

Ethics in Emotion-Oriented Systems: The Challenges for an Ethics Committee

Ian Sneddon, Peter Goldie, and Paolo Petta

Abstract The development of emotion-oriented systems has the potential to raise a range of ethical issues, not all of which were clearly understood at the inception of the project. This chapter discusses these issues and details the practical measures taken and the challenges faced in addressing them. An ethical audit revealed a lack of consistency in ethical procedures across the institutions and disciplines involved in the network, and HUMAINE established its own ethics committee to offer ethical advice and scrutiny when required. In addition, space was provided within the project for discussion of ethical issues – a process that allowed the emergence of a wider understanding of the issues themselves and of the sensitivities of different disciplines and users to them.

1 Introduction

It was a distinctive feature of HUMAINE that it established its own ethics committee. The intention of this chapter is first to explain why that was done; second, to examine some of the issues that should be addressed when establishing a similar ethics committee in the context of an interdisciplinary research programme; third, to chart the various initiatives and approaches to ethics adopted during the life of HUMAINE; and finally, to touch on some of the key conceptual issues that an ethics committee in the area has to deal with. There are two types of issues to address. One arises simply because some research on emotion-oriented computing involves human participants, and the ethical issues surrounding their participation have to be addressed, just as they do in any other field that uses human participants in experiments. The other is unique to emotion-oriented computing, and it arises because of the particular type of application that the area aims to develop.

To say that there are specific ethical issues is not to imply that the area is fraught with ethical problems. The HUMAINE project was at pains throughout to emphasise

I. Sneddon (✉)

Department of Psychology, Queen's University, Belfast, Northern Ireland, UK
e-mail: I.Sneddon@qub.ac.uk

that the aim of the research is benign. It is self-evident that many of our current interactions with machines are limited and often at best unsatisfying, at least in part due to the very limited communication abilities of current systems. There are of course human-machine interactions that would not benefit from the machine having any sensitivity to the emotional state of the human. However, just as with human to human communication, there are many situations where success in communication is improved by the ability of one party to interpret the emotional status of the other. It is easy to see that many machine systems involved in, for instance, education or entertainment would simply work better if they had a capacity to interpret changes in the emotion of the person and to respond in an appropriate emotional manner.

However, even with the best of intentions, there is a potential for unintentional harm or intentional abuse of emotion-oriented systems. There is also a problem of perception. Dystopian futures inhabited by machines with just such abilities are a common theme in science fiction and such imaginary worlds do have an impact on public ideas about this kind of research; and allaying public fears is a serious issue in itself.

2 The HUMAINE Ethics Committee: A Brief History

Increasingly, research programmes involving human participants have to make detailed provision for the treatment of ethical issues that might arise either during or as a result of their research. Five years ago, at the planning stages of the HUMAINE programme, it was not at all clear what provision the project would need in that area. At that time, we had a clear idea neither of the ethical climate existing in each of the partner countries nor of ethical practice and awareness in the various disciplines from which partners came. It seemed likely that their backgrounds would range from areas where the ethical issues were thoroughly familiar, to areas where they were never considered.

Against that background, the project adopted a proactive strategy. From the first plenary meeting of the network, researchers were alerted to some of the potential ethical problems that could arise from the development of emotion-oriented systems and asked to think about any possible ethical implications in their own research. Philosophers were included in the research programme to ensure that the response had a sound intellectual basis. Their input has proved invaluable throughout the lifetime of the project, particularly in dealing with the novel issues that the research raises.

To clarify the starting position, an ethical audit was conducted during the first year of the project (Goldie et al., 2004). This revealed that the majority of partners were conducting research that involved human participants in some way, but that they were operating under a very wide range of regulatory regimes. In the UK at the time, all research institutions were at various stages of implementing the policy that all research on human participants must be approved by a research ethics committee and the pace of demand for such change had quickened over the previous few years.

In partner institutions in other countries the picture was more varied. In some countries researchers did not yet have access to research ethics committees and in others, although access existed, submission of research for scrutiny was only required if the research would threaten the participant's health or psychological well-being – effectively introducing a self-selection element into the process. There was also variation across disciplines. Most social scientists were used to submitting their work to ethical scrutiny by others, but the concept was alien to most engineers and others from an IT background. It was the lack of consistency revealed by the ethical audit that suggested that HUMAINE itself might usefully set up a standing ethical committee.

There were mixed feelings about the proposal, and it was not at all clear to some members of the network what the function of such a committee might be. There was a strong view that a heavy-handed or over-regulatory approach would be counter-productive. Rather than encouraging researchers to reflect on the ethical dimension of their research, any attempt to impose regulation from above (without the legal basis for enforcement) would clearly have resulted in many researchers disengaging from any discussion of ethics altogether. There was also debate about the proper composition of a committee.

With this in mind, a group from diverse countries, disciplines and backgrounds was identified and approached to form an initial committee that held its first face-to-face meeting in Athens in June 2006. In the light of the concerns expressed above, the first matter for discussion was obviously the role of the committee. After much discussion, both between the members of the committee and between the committee and the other members of the network attending this 'cross-currents' meeting, it was decided that it could probably be of most benefit to HUMAINE researchers by taking on two main roles.

The first role was to answer questions on matters of research ethics and to offer advice on procedure and best practice for those involved in research on human participants. It was strongly hoped that the network would see the committee as a useful source of help and advice rather than as an unnecessary piece of bureaucracy. Informal requests for information or advice were, in the first instance, channelled through the chair of the committee. This role of the committee has been reasonably successful, with a number of requests for mainly practical advice on ethical matter being submitted to the committee.

The second possible role of the committee was to review any HUMAINE research involving human participants that had not already been scrutinised by an institutional research ethics committee. Although the committee had no legally constituted role for the formal approval or rejection of research proposals, it could nonetheless act as an advisory body and help researchers to ensure that their practice and procedures conformed to appropriate ethical guidelines.

Although the committee dealt with only a small number of full research submissions, a larger number of researchers asked for advice on specific aspects of a project. As with almost all administrative structures in large international projects, most business was conducted via e-mail or by telephone and after the committee had been established there was felt to be little need for face-to-face meetings.

It is revealing that the committee's last activity involved a request from another project looking for guidance that was not available in some of its participant institution. There does appear to be a useful function in the area for a properly constituted ethics committee that can cover gaps in local provision. It remains to be seen how the need will be met after the end of the HUMAINE project.

3 General Principles Underpinning an Ethics Committee

This section covers issues that would be faced by any project or group considering establishing an ethics committee. These range from highly abstract to extremely practical.

3.1 *The Intellectual Framework of the Committee*

Early debates about ethics in HUMAINE were constantly sidetracked into the maze of meta-ethical theory. Discussion visited and revisited questions about the possibility of justifying any ethical stance at all and about the relative merits of ideas derived (not always expertly) from Aristotle, Hobbes, Kant, the Utilitarians, Rawls and others. Fascinating as these debates are, they do not offer a basis for practical responses in a finite time.

HUMAINE addressed the difficulty by agreeing to base its practical analysis on the framework known as principlism (see Beauchamp and Childress, 2001, *Principlism: A Method for the Ethics of Emotion Oriented Machines* by McGuinness, this volume). Beauchamp and Childress reconstruct the core of morality in terms of a four-level model where each level is characterised by one of the following principles: *nonmaleficence*, *autonomy*, *beneficence* and *justice*.

Principlism was first designed for biomedical ethics, but it can easily be applied to other areas of ethical research, including research on the ethics of emotion-oriented systems. In fact, it has become one of the most influential approaches in the so-called *applied ethics*. At first glance it may seem as if 'application' means to apply a general ethical principle like the Utility Principle of Utilitarianism or Immanuel Kant's Categorical Imperative to a particular situation. Things are much more complicated, however. What familiar problem cases show is that there are moral intuitions that we all share, and against which we measure general ethical principles. By that measure, all of the familiar accounts can be found wanting. Against the Utility Principle, for instance, Bernard Williams offered the example of a botanist who wanders into a village in the jungle where 10 innocent people are about to be shot. He is told that nine of them will be spared, if only he will himself shoot the tenth (see Williams, 1973). In a similar way, many have argued against Kant's claim that we have a 'perfect duty' not to lie (or, a perfect duty to be truthful) that sometimes lying can be the only way to save an innocent life from death.

In contrast, principlism is a form of perception model (see Quante and Vieth, 2002). The principles that Beauchamp and Childress employ are not first justified through a specific ethical theory. They enable us to articulate what we perceive as being ethically sound or unsound in a particular situation. They are rules of thumb, or prima facie duties, which reflect the core stock of moral beliefs held in common in a modern pluralistic world: a framework for expressing the moral intuitions we all share.

Naturally, principlism has its critics (e.g. Clouser and Gert, 1990; Gert et al. 1997). Nevertheless, it provides a very valuable service by providing a rationale for closing down debates that are intriguing, but fruitless.

3.2 The Membership of a Committee

The guiding principle for membership is that a committee should reflect as wide a range of expertise as possible. That allows members to offer expert advice on as many research areas as possible and also makes the committee less likely to be unduly swayed by the interests of any one group. Perhaps less obvious, but common practice on medical ethics committees and increasingly on social science committees is the inclusion of a 'lay' person. On medical committees it is felt to be important that the lay representative is from outside the medical profession altogether. In the case of interdisciplinary projects such as HUMAINE, where members come from many diverse research cultures, there is less risk of developing a single 'establishment' view that needs to be challenged by someone from outside the community altogether. Nevertheless, it is important to have a lay representative who has no direct stake in the project; and preferable, even in intellectually diverse projects, one who is independent of the disciplines involved.

3.3 Choosing a Code

The most significant decision was to adopt a code derived from psychology rather than medicine. Medical codes are geared to situations where success may be life-preserving, unforeseen effects may be life-threatening, and there is an extreme disparity in power between clinicians and patients. As such, they tend to be more restrictive than codes directed to the less extreme situations that are the norm in psychology.

There are various psychological codes for research involving human participants. HUMAINE decided to adopt the guidelines of the British Psychological Society (BPS) as described in the 'Code of Ethics and Conduct' (2006), the 'Guidelines for Minimum Standards of Ethical Approval in Psychological Research' (2004) and the Code of Conduct, Ethical Principles and Guidelines (2000). There is a great deal of common ground between the codes of practice available in several countries and even in different disciplines. Other codes are mentioned later in this chapter.

3.4 The Value of Scrutiny

It is useful to sketch what, in practice, research ethics committees are likely to achieve. In an area where people are new to ethical scrutiny, they will clearly have an educational function. However, experience of reviewing research applications as part of a psychology research ethics committee suggests that research psychologists actually have rather a good grasp of most of the ethical issues in principles, and yet they repeatedly fall down in turning these abstract guidelines or codes of conduct into good research practice. It is as if having grasped the big picture, they assume that the details can look after themselves. Unfortunately it is the details that can make the difference between participants leaving an experiment with a positive view of science and scientists – or leaving upset, vowing never to return – and perhaps transmitting a negative image of the discipline.

It seems to be increasingly accepted that it is unwise to leave decisions about the conduct of experiments with human participants to the researcher alone. Researchers tend to be enthusiasts. Nobody advances as an academic researcher without some degree of passion for the research. That enthusiasm drives research forward; but on the other hand, it can make the researcher (with a clear focus on the research itself rather than on the participant) a poor judge of some of these issues.

For example, the principle of informed consent implies that the participant gives consent to take part on the basis of full knowledge of what they are likely to experience. But often the experimenter does not think of the situation from the viewpoint of the participant at all, leaving the participant to decide whether to continue based on very patchy and incomplete information. The ability to view the experimental procedure from the outside is one of the most valuable contributions that an ethics committee can make to the process of designing research that works.

3.5 Due Care and Indemnity

University ethics committees typically have two functions: to prevent undesirable things from happening and to provide a kind of indemnity if they do (so that an institution, or preferably its insurers, will bear the cost if anything does go wrong).

Unless there are radical changes, groups like the HUMAINE ethics committee will have no role in providing indemnity. Linked to that, approval from them will not satisfy institutions that require employees to go through their own ethical approval procedure. For that reason, some people have suggested that the committee was useless or worse than useless.

That ignores the function of preventing undesirable things from happening. What an ethical committee like HUMAINE's can do is to identify potentially unethical things that are entailed in an original proposal and to direct the research onto sounder tracks. In other words, that kind of committee is of use to people who do not want to do unethical things or who want to minimise the risk of doing things that would be unethical because they would do harm. It is of no use to people who simply want a fireproof wall if something does go wrong.

Even that may not be quite true. It remains to be tested who is legally responsible if a partner in a large project causes harm to participants. If the project has not taken due care, including appropriate ethical scrutiny, the whole group may be considered liable.

4 Specific Issues in Studying Human Emotion

HUMAINE aims to develop emotionally oriented technology, and so there is clearly a requirement for suitable examples of human emotion. Systems capable of interpreting human emotion or of sending emotionally appropriate signals must have examples from which to learn. Such examples have to be gathered, stored and labelled or annotated in a form that is comprehensible to the artificial system. For a number of theoretical reasons, an explicit decision was taken at the start of the project that the examples used would not be based on acted emotion but would, rather, be as natural as possible. This poses a number of practical difficulties for researchers, as natural episodes of emotion are neither predictable nor particularly common. The practical challenge of capturing, storing and annotating natural emotion also poses a number of particular ethical problems.

4.1 The Ethics of Inducing Emotion

The first approach to solving the problem of the rarity of natural emotional episodes might be to find some method of inducing the emotion. Of course the consequences of inducing emotion in a participant can vary from the trivial to the profound, depending on the nature and intensity of the emotional state generated. However, induction of emotion, particularly negative emotion, is generally not a process to be undertaken lightly. The BPS' revised Ethical Principles for Conducting Research with Human Participants (2000) indicate that on completion of an experiment that has involved induction of a negative emotional state, simply offering a descriptive information-based debriefing is not sufficient. The researcher should ensure that the participant is returned to at least a neutral emotional state before leaving. Thus, if a negative mood has been induced then it should be dispelled or a positive emotional state induced before the participant leaves. However, experience would suggest that this is a process that is easier to achieve with some emotions than with others. Some emotions such as anger seem to be relatively easy to dispel by using a short film or story to invoke amusement or pleasure. Other emotions can be much harder to dispel easily. For instance, the common emotion induction technique of asking someone to recall an emotional event from their past could, if focussing on sadness, easily unlock quite profound memories from their past that neither the participant nor the experimenter is prepared to deal with. Such eventualities may be rare, but when dealing with human participants with all of their frailties, there is a real responsibility on researchers to try to anticipate such possible problems and put safeguards in place

for their protection. It is the sharing of such infrequent but important experiences and insights among the researchers in HUMAINE that has been one of the most significant roles of the ethics strand within HUMAINE and of the ethics committee in particular.

4.2 The Ethics of Observing Emotion

If our aim is to collect examples of emotion that are as natural as possible, there is a second problem with inducing emotion: it may be genuine, but it is not natural. For that reason, some teams opt for observing people as they go about their daily business. In adopting this approach you might or might not choose to inform the participant in advance that they are being observed. The former method is open to the criticism that the participant's behaviour could be affected by the knowledge that they are being observed as part of a scientific study. The latter method, of course, raises the ethical issue of an individual's right to privacy.

Indeed, many codes of conduct recognise the particular difficulties of observational research, and the BPS principles state that researchers should normally only make recordings of participants with their full consent both to the recording itself and to any subsequent use(s) of the recordings. However, they do suggest that observational research may be acceptable in situations "where those observed would expect to be observed by strangers". It is suggested that in a public context it may be permissible to record their behaviour and to then ask for permission afterwards – a process that clearly involves 'incomplete disclosure'. However, in the introduction to their revised ethical principles (2000) the BPS make it clear that in the case of deception "the central principle was the reaction of participants when deception was revealed. If this led to discomfort, anger or objections from the participants then the deception was inappropriate."

Thus the guidance seems to be that if you are going to record people's public behaviour without their knowledge, they need to be involved as soon as possible and debriefed. If at that point you encounter negative reactions, then you need to rethink.

Realistically, someone who has been recorded like this should be debriefed as soon as is practical and should be informed about the research. The process naturally goes through many of the steps that would be involved in looking for informed consent in advance of a recording. Of course the participant's real safeguard here is that at this point the investigator must offer them the possibility of withdrawing and of having their data destroyed.

Best practice for a debriefing like this suggests that the researcher should try to involve the participant, to give a sense of having contributed something to the advancement of our knowledge about emotions. There is also an onus on the researcher to find out how the participant feels about the study and their place in it. It would be only too easy to ignore the fact that the exercise has made someone feel uncomfortable or upset in these circumstances – so rather than wait for the participant to express negative feelings, the researcher should ask.

4.3 Alternatives to Induction and Covert Observation

There is an increasingly vocal body of opinion that deception or incomplete disclosure should be avoided if any viable alternatives exist, and it does seem that there is a wide range of situations where researchers could use naturalistic observation after getting informed consent. People regularly take part in events or experiences where both the researcher and the participant know in advance that the participant is likely to experience and express a range of emotions. There is a wide range of situations such as weddings, sporting events (both participating and spectating), parachute jumps, public speaking, emerging after important exams and so on where observation would be possible.

In all of these cases, the participant could be asked in advance to consent to being filmed. It is not yet known whether any of these situations could provide sufficient diversion for the participant to become unaware of the observation and to behave in a natural manner but this is an empirical question and such possibilities should certainly be explored.

4.4 Privacy, Confidentiality and Anonymity

Even once the data are collected, there are further ethical issues raised by the particular nature of the recordings of emotion and the fact that these must be stored and then used in the development of emotionally sensitive systems. The particular problems here are privacy and confidentiality. For most psychological research these should not be major concerns. Usually the individual is included in the research as a representative of a larger group – either as a member of an experimental treatment group or as a representative of an age or gender group or social class and so on. The point is that the data are usually aggregated and there is no need or desire on the part of the researcher to ever reveal the performance of an individual. Even studies of patients with particular psychological problems are usually interesting because of the problem and not the individual, and their identity can be kept confidential with no difficulty. However, in the case of HUMAINE research that simply is not true – the point of the observing and recording emotional behaviour is to allow us to identify and label occurrences of emotion. In most cases the recordings will of course reveal the face or the voice and allow the individual (at least potentially) to be identified. In addition, the BPS ethical principles (2000) state that “in the event that confidentiality/anonymity cannot be guaranteed, the participant **MUST** be warned of this in advance of agreeing to participate”. This brings us to probably the most important of the ethical principles that we will need to consider when we begin collecting and handling emotional data – the principle of informed consent.

4.5 Informed Consent

Informed consent is seen as an essential tool for safeguarding the rights of the individual. Most modern codes of conduct place great emphasis on the need for

psychologists to avoid deception of participants whenever possible and the Canadian Code of Ethics for Psychologists (Canadian Psychological Association, 2000) explicitly equates incomplete disclosure with deception – emphasising the need to get fully informed consent before starting. The Canadian code also makes the point that we should approach this exercise as more than just getting the form filled in – we should see informed consent as “the result of a process of reaching an agreement to work collaboratively”. Unless there are very good reasons for not doing so, obtaining fully informed consent in writing is seen as an essential prerequisite to most human research.

A typical consent form takes the participant through the essential elements that need to be addressed before asking someone to take part in a study. First they agree that they consent to take part in the study and that they are not being coerced or induced to consent. There is then usually some mention of the consent being genuinely informed. A common way to achieve this is to give participants a short written summary of what the participant should expect and to reassure them that no physical or psychological harm will occur to them. This summary might also briefly describe the aims and nature of the research but, in some cases where this knowledge is thought likely to influence the behaviour of the participant, that information might be withheld until the debriefing at the end of the experiment. There is usually an item that draws the participant’s attention to the principle that they can withdraw from the research at any time with no adverse consequences – in effect the participant does not just have to consent at the beginning of the study, they must continue to consent throughout. They can withdraw their consent at any point and this also means that they can withdraw their data at any time. Finally on a typical consent form the issue of confidentiality is commonly addressed. If it will not be possible to guarantee the anonymity or confidentiality of the data then this would be an ideal opportunity to make the participant aware of that fact in advance of them agreeing to take part in the study.

4.6 Storage of Emotion Data

If it is really not possible to guarantee the anonymity or confidentiality of the data then there is an added responsibility to anticipate the uses to which the data are likely to be put. Will other research teams have access to it? Might it be used in public dissemination of research results? There are some notable picture sets of facial expression of emotion that have been very widely used indeed and it is not difficult to imagine that some participants might find that sort of exposure disturbing. We have found that while most participants have no problem consenting to a wide range of possible uses for such data a substantial minority are uncomfortable, particularly with the idea of possible public dissemination of their image or voice. Even if participants are willing to have their faces or voices appear in a variety of circumstances, it is conceivable they may have an objection to being portrayed in a negative way (for instance through the researcher labelling them or their behaviour as angry or ashamed or fearful).

A final issue about the storage of data raised in the Canadian Code of Ethics for Psychologists (2000) is about the security of data. During the process of giving consent to participate in the study and for any uses of the data specified, it is assumed that researchers will safeguard that data. However, the Canadian code asks if the researcher's responsibility ends when anything happens to them (e.g. if they fall ill or lose their job or, in extreme circumstances, if they die). The implication in the Canadian code is that researchers have a responsibility to anticipate such events and should make provision for these eventualities – ensuring, for instance, that there is another individual who can take over responsibility for the data or have it destroyed.

4.7 Who Are the Examples?

When research is compiling databases of emotional behaviour to be used as examples, it begs the question 'Who are to be used as examples?' HUMAINE was sensitive to this question from the beginning and concerned that examples of emotional behaviour should represent people of different age, gender, culture and so on. A broad collection is crucial for many reasons, not least that the developed systems will work only poorly if provided with a narrow range of emotional examples. The inclusion of examples of people of different ages implies the inclusion of children and, of course, the emotional development of children is surely one of the most theoretically interesting areas of study. However, in the UK at least, there has been a marked increase in sensitivity about anyone having contact in a professional capacity with children. Anyone now wishing to conduct research with children must go through a police vetting procedure to make sure they do not have any previous offences against children. In terms of ethical procedures, it is obviously necessary to get the informed consent of the parents as well as the consent of the child, and researchers would need to be much more sensitive to the dangers posed by inducing emotions.

Where researchers again stray beyond the usual experience of collecting data from child participants is in the recording and storage of data by which the child can be identified. In naturalistic studies it is difficult to imagine a research ethics committee granting permission for a study where the researcher planned to film children without their knowledge and only then to ask for consent from the child and the parents. Again, as with natural observation of adults, researchers may have to be much more creative in anticipating suitable emotional situations and gaining prior consent.

5 Predicting Future Ethical Issues

Of course the underlying rationale for the ethical scrutiny of any research is to protect people from any harm that might arise from the research. In most of the cases we have been considering, the focus of such scrutiny is on the research environment

itself and on the protection of the participant while the research is being conducted. However, in the case of research on emotion-oriented systems, it was recognised early on in the HUMAINE project that the research might generate applications or products that could potentially have ethical implications far beyond the research environment itself.

Throughout the life of HUMAINE the researchers have tried to predict the ethical issues related to their research and have openly discussed the potential pitfalls and possible abuses of any projected applications (see also “The Ethical Distinctiveness of Emotion-Oriented Technology: Four Long-Term Issues” by Goldie et al., this volume). The issues are relevant in principle to an ethical committee: it would not be ethical to approve a project designed to produce technology that would clearly be damaging to substantial numbers of people. That issue has not arisen in HUMAINE because there is a benign motivation behind all the technology, as the introduction to this chapter pointed out. However, in an area of newly developing technology like this it can be difficult to predict how people will actually use any systems that are developed.

It is not surprising, therefore, that some of the discussions that have taken place over the last 5 years have spilled over into ‘science fiction’ rather than science. Such speculation is a key part of trying to foresee future developments of technology and its uses. It is increasingly expected that when research involves technological innovation that may have ethical implications, the scientists involved in its development take some responsibility for trying to foresee what the problems might be.

On the other hand, ethical committees should certainly not block research because of speculation that belongs in the realm of science fiction. The sections that follow try to give a sense of what the real issues are in core areas of activity.

5.1 Systems That Can Recognise Emotion

Several types of issues arise in the context of systems that are designed to be able to recognise some features of human emotion. Some of the most pressing hinge on overestimating what they can do (or will soon be able to). There has been a common historical tendency for us to attribute ‘objectivity’ to machine systems such as the polygraph that can recognise some aspect of human behaviour. Of course the arguments about the abilities and limitations of the polygraph have dragged on for over half a century and it is crucial that, for any new technology in this area, we recognise and spell out the limitations as clearly as possible. In fact the current ability of emotion-oriented systems to accurately and reliably read human emotion falls very far short of the human ability to do the same thing. This seems unlikely to change in the near/medium term and it is important that researchers ensure that their systems are accurately portrayed. Overselling the abilities of such systems would be ethically wrong and could lead to quite inappropriate decisions being taken and harm done.

A related issue is the extent to which people will regard the monitoring of their emotions as an intrusion into their private lives. In an extension of a point made

above in relation to the observational study of human emotion, it is not clear if the use of machines to monitor people in public places without their knowledge should be regarded as a similar invasion of privacy. In addition, the use of technology such as semi-intelligent information filters (SIIFs) means that such monitoring is unlikely to remain purely descriptive, but will allow judgements to be made about emotional changes. This of course begs the question about the uses to which such technology might be put to. The monitoring of call centre employees might allow workers to be pulled off the job as their anger levels rise. This is a rather straightforward example of a much wider reaching ethical problem. After monitoring a large sample of workers we may be able to make some sort of probabilistic statement about the relationship between the emotion (e.g. anger) and problem behaviour (e.g. rudeness). However, the point is that the emotion does not *inevitably* lead to a change in behaviour. A worker may become very angry indeed with a caller, but manage to keep her feelings under control and not let the anger alter her behaviour towards the caller. So, even if it does become possible to identify emotional states reliably, then it is important that the information is not used unreasonably.

In the first report (D10a) on ethics for HUMAINE (Goldie et al., 2004) the point was also made that the very existence of machines that can monitor or probe inner mental states in ways that may be poorly understood could in itself be frightening to many people. The research community has a responsibility to reduce the risk that they will be used to create fear.

5.2 Systems That Can Manipulate Emotion

Next we come to systems that are designed to manipulate human emotions. Indeed, one area that has been the focus of an entire work-package within HUMAINE is persuasion. Of course this area raises many ethical issues – some of which were raised at the start of the project. At the first plenary meeting of the network members, in a talk intended to introduce many of the ethical issues that might lie ahead, the network coordinator warned of a ‘nightmare possibility’ – a system that could read user emotions and selectively reinforce or attenuate them without conscience or empathy.

However, although it is possible to conceive of such an extreme system in theory, most of the problems associated with persuasive systems seem to be more mundane. In fact the existing legal controls and guidelines applicable in most western countries already cover the most likely abuses. The replacement of a person with a machine in the role of the persuader does not alter the underlying ethical issues.

5.3 Systems That Can Express Emotion

Finally we come to systems that are capable of sending emotional signals (ultimately with the aim of simulating human emotion). Although some of the discussions regarding ethical issues in this area are definitely ‘crystal-ball gazing’ the ethical

problems posed by machines that can simulate human emotion and one day might be mistaken for human should probably begin to be discussed now.

An issue that was raised at the first HUMAINE plenary meeting was that when people's expectations of a machine include emotional responsiveness, then you immediately enter an ethically sensitive situation – if that expectation exists then it becomes possible for the machine to respond inappropriately (either by failing to respond, or by responding with the wrong emotion) – for instance if someone is experiencing and expressing sadness or grief and receives a flat, unemotional response or even worse, a response of laughter, the result could be distressing.

Of course, the way that *we* respond to a machine that is expressing emotion also raises tricky questions. While it is true that the machine is only simulating rather than experiencing emotion and cannot 'feel' emotional distress if we react inappropriately, it has been argued that such behaviour may, in the long run, be harmful to us – by getting us into the bad habit of treating emotional beings as mere things.

These questions, related to the possible uses that people might eventually find for emotion-sensitive technology and how they might interact with such systems, leave us gazing into an uncertain future. Predictions about how the world might be influenced by developing technologies are notoriously unreliable. Such speculations definitely go beyond the activity of a traditional research ethics committee.

6 Conclusion

HUMAINE followed a dual-track approach: establishing an ethics committee to deal with immediate issues raised by ongoing research on one side and, on the other, allowing sufficient time and resources for discussion of the ethical implications of the systems that might develop out of the current research. The approach seems to have been very productive.

Over the lifetime of the project, attitudes towards the process of ethical scrutiny have changed considerably as the network members have become more familiar with the issues involved. The point was made at the beginning of this chapter that the aims of HUMAINE have been benign, but it is undoubtedly true that research on emotion-oriented systems has the potential for controversy. The community is far more likely to maintain public support for this research and the technology that emerges from it, if it is open about discussing the ethical issues and it communicates these discussions as widely as possible.

References

- Beauchamp TL, Childress JF (2001) Principles of biomedical ethics, 5th edn. Oxford University Press, Oxford
- Canadian Psychological Association (2000) Canadian code of ethics for psychologists, 3rd edn. Canadian Psychological Association, Ottawa, ON

- Clouser KD, Gert B (1990) A critique of principlism. *J Med Philos* 15:219–236
- Gert B et al (1997) *Bioethics: a return to fundamentals*. Oxford University Press, New York, NY
- Goldie P et al (2004) Report on ethical prospects and pitfalls for emotion-oriented systems. HUMAINE report D10a. <http://emotion-research.net/projects/humaine/deliverables/D10a.pdf>. Last visited 17 May 2010
- Quante M, Vieth A (2002) Defending principlism well understood. *J Med Philos* 27(6):621–649
- The British Psychological Society (2000) Code of conduct, ethical principles and guidelines. The British Psychological Society, Leicester, UK
- The British Psychological Society (2004) Guidelines for minimum standards of ethical approval in psychological research. The British Psychological Society, Leicester, UK
- The British Psychological Society (2006) Code of ethics and conduct. The British Psychological Society, Leicester, UK
- Williams B (1973) A critique of utilitarianism. In: Smart JJC, Williams B (eds) *Utilitarianism for and against*. Cambridge University Press, Cambridge, MA, p 98

Additional References (on Principlism)

- Bryson J, Kime P (2003) Just another artifact: ethics and the empirical experience of AI. University of Edinburgh, Edinburgh
- Dancy J (1993) *Moral reasons*. Blackwell, Oxford
- Döring SA (2004) Gründe und Gefühle. Rationale Motivation durch emotionale Vernunft. Habilitationsschrift, Universität Duisburg-Essen. (in German)
- Goldie P, Döring S (2004) Categories of emotion: everyday psychology and scientific psychology, with Peter Goldie. In: Proceedings of the first HUMAINE workshop, Geneva 17–19 Jun. <http://emotion-research.net/ws/wp3/ExtraMaterial/HUMAINE-Goldie.pdf>. Last visited 17 May 2010
- Goldie P, Döring S (2005) Emotions as evaluations. In: Proceedings of the AISB'05 symposium on *Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action*, Hertfordshire University, 12–13 April 2005, SSAIB, pp 45–50
- Goldie P, Döring SA, WP10 members (2005) D10b Interim report to plenary meeting on ethical frameworks for emotion-oriented system. <http://emotion-research.net/project/humaine/deliverables/D10b.pdf>. Accessed 17 May 2010
- McDowell J (1998) *Virtue and reason*. Reprinted in his *Mind, value, and reality*. Harvard University Press, Cambridge, MA, pp 131–166
- Quante M, Vieth A (2001) Wahrnehmung oder Rechtfertigung? Zum Verhältnis inferenzieller und nicht-inferenzieller Erkenntnis in der partikularistischen Ethik. *Jahrbuch für Wissenschaft und Ethik* 6:203–234. (in German)
- Wiggins D (1998) *Deliberation and practical reason*. Reprinted in his *Needs, values, truth*, 3rd edn. Blackwell, Oxford, pp 215–237