubtful provenance and best understood in its own terms and without direct reference to the external non-digital world. We will come back to this point in considering why and how to apply information ethics in an Actor-Network 'worknet' characterisation of 'bots'.

To assist in conceptualising how to trace and understand 'bots' in an ANT context, it is worth digressing slightly to a more 'low-tech' example: Sismondo relates a useful example [3] in science

[5] Latour also argues that intentionality is a distraction, an attempted bridging for the 'subject/object' divide that he considers to be unnecessary philosophical baggage and the source of misery for most epistemologists.

[6] see for example Welchman, A, 'Machinic Thinking' (ch 12) in 'Deleuze and Philosophy: The Difference Engineer'. Ed. Keith Ansell Pearson. London: Routledge, (1997)

for comprehending ANT via Latour's description of Rudolf Diesel's attempts to create a new kind of engine. Diesel initially designed his engine for use with any fuel at high pressure. But not every fuel ignited under such conditions and Diesel had to revise his plans. Latour refers to these physical objects, substances and limitations as "allies" and as Sismondo puts it, for Diesel: "Diesel's alliances include entities as diverse as kerosene, pumps, other scientists and engineers, financiers and entrepreneurs, and consumers. The technoscientist needs to remain constantly aware of a shifting array of dramatically different actors in order to succeed." This illustrates the potentially ever changing nature of many worknets. Diesel's eventual success relied upon a continually changing plethora of human and non-human actors and differing "alliances" between them. As Latour originally stated it: "This ally, which he had expected to be unproblematic and faithful, betrayed him… So what is happening? Diesel has to shift his system of alliances"[7]

The properties of physical objects, such that they can be determined, appear in the context of tests, not as properties in and of themselves in isolation. Realism however is definitely prior in the ANT conception because physical objects are able to "push back". If one thinks of human and non-human actors as equivalent, it is easy to see what Latour means when he describes physical objects of study as having "interests". Researchers in the 'hard' sciences attempt to isolate and manipulate the 'interests' of these objects, however they (the objects) resist that manipulation and push back against the worknet trying to manipulate them (of which the scientist is part). One would think naturally that the opposite is the case with humans, however it is not. The hard sciences have also become powerful because their translations tend to be rigid (unlike the social sciences) and alternative translations are easily detected and/or amended where necessary.

Such attempts to enforce rigid translations can be seen in computing: Internet and networking technologies are normally portrayed in computer science using the seven layer OSI model[8], or similar. The most basic layer is the physical – the literal hardware used to transmit the electronic communications. This layer 'pushes back' in the Actor-Network sense. Every layer above this, however, is a relatively unknown area, and is determined largely by its participants – both human and non-human; constrained only by the limitations imposed by the hardware – limitations themselves that are continually changing (for example, the jump from dial-up 56k modems to fibre-optic internet connections in less than two decades – and to paraphrase Shirky, with regard to internet and communications technologies, more isn't just more. More is different[4]). Everything in layers above the physical layer is a series of negotiations – and choices and trade-offs in software, protocols and so on. The layers are best understood as providing a series of constraints that must (generally speaking) further constrain possible activities on higher layers. It is important to note however that causation is bi-directional – something modelled very well in ANT, which argues that an actor or worknet on one level of abstraction can affect actors or worknets on other levels of abstraction above or below it if enough "allies" are gathered. The OSI model is also only one translation model / protocol amongst many (for example the TCP/IP protocol[9]) – all of which would provide very interesting worknets in themselves for future Actor-Network analysis and research. It is possible for example to make the most basic separation – between a basic hardware layer and a virtual software layer that sits on top of it. However, aside from being less interesting such a two-tier model also teases out far fewer of the important aspects that would be of concern to anyone studying the epistemological and ethical status of virtual objects and actors.

This is where Actor Network Theory can really help to articulate the problems in the philosophy of computing and technology and possibly propose some solutions, though it may require indulging some philosophical behaviour that is more awkward than most analyses of these issues carried out so far: namely insisting on only ever considering the issue from the point of view of an assemblage – and one that, for reasons already identified, cannot rely heavily on a foundationalist or correspondence epistemology. Commenting on either just the 'bot' software itself or the environment in an isolated manner would mean insisting on only the most limited conclusions. Meaningful extrapolations of what the 'bot' may or may not do must be reserved for considerations of the total assemblage, rendering the kind of isolated software study advocated by Berry as somewhat limited and perhaps more suited to discussion of software in terms of a form of literature, especially as Berry's analysis makes a number of assumptions regarding software development [5] – such as for example, the assumption that most software will have gone through the full software development life cycle. This cycle, whilst taught on many computing courses, is nevertheless an ideal and only ever carried out when there is time and money to spare. Most software development in reality is – to use the lexicon of Perl programmers – 'quick and dirty' and any academic analysis of it (and in particular of the comments left within software code by programmers) should start from this understanding or go badly astray with misplaced assumptions.

Philosophical accounts may not be so arduous when tracing one worknet or another that involves the use of 'bots'. However adding the ethical and legal dimensions creates additional layers of complexity and it is at these two levels (the legal in particular) that most of the cutting edge research and commentary is taking place. The intractable legal issues already making headlines on digital technology issues for some years now mean that legal academics and professionals deal with direct conflicts and distinct shortfalls in the underlying philosophies available. It is no surprise then that much of Benkler's work, to use an example, [6] delves quite deeply into philosophical issues out of necessity. Their contributions could be much better informed and improved through basing them first on an Actor-Network analysis. The subject/object philosophical baggage that Latour so consistently dismisses [6], for example, haunts the work of Benkler, Lessig and other legally focused researchers in this area.

[7] Latour (1987), p.123 (original emphasis), cited in Sismondo (2003) [3]

[8] See for example the Wikipedia page, which provides a comprehensive description of the OSI model: http://en.wikipedia.org/wiki/OSI_model (last accessed 25/05/12)

[9] A particularly interesting difference between OSI and TCP/IP is that the former was intended to be prescriptive whilst the latter was intended to be descriptive.

3 A normative direction for ANT-'bots'

Luciano Floridi's early work on Information Ethics (IE) [8] is a helpful framework in which to discuss the kind of broad stroke philosophical engagement being attempted here. This is especially true, as we are attempting to link ANT and IE, with ANT facing the same kind of difficulties of acceptance and breakthrough now in mainstream anglo-analytic philosophy that IE faced [9]. Floridi makes a solid case for arguing that computing technology and the ethical theories (the 'macroethics' theories) mark a fundamental break with the past history of ethics in philosophy: "ICT, by transforming in a profound way the context in which some old ethical issues arise, not only adds interesting new dimensions to old problems, but may lead us to rethink, methodologically, the very grounds on which our ethical positions are based.".

The typical view of macroethics, Floridi argues, dismisses the philosophical significance of both ICT and a specific branch of philosophy dedicated to it, holding that it is at most a microethics within larger, already well established macroethical structures.

Floridi argues that ICT related problems actually "strain the conceptual resources of action-oriented theories more seriously than is usually suspected". The reason for considering a new series of approaches to philosophy in general, and his Information Ethics in particular, is that other ethical philosophies are both "agent" and "action" focused. Computer and communication systems end up being anthropomorphized inappropriately and human agents' sense or moral responsibility ends up being diluted for failure to take seriously the mediating and virtual role and characteristics respectively of ICT. The processes are remote and immaterial and so difficult to conceptualise correctly in the anthropomorphic frames set by the classical macroethics philosophies (not to mention that the same applies to epistemology and other philosophical domains too which is why we argue that ANT provides a necessary palliative in those other areas).

Floridi summarises it thusly: "a person may wrongly infer that her actions are as unreal and insignificant as the killing of enemies in a virtual game." And indeed this can be seen illustrated graphically in the numerous data protection and data loss incidents reported over recent years[10] – these largely happen as a result of this way of thinking as the issues are not regarded sufficiently seriously.

By way of corollary similarly the diffuse and anonymising capabilities of such technologies make violations difficult to detect (and – importantly – less public) and sanctions very difficult to impose. Action and human oriented philosophies do not treat information and virtual entities more generally as real objects and are thus necessarily myopic.

Floridi interestingly referred originally to his philosophy as "object-oriented" – just as Harman in describing Latour's ANT metaphysics refers to that also as "object-oriented philosophy"[9]. This is one of several striking pieces of continuity and compatibility between the two philosophies. Latour argues for a conceptual basis of a 'parliament of things' (and, we would argue, more concretely and accurately describes this kind of ontology than Floridi does through notions such as 'black boxes', 'allies' and 'trials of strength' allowing for explanations of divergences in relative capabilities and effects at whatever abstract level one picks). Similarly Floridi argues for a computer ethics based on a level playing field of information entities and also regularly invokes the notion of 'Levels of Abstraction' for apprehending info-sphere entities.

Such an equalising philosophy allows then for what Floridi refers to as a 'patient focused' ethics. Classical macroethics generally focuses only on the agent and even in the form of consequentialism relates everything back to the agent and also commits them to the difficulties of maximal and supererogatory outcomes. Information Ethics is "allocentric" – focusing on the entity itself that is on the receiving end of the action. Information and not just agents that are alive are here regarded as possible recipients.

Diving into the meat of the IE directives then, Floridi argues that they deal not so much with 'right' or 'wrong' as what is actually better or worse for the 'infosphere'. And – following Latour's principles of 'black boxes' and various levels of abstraction – this can refer to arbitrarily sectioned off elements of the 'infosphere' (e.g. the content on a single website). Further, he argues that despite widening the objects of ethical concern beyond the purely anthropocentric and thus being improvements in that sense, bio and environmental ethics nevertheless disregard "what is inanimate, lifeless or merely possible".

Whilst Floridi does attempt to consistently stretch this conception to all entities – and many (we included) do not find this comprehensively or consistently defensible – it is not necessary to do so for discussions dealing primarily with the realm of ICT. As an ethical tool, IE is tremendously useful and consistent when referring to data entities and the agents that interact with them (and for the purposes of this paper we will consider human actors as acting from the point at which they interface with an ICT system) so as not to have to address the concerns with a broader and necessary appeal for IE beyond ICT objects themselves. As Sara Baase argues in 'A Gift of Fire', with ethics in general and the ethics of technology in particular, it is best to pick the right tool for the job in hand. [10]

IE articulates a list of directives that all focus on maintaining the integrity[11], and – if possible – improving the quality of information in the 'infosphere' with which one is dealing. The primary locus – and test - of morality here is to consciously avoid imposing entropy and indeed to minimise it where possible. Slotted within an ANT view of technological and informational objects then, this gives us a clear normative direction that is lacking in ANT on its own, even if (just like ANT) it has to be worked through for each unique situation and worknet (or 'infosphere') in question. An ANT analysis provides an excellent framework for describing and conceptualising what the object(s) of study is. IE can then be applied to understand what agents should then be able to do within it.

[10] See for example 'List of UK government data losses' - http://en.wikipedia.org/wiki/List_of_UK_government_data_losses - last accessed 25/05/12

[11] Those principles being (from [8]) - 1. uniformity of becoming, 2. reflexivity of information processes, 3. inevitability of information processes, 4. uniformity of being, 5. uniformity of agency, 6. uniformity of non-being and 7. uniformity of environment

4 Practical examples – 'Churnalism' and ''bot' manners'

One domain of seemingly intractable problems in this area is created by the fact that a 'bot' that may, on its surface, appear able to engage in only perfectly legal and ethical behaviour when analysed both by computing professionals and in the more philosophical-descriptive analysis following Berry [5], yet once out in the network and actually combined with said network may in fact result in consequences that would be considered illegal or unethical. We suspect a Latourian analysis of such cases would be to simply see the legality or illegality as yet another interesting dimension to describe, not a show stopper as it is for Benkler and others. And this of course, returns us to the 'amoral' charge levelled at ANT analyses. So now that we have outlined a basis in ANT with which to consider how to apprehend what a 'bot' is, we now turn to considering briefly whether and how Floridi's Information Ethics could fill the gap for some example cases.

A pressing concern in computer ethics and for civil liberties is the advent of 'mass dataveillance'[12], where the monitoring of an individual becomes far less significant - in both import and consequences - than the gathering of masses of data about lots of individuals [11]. This is because those masses of data are able to reveal patterns and links that the individual members of the data set are very likely unaware of. Our purpose is not to discuss the privacy implications here (nor Floridi's treatment of such in IE). Instead we wish to highlight how such 'mass dataveillance' is in fact a practice open to any individual who learns to program in a web language and how this may fit into an ANT and IE understanding. There is a tremendous amount of information in the public domain that is awaiting collection and analysis. IE gives us directives to follow and ANT a way of contextualising such work. Such work however may already lead one to see why the 'bots' created to carry it out may already be regarded as 'bad', even if under IE they are carrying out a morally commendable task.

Nick Davies, in his book, Flat Earth News, [12] popularised the term "Churnalism", in which he described how journalists "....are no longer gathering news but are reduced instead to passive processors of whatever material comes their way, churning out stories, whether real event or PR artifice, important or trivial, true or false".

In IE terms this is very bad. More entropy is creeping into the media infosphere, especially as more content becomes available, and is shared, online. Davies' book was substantially based on research he commissioned at the University of Cardiff [13]. In the course of that research, amongst other things, what they found "suggests that 60% of press articles and 34% of broadcast stories come wholly or mainly from one of these 'pre-packaged' sources." Upon reading this research over a year ago, it was realised that analysis of this could be automated. And not only could it be automated, but it was hypothesised with enough data, direct patterns of bias and influence could also be detected.

Churnalism.com is a website that plugs into the 'Journalisted' database[13], where every single press item is archived online. The Churnalism engine uses an algorithm that enables anyone to manually copy and paste text from any source (though most usually a press release) and compare it quickly to the entire 'Journalisted' database. It is then able to report back any cases that are likely to have been copied, along with an estimate of the proportions copied. It allows people to effectively trace the provenance of many news stories back to press releases and also assess how much has been copied directly into the article.

This facility was used by automated 'bots' that were directed at two major organisations' press releases and submitted automatically to the Churnalism engine for analysis, and the results then collected en masse. These organisations were picked (with the intention that the work be easily replicable for other organisations), because they had been highlighted in the news at the time for issuing press releases with doubtful claims that were then repeated throughout the British mass media – DEFRA and the Environment Agency[14].

The full results are available online[15], but, having collected every single press release available on the websites of both organisations it was possible to detect distinct patterns in which media outlets were more inclined to copy and paste the press releases, and even to identify the general tendency to copy more or less content. This showed how, were other people to carry out a similar analysis of other organisations that issue press releases online, a citizen research (crowdsourced) analysis of publicly available data could be useful in 'mass dataveillance' terms and sits quite easily within an ANT ontology and an IE ethical framework. The data – once collected and analysed – revealed previously hidden, and informative, patterns.

At first glance, in IE terms this is clearly a desirable practice. Entropy has been revealed and potential corrections offered in the online media infosphere related to stories originating from the two organisations in question. The story becomes a little more interesting however with regard to the behaviour of the 'bots' and the response to them:

How 'bots', in this context information gathering 'bots' (web spiders), should behave online is still an evolving and wonderfully vague grey area. Generally some kind of civility is maintained between 'bot' writers and webmasters by following informal rules of netiquette, some examples being: that spiders should read and respect requests contained in the 'robots.txt' file to not index or collect data from the proscribed folders and pages, that the number of requests submitted by the 'bot' in any one time period should be limited (one proposed limit is once every three seconds for a request), that data collection should if possible take place at a time when the server is less likely to be busy, that the 'bot' should identify itself as a 'bot' and not a browser (thus potentially imitating a human) and so on.

'Bots' that break these rules can be blacklisted or restricted, the IP addresses from which they are launched blocked and those who write or launch them can even be prosecuted for interfering with the operation of computer systems. The problem is that none of these aspects are anywhere near clearly delineated sufficiently. Take the last case for example. In the recent spate of Denial of Service attacks on numerous websites by the hacker

[12] The chapter on 'Privacy 2.0' in [11] is particularly informative on this issue.

[13] http://journalisted.com/ - last accessed 12/05/12

[14] For example – from the Environment Agency 'Llamas help protect ice age fish'- http://churnalism.com/rk943/ and from DEFRA 'New service for householders to stop unwanted advertising mail' - http://churnalism.com/tt696/ - both last accessed 25/05/12

[15] http://www.censoring.me/churnalism

group Anonymous[16], the intent was very clear – and the attacks (mis)used a characteristic of the underlying networking technologies to achieve their goal – by flooding the target servers with SYN/ACK requests, something that can be done at very high speeds. And yet perfectly legitimate data-gathering requests could be sent by another party using a clearly identified 'bot' that sends requests only infrequently in a situation where the server happens to be under a large amount of strain. If the server subsequently crashed or stopped responding, some webmasters may regard the 'bot' activity as hostile.

Similarly, consider the 'bots' deployed for the Churnalism research above. They were launched at times when the servers in question would be under less strain and maintained a choke of one request every three seconds. Despite this, there was a response from the webmasters of both the Environment Agency and DEFRA websites – an attempt was made to block access from the IP address used to the sections of the website that contained the press releases. This raises questions as to why – and whether they have a moral case – for doing this. Under common 'netiquette' for 'bots', these are being regarded as 'bad' 'bots', even if they are 'good' in an IE sense. IE is only of limited help here also as it would nevertheless advocate the continuation of data collection especially as no direct harm was coming to the 'information entity' that was the target server, entropy was potentially being reduced and the data is also publicly available (and publicly funded in these cases). This is where ANT comes to the rescue: the entire situation must be regarded as an 'assemblage' of the webmaster, the server, the 'bots' and the programmer launching the 'bots'. As highlighted earlier in the paper, the 'meaning' – and even one might say, the 'identity' - of the 'bot' now only becomes clear in this totalising context. The 'bot' cannot necessarily be identified beforehand as a badly behaved program without the context of the server it queries and the state that is in, along with the competing priorities of the webmaster – which appeared to be in this case, a realisation that the 'bots' were scanning every single press release and an – as yet undisclosed – reason for stopping it[17]

By way of comparison, when the press releases were submitted en masse to the Churnalism site, there were no attempts to restrict access. It is also worth bearing in mind that the site received twice as much traffic from the 'bots' as data for the target sites were both submitted to this one. This presented an entirely different assemblage and indeed the team behind the Churnalism.com facility have expressed their desire previously for people to make as much use of the site as possible. Whether the 'bots' would have been regarded as 'good' (or at least neutral) or not by the Churnalism webmaster is unknown at present, even if IE would also assess them to be so.

[16] See for example, BBC News, 9 December 2010, 'Anonymous activists say Wikileak war to continue' - http://www.bbc.co.uk/news/technology-11935539 - last accessed 15/05/12

[17] It is worth consulting the Brooke (2010) [14] work on this, particularly chapter 5, on how similar efforts have been stymied in the past such as using publicly available data to create 'the public whip' website (http://www.publicwhip.org.uk/).

5 Conclusion

This paper has shown how Floridi's Information Ethics and Latour's Actor-Network Theory can be used to usefully describe and analyse Web based 'bots'. It is important to be able to consider the 'bot' outside of its immediate code and to visualise it in broader terms of the environment in which it is placed in order to understand its ethical impact. Such an 'assemblage' environment ought to not only include the technical and socio-technical systems but the humans involved, as their understanding and view of 'bots' will have an effect on the behaviour and thus impact of the 'bot' in question. The examples, show, using this framework, that even if 'bots' can be developed that are, in an IE-sense, 'good', they can be determined as 'bad' according to the 'bot's target. Further research and analysis is required to establish the efficacy of the ANT-IE analysis framework for 'bots' in more complex socio-technical systems, and to understand the reasons behind decisions made by the human actors in the 'assemblages'.

6 Acknowledgements

This work was funded as part of the RCUK Digital Economy project PATINA, grant number EP/H042806/1.

REFERENCES

[1] W.Wallach and C.Allen. Moral Machines: Teaching Ro'bots' Right from Wrong. Oxford University Press. (2009).
[2] B. Latour. Reassembling the Social: An Introduction to Actor-Network-Theory. OUP Oxford. New Ed edition (2007).
[3] S. Sismondo. An Introduction to Science and Technology Studies. Wiley-Blackwell; 1 edition. (2003)
[4] C. Shirky. Here Comes Everybody: How Change Happens when People Come Together. Penguin. (2009).
[5] D.M. Berry. The Philosophy of Software: Code and Mediation in the Digital Age. Palgrave Macmillan (2011).
[6] Y. Benkler. The Wealth of Networks. How Social Production Transforms Markets and Freedom. Yale University Press (2007).
[7] G. Harman. The Prince of Networks: Bruno Latour and Metaphysics. Re.press. (2009).
[8] L. Floridi. Information ethics: On the philosophical foundation of computer ethics. Ethics and Information Technology 1: 37–56. (1999).
[9] G. Harman. The importance of Bruno Latour for philosophy. In Cultural Studies Review Vol 13, No 1. (2007).
[10] S. Baase. A Gift of Fire: Social, Legal, and Ethical Issues for Computing and the Internet. Third Ed. Pearson. (2009).
[11] J. Zittrain. The Future of the Internet. Penguin. (2008)
[12] N. Davies. Flat Earth News: An Award-winning Reporter Exposes Falsehood, Distortion and Propaganda in the Global Media. Vintage (2009).
[13] B. Franklin, J. Lewis, N. Mosdell, J. Thomas, and A. Williams. The Independence and Quality of UK Journalism. Cardiff: Cardiff University. (2006)
[14] H. Brooke. The Silent State. Windmill Books. (2010).