Ethics and Information Technology (2006) 8:205–213
DOI 10.1007/s10676-006-9112-4

Responsible computers? A case for ascribing quasi-responsibility to computers independent of personhood or agency

Bernd Carsten Stahl
Faculty of Computer Science and Engineering, Centre for Computing and Social Responsibility, De Montfort University, The Gateway, Leicester, LE1 9BH, UK
E-mail: bstahl@dmu.ac.uk

Abstract. There has been much debate whether computers can be responsible. This question is usually discussed in terms of personhood and personal characteristics, which a computer may or may not possess. If a computer fulfils the conditions required for agency or personhood, then it can be responsible; otherwise not. This paper suggests a different approach. An analysis of the concept of responsibility shows that it is a social construct of ascription which is only viable in certain social contexts and which serves particular social aims. If this is the main aspect of responsibility then the question whether computers can be responsible no longer hinges on the difficult problem of agency but on the possibly simpler question whether responsibility ascriptions to computers can fulfil social goals. The suggested solution to the question whether computers can be subjects of responsibility is the introduction of a new concept, called ''quasi-responsibility'' which will emphasise the social aim of responsibility ascription and which can be applied to computers.

Key words: agency ascription, computer, personality, quasi-responsibility, responsibility

Introduction

Computers can do morally good things, such as help autistic children develop their potential and become responsible members of society (Anonymous 2004a). They may revolutionise education and help us perceive reality in a more complete way (Anonymous 2004b). At the same time it is conceivable that they can be used for immoral purposes. One can therefore say sentences such as: ''This technology is good/bad'' and these sentences will be comprehensible to most potential listeners. The examples show that we find it fundamentally possible to perceive non-human entities in moral terms. Some academic research is interested in the question whether technology can have human properties. Picard (1997), for example, investigates what it means for computers to develop emotions. Other research (Brooks 2002) suggests that interaction with robots can acquire properties traditionally reserved for interaction with humans. Does this mean that we will start treating computers as responsible entities?

This raises the question whether artefacts can be held responsible. When a human being kills another human being using a hammer then we will usually hold the human responsible, and not the hammer. When a human being shoots another human being, then there are people who believe that the gun is at least partly to blame, possibly because guns make it easier to kill people than hammers do, or possibly because the social structures that allows the production of guns hold a share of the responsibility. This sort of reasoning can be the justification for suing weapons manufacturers after a shooting spree (cf. Kairy 2003). When a human being uses a highly complex piece of technology to kill another then the individual's responsibility can seem relatively minor in terms of overall responsibility. An example here might be the use of an industrial robot which accidentally kills a human being. In this case it is not clear how responsibility should be distributed among humans. Does this mean that we should ascribe responsibility to the robot? Such questions are of high economic and moral importance with regards to computer hardware and software which are highly complex and do not allow for linear ascriptions of causality and responsibility.

Many philosophers argue that being moral or immoral are purely human characteristics (Lewis 1991). Following this sort of reasoning would lead to the exclusion by definition of computers from the realm of morality. Given the (moral) importance of computers in modern societies, this paper aims to approach the question of computers' responsibility from a different direction. The paper will start out with an analysis of the concept of responsibility and it will

stress the nature of responsibility as a social construct of ascription, aimed at achieving certain social goals. Based on this notion of responsibility it will then discuss why computers or technical systems in general might be considered subjects of responsibility. It will suggest the concept of "quasi-responsibility" that would be applicable to computers and information systems. The paper will discuss the advantages of this approach and consider the relationship between computers' quasi-responsibility and humans' traditional responsibility.

Responsibility

The question whether computers should be held responsible is usually answered quickly with a "yes" or a "no". Most people are intuitively clear about their attitude towards the moral status of computers. Unfortunately, however, the individual clarity of the answer does not lead to collective unanimity on the subject. The reason for that is the fuzziness of the notion of responsibility and the multitude of different meanings, conditions, and results it can have. To give a correct account of computers or information systems as responsible subjects we will therefore have to start with an analysis of the notion of responsibility.

A first definition of responsibility

The term "responsibility" is not easily defined. A first hint concerning the meaning of responsibility can be found in its etymology, in the "response". Responsibility has something to do with answering. Responsibility stands for duty to answer to somebody for something.[1] This carries some implications. If responsibility has to do with answering then this means that it is a social construct. Answering implies that there is something (we call it the subject) that has the ability to answer. The subject must fulfil some conditions. The first one is that it is similar to me (the person demanding an answer) in at least the respect that it / he / she can understand me. This similarity between the "other" and me is the foundation of ethics, especially in French philosophy of the 20th century (cf. Ricoeur 1990). Responsibility relations in real life tend to be broader than a dialogue between two individuals but their root is nevertheless the ability to answer. The nature of giving the answer, of being responsible, of ascribing responsibility includes the establishment of a link between the subject and an object. This link typically leads to the attribution of sanctions.

There are different types of responsibility all of which are based on answers. They differ in the sort, settings and surroundings of how the answer is given and by whom. One classical type of responsibility is role responsibility. In this case social expectations are condensed to a role and the person holding the role has to answer accordingly. Roles are often linked to professions and professional roles determine actions and communication. An engineer, for example, has certain responsibilities that are defined by his job. Different types of responsibility can be co-located in a single person, where they can come into conflict. One can be a mother, a politician and an engineer at the same time.

For the purpose of discussing a computer's responsibility it will suffice, however, to concentrate on the two most important forms of responsibility, on moral and on legal responsibility. Legal responsibility, as opposed to all other sorts, has the advantage of being clearly defined in theory and having verifiable results in practice. This does not mean that material responsibility ascriptions are a priori clear. Rather, it means that the structures and procedures that will lead to the ascription of responsibility are clearly defined and that there are established ways of clarifying questions regarding these structures. The legal sphere is also the root of today's ubiquitous use of the term responsibility. Whoever is legally responsible first of all has an obligation to answer (Trigeaud 1999). Legal responsibility can in most countries be divided in responsibility according to criminal or civil law which lead to different consequences and sanctions. While the term responsibility is most easily understandable in its legal use (Ricoeur 1995) there is a relationship to moral responsibility. There is a close link between criminal law and ethics and thus between the two types of responsibility (Neuberg 1997). Murder, rape, and fraud are immoral and thus

[1] For a more detailed discussion of the concept of responsibility that would exceed the limits of this paper see: H. Jonas *Das Prinzip Verantwortung*. Suhrkamp, Frankfurt a. M., 1984; K. Bayertz. Eine kurze Geschichte der Herkunft der Verantwortung. In K. Bayertz, editor, *Verantwortung: Prinzip oder Problem*? pp. 3–71, Darmstadt: Wissenschaftliche Buchgesellschaft, 1995; J.M. Fischer. Recent Work on Moral Responsibility. *Ethics* (110:1) 93–139, 1999; P.A. French. *Responsibility Matters*. Lawrence, Kansas: University Press of Kansas, 1992; L. May and S. Hoffman, editors, *Collective Responsibility: Fife Decades of Debate in Theoretical and Applied Ethics*. Rowman & Littlefield Publishers Inc., Savage, Maryland, 1991; M. Neuberg. Introduction à "La responsabilité". In M. Neuberg, editor, *La responsabilité — questions philosophiques*. pp. 1–24, Presses Universitaires de France, Paris, 1997; E.F. Paul, F.D. Miller and J. Paul, editors, *Responsibility*. Cambridge University Press, Cambridge et al, 1999; R.J. Wallace. *Responsibility and the Moral Sentiment*. Harvard University Press, Cambridge, Massachusetts/London, England, 1996. For an in-depth analysis of the relationship of ethics, morality, and responsibility as well as for a discussion of the application of responsibility to information systems cf. B.C. Stahl. *Responsible Management of Information Systems*. Idea Group Publishing, Hershey, 2004a.

they are also illegal. This is not the place to discuss the relationship of law and morality but it should be clear that the two are related. When we talk about computers as subjects of responsibility then this implies a moral as well as a legal background and therefore moral as well as legal consequences.

There are many more aspects of responsibility that should be discussed for a complete picture of the subject. Responsibility can be external and internal. It can have descriptive or normative aims. It can have differing temporal directions; that means we can be responsible for things to come or for events of the past. We should keep in mind that responsibility is an ascription. It is a relational notion involving at the very least an object and a subject. The subject is the "who?" of responsibility, the object the "for what?" In the sentence "the computer is responsible for the data loss", the computer is the subject and the data loss represents the object. Usually there are other dimensions involved in the social process of responsibility ascription. We can often hear of an instance, an authority. In the case of legal responsibility this is the judge or the jury who will decide about the ascription and its consequences. Furthermore we need some sort of normative background, the law in legal and morality in moral responsibility. Finally we need somebody, some group, or some process that initiates and executes the process of ascription. Again, all of these dimensions are complex and cannot be discussed here in greater length. (For a more comprehensive discussion of these issues, see Stahl 2004a). This paper will concentrate on the subject of responsibility in order to clarify whether computers can fulfil this role. In order to answer this, we will first have to discuss why we ascribe responsibility at all.

Objectives of responsibility

The main purpose of this paper is to explore the question whether computers can be viewed as subjects of responsibility. The answer to this question typically hinges on considerations of personhood or agency, as will be explored below. This paper puts forward a different argument based on the social function of responsibility. In order to do so, we need to understand why responsibility is ascribed and what functions it fulfils in society. Briefly, the purpose of responsibility is to effect a socially desirable state. This is usually achieved through the imputation of sanctions, be they positive (rewards) or negative (punishment). Sanctions are an integral part of responsibility and in most cases the sanctions take the form of punishment. The concept of responsibility results from the need or want to find the guilty party in the face of negative results (Bayertz 1995, p. 22) and punish them. At the same time, the threat of punishment has the purpose of motivating people to act responsibly (De George 1999, p. 118).

The sanctions depend on the type of responsibility. While legal responsibility is institutionalised and attributed according to legal schemes, moral responsibility is not institutionalised, often internal and in many cases consists of blame (cf. Collste 2000a, p. 126; Hausman and McPherson 1996, p. 223). Why do we punish the subject of responsibility? Hart suggest a number of reasons for punishment. Men punish "to secure obedience to laws, to gratify feelings of revenge, to satisfy a public demand for severe reprisals for outrageous crimes, because they believed a deity demands punishment, to match with suffering the moral evil inherent in the perpetration of a crime, or simply out of respect for tradition" (Hart 1968, p. 73).

Fauconnet (1928) argues that every punishment has to fulfil a moral purpose. This purpose is usually the improvement of social circumstances. How can society improve its workings through the use of punishment? By keeping people from doing what is considered bad and by giving them reasons to act in a morally good way. The prime moral value of punishment is deterrence, a thought widely shared by many philosophers from different schools (cf. Schlick 1930; Bayertz 1995). Deterrence seems to be the prevalent justification of punishment. Restitution, which is a sanction in civil law, can serve a similar purpose of deterring individuals from committing certain acts (cf. Long 1999).

A last remark on deterrence as objective of responsibility: While it is certainly plausible to most of us that the knowledge of punishment can keep a person from doing something, this idea contains several assumptions concerning the subject. Those are often discussed under the heading of "economy of threats". The subject must be able to calculate the consequences of his actions and if the expected value is negative he will refrain from doing the deed (cf. Wallace 1996). This presupposes a high degree of rationality, certainty, and knowledge. It also makes assumptions about humans' ability to act rationally, to understand the world, and even more fundamentally, about freedom of mind and action, which are deeply contested among philosophers.

A functionalist view of responsibility that will be used for the argument of this paper can concentrate on the purpose of responsibility ascriptions of achieving specific desired outcomes. The conditions of responsibility which will be discussed in the following section can then be interpreted as mere characteristics that are necessary for the function of the ascription to come to pass.

Conditions of responsibility

In order for responsibility ascriptions to fulfil their social goals, the subject is usually deemed to need to fulfil a number of conditions. Most of these involve deeply contentious philosophical ideas or assumptions. This paper will not be able to address all of these. Indeed, as will become clear during the introduction of the concept of "quasi-responsibility", it is the purpose of the paper to move beyond the philosophical pitfalls surrounding the conditions of responsibility. In order to appreciate this turn of the argument, we will nevertheless have to reviews these issues briefly.

A first condition of responsibility is causality (May 1992; Nissenbaum 1995; Moore 1999; Lipinski et al. 2002; Scheines 2002). There must be some kind of causal chain that links the subject to the object. The reason why we think that causality is important for responsibility is that we need causality to have some kind of power over the outcome, which is the next condition of responsibility. What we cannot change, what we have no control over and cannot influence cannot be an object of responsibility (Birnbacher 1995). It may not be necessary to have complete control over the outcome but, in order to be subject of responsibility, one must at least be able to avoid the outcome or parts of it (cf. Bayertz 1995). Power, in turn, is based on knowledge. The subject must know what is happening in order to influence it. Furthermore, in order for the claim of power to make sense, the subject must be free to act on his or her knowledge. Freedom is therefore another vitally important precondition of responsibility (Johnson and Powers 2005). Being able to act voluntarily is the first one of several mental properties that subjects of responsibility are supposed to have. It is similar if not identical with the legal term of "*mens rea*" (the guilty mind). This means that the subject must act intentionally (Collste 2000b).

Maybe the most difficult and controversial conditions a subject of responsibility has to fulfil concern its inner state or mentality. Some authors think that subjects need emotions (cf. Sherman 1999; Wallace 1996). One reason for the necessity of emotions is that the subject must be concerned by what s/he does, there must be a personal connection to the object (Bierhoff 1995). Maybe more important, in the light of the economy of threats, the subject needs emotions in order to translate the abstract moral calculus into concrete behaviour.

Computers and responsibility

These conditions are aimed at the traditional subject of responsibility, the adult rational human being, which explains why the "whole conceptual vocabulary of 'responsibility' and its cognate terms is completely soaked with anthropocentrism" (Floridi and Sanders 2004, p. 366). But even with regards to humans they are deeply problematic. Apart from difficult philosophical questions such as "are humans free?" or "can humans act according to their will?" even the apparently simpler conditions are rarely met. If we look at the big problems of our time such as environmental problems, e.g. global warming, depletion of the ozone layer, societal problems e.g. globalisation, poverty on national and global scales, or just the problem of organising one's life in the face of growing complexity, single humans quickly reach their limits. In many cases we may be causally responsible but those causalities are rarely known. Even in those rare cases where we have knowledge of causality, e.g. the use of cars and global warming, we are individually powerless to change things. The mental conditions for being a subject of responsibility are rarely fulfilled. We may be emotionally concerned but often not about the things that we can change. When we execute problematic acts we may not do them intentionally. Even if we act intentionally, the economy of threats may not work because we are not aware of consequences and sanctions. These and other problems have led to the idea that more than just humans should be considered subjects of responsibility. One type of entity to which this question can be applied is the computer.

Why the computer cannot become a subject of responsibility

After the above discussion of the concept of responsibility the immediate reaction to the idea of computers as subjects is negative. The main arguments against the admission of computers as subjects of responsibility are to be found in their lack of fulfilment of the conditions of responsibility as described in section "Responsibility". First of all, computers do not fulfil the individual conditions. They lack consciousness in the human sense and therefore cannot be said to have intentions. The *mens rea* criterion does not apply to computers. The lack of *mens rea* coincides with the lack of a conscience and any sort of emotions that make humans susceptible to acting responsibly. While computers are better than humans at doing calculations, the economy of threats cannot work in their case, at least not in the way we envisage it for humans. Computers cannot be punished because they lack fear. Information systems lack another vital component of responsibility; they are not free. Computers are clearly determined by hardware and software and even though we may not

understand them any more and therefore may not be able to predict what they will do, they have neither freedom of will nor of action. Another weakness is that they are unable to really understand human beings (cf. Stahl 2004b) and are not able to originate an act worthy of responsibility because of their lack of a human body, which can be viewed as a vital component of responsibility (Velasquez 1991).

Computers, one can summarise, have none of the characteristics of persons, cannot be persons and can therefore not be subject of responsibility, especially not of moral responsibility. Much of the current literature regarding the responsibility of computers (or related entities such as artificial agents, software bots etc.) aims to create more clarity in this conceptual jungle. Authors try to shed light on which conditions need to be fulfilled in order to be a subject of responsibility. They analyse the components of agency or personhood in order to identify which ones of those may apply to computers or under which conditions computers may count as agents or moral subjects. Floridi and Sanders (2004), for example, suggest that computers can be seen as artificial agents, which leads them to suggest that material moral norms such as codes of ethics can be implemented for them. Johnson and Powers (2005) and Johnson (this issue), on the other hand, explicitly deny that computers can be agents but nevertheless argue that they are moral enteritis. Allen et al. (2000) and Allen et al. (2006) believe that computers can actually be moral agents and discuss how such morality could be technically implemented.

There seems to be little agreement between different authors. The reason may be that we are touching on deep philosophical issues which by no means are solved for human beings. This may make it even harder to pose similar questions for technical artefacts. And while these debates are intellectually stimulating and academically legitimate, this paper will not engage with them but suggest a different route to address the issue of computer responsibility.

Why the computer should become a subject of quasi-responsibility

The traditional philosophical view is that non-humans cannot be subjects of responsibility. However, new technical artefacts, particularly computers and their derivatives, display properties that make us doubt this conclusion, even though this may be "ethically troublesome" (Bloomfield and Vurdubakis 2003, p. 27). They are adaptable, able to learn, autonomous, and possibly even intelligent (cf. Korienek and Uzgalis 2002). Because of such properties, computers can usefully be described as artificial agents (Floridi and Sanders 2004) and maybe even as artificial moral agents (Allen 2002; Allen et al. 2000). This paper refrains from engaging in this debate but takes from it the intuition that responsibility ascriptions to computers can be viable. There are thus good reasons to question the restriction of the concept of responsibility to humans.

From a functionalist perspective, furthermore, one could argue that what matters with regards to responsibility ascription is not whether the subject fulfils the conditions but rather whether the ascription leads to the intended consequences. Ascribing responsibility to computers would then be desirable because of the good that comes from it and should thus be considered. However, most philosophers would tell us that arriving at the conclusion that computers can be responsible because they should be is a categorical mistake. It is what has been termed a normativist fallacy, the impossible inference from ought to is. We therefore have to find better reasons.

The first reason is that computers do fulfil one central condition of being subjects, namely that they make decisions. We are confronted all of the time with decisions that are made by machines, especially computers. The red light that tells us to stop, the ATM that decides to give us money or not and the defence system that decides to shoot down the airplane make factual decisions. These are highly contentious statements. Do machines actually make decisions? Is this an example of inadmissible anthropomorphism? Is the decision not made by humans somewhere at the end of a causal chain? These are very difficult questions and, thought through to the end, they will again point to questions concerning human beings. Do we make decisions? Are we free? The suggested solution here is to ignore these questions and return to the functionalist view of responsibility and related facts. From this viewpoint one could argue that computers do make decisions that, in terms of their social consequences, are comparable to human decision. Whether a red light tells me to stop or a policeman does so can (possibly should) have the same effect. This says nothing about the inner states of either, just about associated behaviour. The argument is based on a certain Level of Abstraction (LoA) (cf. Floridi and Sanders 2004). This means that it does not claim that computers do take decisions in an objectivist understanding of the world. Rather, there is a plane of description on which one can usefully speak of computers taking decisions. Given the constructivist nature of responsibility ascription, the determination of such a Level of Abstraction is completely sufficient. Bechtel (1985) argues that information systems have something like an internal decision structure which can be viewed as a supporting argument of this line of thought.

Another argument for the computer's responsibility is that it may perform its social function. Why do we consider the conditions of responsibility as relevant? Because they guarantee that the desired results can be realised. If the social construct of responsibility is interpreted in light of its social functionality then what counts is less the abstract definition of admissible entities and more the social consequences it produces. The subject then has to be rational and endowed with emotions in order to be able to react adequately to the threat of ascription. Rationality, for example, then is no end in itself but a condition for deterrence to work and thus for people to act according to the social aims that motivate responsibility ascriptions. This argument for responsibility is purely functionalist and thus differs from our typical understanding of the term. It considers no other criteria for an ascription other than its social usefulness. It is thus also formal because it cannot determine what "social usefulness" means. This is a question that must be decided by the involved stakeholders during the ascription. The argument represents an attempt to reduce the complexity of responsibility ascriptions by focusing on just one aspect, namely their assumed aim. Again, these considerations can be phrased in terms of Levels of Abstraction. One important aspect of LoA is that they define which variables are relevant in a situation and which ones can be observed. The functionalist LoA suggested here simply abstracts from the question of agency or personhood and concentrates on observable variables, namely the social consequences of computer use.

In the light of these considerations, it might be possible to ascribe moral responsibility from this instrumentalist or functionalist perspective. However, many philosophers would not agree to this and it might lead to equivocations. The central problem is that such a functionalist understanding of responsibility, while similar in appearance and structure, does not cover all of the aspects usually covered by responsibility as described in the preceding sections. One could therefore use a different term for the responsibility of computers, for example "quasi-responsibility". This is not a very elegant term and the author would welcome better suggestions. It is nevertheless useful because it indicates that we are looking at something very similar to responsibility which is nevertheless not quite the same thing as the concept of responsibility we usually encounter. The terminology follows Ricoeur (1983) who suggested a "quasi-agency" for historical collectives such as states or nations who can be described usefully as agents even though they are not traditional agents.

The term "quasi-responsibility" indicates that the speaker intends to use the idea of a social construction for the purpose of ascribing a subject to an object with the aim of attributing sanctions (the heart of responsibility) without regard to the question whether the subject fulfils the traditional conditions of responsibility. It shows that the focus of the ascription is on the social outcomes and consequences, not on considerations of agency or personhood. The concept was developed using computers as a main example but there is no fundamental reason why it could not be extended to other non-human entities, including animals.

This sort of quasi-responsibility seems less controversial than the application of traditional responsibility to computers. We have seen that responsibility is a social construct of ascription. If all those involved agree that there is a sort of responsibility where computers and information systems are considered legitimate subjects, then this ascription is possible. It can then be used as a tool to achieve socially desired results. If we find computers quasi-responsible for an undesirable result we can then proceed to develop sanctions or other consequences to meet them. We could for example sentence them to death or make their use a morally blameworthy action.

In the end, of course, all of this would translate back into responsibility by our classical subject, the person. If we, for example, decide to hold a particular information system quasi-responsible then this would lead to sanctions, such as outlawing the use of the system, which would eventually affect human beings. Why, one might ask, go through all the trouble in the first place? Why not say that humans are responsible for the actions of a computer? The reason is that the distribution of responsibility is often no longer clear. If a central military computer makes a mistake, who is to blame? The government, the state, the people, the user, the programmer, the vendor? Even if we could decide this theoretically, we would find that in most cases individuals did not know the results and are themselves lacking the conditions of responsibility. The construct of computer quasi-responsibility might be a step in overcoming these difficulties because it would facilitate mid-range solutions.

Mid-range solutions here means that sanctions are facilitated in an environment where traditional individual responsibility is no longer viable. They can be called "mid-range" because they move beyond the micro-level of the individual human being but they stop short of the macro-level of societal or global issues. The application of quasi-responsibility as developed here seems most promising in organisational contexts where moral problems may arise out of the actions of computers but where there is currently no way of attributing sanctions in order to improve the situation. An example may help visualising this. Let us imagine an enterprise resource

planning system, which forms the backbone of a large multi-national corporation. Such systems collect data, model markets and structure decisions in a wide range of situations. Their use has implications for the economic viability of the corporation but they can also be morally relevant, leading to replacement of human work, increasing (or decreasing) profits and all the impacts these may have on individual and social lives. Such systems are also hugely complex and expensive. They incorporate centuries or millennia of man-hours and it will rarely be possible to attribute their activities to individuals. They may be able to adapt to the environment and their outcomes or decisions are not always predictable.

Holding such a system quasi-responsible means that the system itself can be treated as a subject, which means that objects (such as financial implications but also issues of quality of life etc.) can be attributed directly to the system. This means that the sanctions, such as moral blame (or praise) or financial liability will be linked to the system. The interesting question then is whether this will improve the status quo. This might be the case if, because of moral blame, people will refrain from using it or if it were sentenced to death (i.e. if legal processes precluded further use of the system). We could then say that the system was quasi-responsible for the object, which led to certain consequences, which, in turn, had an impact on the social situation in which the system could be found.

Such quasi-responsibility will be linked to other types of responsibility ascriptions, be they traditional or quasi-responsibility themselves. Holding a computer system quasi-responsible will have manifest consequences for individuals, e.g. the CEO, the shareholders of the company, the users, the developers, the vendor etc. Quasi-responsibility should thus not be seen as an isolated instrument. It is always part of a larger net of responsibilities and can link to responsibility on the micro as well as the macro level. Its advantage is thus that it offers a Level of Abstraction on which the processes of responsibility can be applied independent of the problematic issues of personhood or agency.

Conclusion

If we agree with the argument of this paper and concede that a sort of quasi-responsibility of computers would be useful then this is of course not the solution to all of our problems. In fact, we would have to start several new lines of inquiry. First of all, we would have to define the computer as the subject of quasi-responsibility. Is it hardware, software, periphery, or the combination? Where does its quasi-responsibility begin and where does it end?

The next important question, and maybe the most difficult one to answer, is the one concerning the distribution of quasi-responsibility and traditional responsibility between computers, people, groups, organisations, and whoever else might be involved. It is clear that quasi-responsibility of computers should not lead to a general exculpation of humans. In fact, it is the other way around. Computer quasi-responsibility should be seen as a means for the facilitation of further responsibility ascriptions. There will clearly be at least one additional responsibility for humans. ''They will bear responsibility for preparing these systems to take responsibility'' (Bechtel 1985, 297). This means that social structures and institutions must be introduced which should allow for accountability of computers designers, programmers, managers, and users. Furthermore, the internal structure of computers might be modified in such a way that they become capable of discharging their quasi-responsibility. This may mean that computers have to become more adaptable to their environment (cf. Bechtel 1985).

The relationship between traditional responsibility and quasi-responsibility will give rise to further conceptual problems. One example of this is the notion of cyborgs. These cybernetic organisms render the difference between humans and computers difficult to identify. If we agree that humans can be ascribed responsibility and computers quasi-responsibility, then the question remains whether cyborgs are responsible or quasi-responsible. Since quasi-responsibility is the weaker concept (which means that the conditions to be ascribed quasi-responsibility are more easily met than for responsibility), one can probably state that they could be quasi-responsible. But that does not really answer the question at what point and under which conditions a cyborg will become responsible in the traditional sense. In light of the fact that cyborgs are no longer purely elements of science fiction (Cerqui 2002), such questions will have to be answered.

Acceptance of the notion of quasi-responsibility will necessitate further conceptual investigations. This paper has argued that the functionalist view of responsibility allows for the establishment of a related concept, namely quasi-responsibility, that can be applied to non-traditional subjects, notably computers. What the paper has not investigated are questions of related notions. One such notion is autonomy. Autonomy is an important concept with regards to the criteria of agency or personhood. The current paper has consciously tried to avoid discussing these, but there is arguably a link between the acceptability

of quasi-responsibility of computers and their autonomy. Autonomy may also be a relevant concept for distinguishing between quasi-responsibility and traditional responsibility. Another term that would need to be related to is "accountability". This term raises almost as many problems as the term responsibility. The literature does not agree on its definition, nor on its relationship to responsibility. There is little doubt, however, that accountability and responsibility are linked (cf. Stahl 2006), which raises the question whether the introduction of quasi-responsibility will require a quasi-accountability as well.

The contribution of this paper is clearly not a completely elaborated new theory of responsibility or quasi-responsibility of computers. It is an essay in Montaigne's understanding of the term; it is an attempt to develop thoughts. Starting from the recognition that responsibility and computers are in a complicated relationship, the paper developed an approach to the question whether computers can be responsibility subjects that differs from most of the literature. Instead of engaging in the questions of agency or personhood and the analysis of when computers can become subjects of (moral) responsibility, the paper introduced a different type of responsibility. This quasi-responsibility encompasses only a limited sub-set of traditional responsibility but it is explicitly applicable to non-human subjects, including computers. We have seen that this can be beneficial in that it facilitates solutions where traditional responsibility is no longer feasible. At the same time, quasi-responsibility raises a host of new questions and issues that will need to be addressed.

All of these problems appear soluble. There is, however, one big question in the background of this topic that was not yet mentioned. It is the question that Weizenbaum (1976) so eloquently asks: What are the limits of what we should let computers do? This is independent of what computers can do and refers quintessentially to the question in how far we are willing to accept changes to our self-image resulting from computer use. Attributing responsibility to computers, albeit only a limited form of responsibility, namely quasi-responsibility, has the potential to seriously affect our self-image. This is a question that this paper cannot answer from a theoretical point of view but that would have to be discussed in a societal discourse. To rephrase this question in the terms of this paper: Can (or should) man assume the responsibility for holding computers (quasi-)responsible?

References

C. Allen. Calculated Morality: Ethical Computing in the Limit. In I. Smit, Iva and G.E. Lasker, editors, *Cognitive, Emotive and Ethical Aspects of Decision Making and Human Action*, Volume I, pp. 19–23, (Workshop Proceedings, Baden-Baden, 31.07.-01.08.2002), 2002.

C. Allen, G. Varner and J. Zinser. Prolegomena to Any Future Artificial Moral Agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 12: 251–261, 2000.

C. Allen, I. Smit and W. Wallach. Artificial Morality: Top-Down, Bottom-Up, and Hybrid Approaches. *Ethics and Information Technology*, 7(3): 149–135, 2006.

Anonymous. Wearable Computers help Autistic Kids. [available: http://www.knowledgedock.com/pooled/articles/BF_NEWSART/view.asp?Q=BF_NEWSART_24460], accessed 09.11.2004, 2004a.

Anonymous. Augmented Reality Technology Promises Breakthroughs In Education And Cognitive Potential. [available: http://www.newstarget.com/001618.html], accessed 09.11.2004, 2004b.

K. Bayertz. Eine kurze Geschichte der Herkunft der Verantwortung. In K. Bayertz, editor, *Verantwortung: Prinzip oder Problem?*, pp. 3–71. Wissenschaftliche Buchgesellschaft, Darmstadt, 1995.

W. Bechtel. Attributing Responsibility to Computer Systems. *Metaphilosophy*, 16(4): 296–305, 1985.

H.W. Bierhoff. Verantwortungsbereitschaft, Verantwortungsabwehr und Verantwortungszuschreibung – Sozialpsychologische Perspektiven. In K. Bayertz, editor, *Verantwortung: Prinzip oder Problem?*, pp. 217 –240. Wissenschaftliche Buchgesellschaft, Darmstadt, 1995.

D. Birnbacher. Grenzen der Verantwortung. In K. Bayertz, editor, *Verantwortung: Prinzip oder Problem?*, pp. 143–183. Wissenschaftliche Buchgesellschaft, Darmstadt, 1995.

B.P. Bloomfield and T. Vurdubakis. Imitation Games: Turing, Menard, Van Meegeren. *Ethics and Information Technology*, 5(1): 27–38, 2003.

R. Brooks, *Flesh and Machines: How Robots Will Change Us*. Pantheon, New York, 2002.

D. Cerqui. The Future of Humankind in the Era of Human and Computer Hybridization: An Anthropological Analysis. *Ethics and Information Technology*, 4(2): 101–108, 2002.

G. Collste. The Internet-Doctor. In G. Collste, editor, *Ethics in the Age of Information Technology*, pp. 119–129. Centre for Applied Ethics, Linköping, 2000a.

G. Collste. Ethical Aspects of Decision Support Systems for Diabetes Care. In G. Collste, editor, *Ethics in the Age of Information Technology*, pp. 181–194. Centre for Applied Ethics, Linköping, 2000b.

R.T. George, *Business Ethics. 5th edition*. Prentice Hall, Upper Saddle River, New Jersey, 1999.

P. Fauconnet. La responsabilité: étude de sociologie. In M. Neuberg, editor, *La responsabilité — questions philosophiques*, pp. 141–152. Presses Universitaires de France, Paris, 1997, (orig: 1928).

J.M. Fischer. Recent Work on Moral Responsibility. *Ethics*, 110(1): 93–139, 1999.

L. Floridi and J.W. Sanders. On the Morality of Artificial Agents. *Minds and Machine*, 14: 349–379, 2004.

P.A. French, *Responsibility Matters*. University Press of Kansas, Lawrence, Kansas, 1992.

H.L.A. Hart, *Punishment and Responsibility – Essays in the Philosophy of Law*. Clarendon Press, Oxford, 1968.

D.M. Hausman and M.S. McPherson, *Economic Analysis and Moral Philosophy*. Cambridge University Press, Cambridge et al, 1996.

H. Jonas, *Das Prinzip Verantwortung*. Suhrkamp, Frankfurt a. M, 1984.

D.G. Johnson Computer Systems: Moral Entities but not Moral Agents. Ethics and Information Technology, 2007 (this issue).

D.G. Johnson and T.M. Powers. Computer Systems and Responsibility: A Normative Look at Technological Complexity. *Ethics and Information Technology*, 7(2): 99–107, 2005.

D. Kairy. A Philadelphia Story. Legal Affairs May/June 2003.

G. Korienek and W. Uzgalis. Adaptable Robots. Metaphilosophy (33:1/2) (Special Issue: Cyberphilosophy: The Intersection of Philosophy and Computing. Edited by J.H. Moor and T.W. Bynum), 83–97, 2002.

H.D. Lewis. Collective Responsibility. In L. May and S. Hoffman, editors, *Collective Responsibility: Fife Decades of Debate in Theoretical and Applied Ethics*, pp. 17–34. Rowman & Littlefield Publishers Inc., Savage, Maryland, 1991.

T.A. Lipinski, E. Buchanan and J.J. Britz. Sticks and Stones and Words That Harm: Liability vs. Responsibility, Section 230 and Defamatory Speech in Cyberspace. *Ethics and Information Technology*, 4(2): 143–158, 2002.

R.T. Long. The Irrelevance of Responsibility. In L. May and S. Hoffman, editors, *Collective Responsibility: Fife Decades of Debate in Theoretical and Applied Ethics*, pp. 118–145. Rowman & Littlefield Publishers Inc, Savage, Maryland, 1999.

L. May, *Sharing Responsibility*. University of Chicago Press, Chicago, 1992.

L. May and S. Hoffman. editors, *Collective Responsibility: Fife Decades of Debate in Theoretical and Applied Ethics*. Rowman & Littlefield Publishers Inc., Savage, Maryland, 1991.

M.S. Moore. Causation and Responsibility. In E.F. Paul, F.D. Miller and J. Paul, editors, *Responsibility*, pp. 1–51. Cambridge University Press, Cambridge, 1999.

M. Neuberg. Introduction à "La responsabilité". In M. Neuberg, editor, *La responsabilité – questions philosophiques*, pp. 1–24. Presses Universitaires de France, Paris, 1997.

H. Nissenbaum. Computing and Accountability. In D.G. Johnson and H. Nissenbaum, editors, *Computers, Ethics & Social Values*, pp. 526–538. Prentice Hall, Upper Saddle River, 1995.

E.F. Paul, F.D. Miller and J. Paul, editors. *Responsibility*. Cambridge University Press, Cambridge et al., 1999.

R.W. Picard, *Affective Computing*. MIT Press, London, Cambridge Massachusetts, 1997.

P. Ricoeur. Le concept de responsabilité — Essai d'analyse sémantique. In P. Ricoeur, editor, *Le Juste*. Editions Esprit, Paris, 1995.

P. Ricoeur, *Soi-même comme un autre*. Edition du Seuil, Paris, 1990.

P. Ricoeur, *Tems et récit – 1. L*'intrigue et le récit historique. Edition du Seuil, Paris, 1983.

R. Scheines. Computation and Causation. *Metaphilosophy* (33:1/2), Special Issue: *Cyberphilosophy: The Intersection of Philosophy and Computing*. Edited by J.H. Moor and T.W. Bynum: 158–180, (2002).

M. Schlick. Quand sommes-nous responsable? In M. Neuberg, editor, *La responsabilité – questions philosophiques*, pp. 27–38. Presses Universitaires de France, Paris, 1997, (orig: 1930).

N. Sherman. Taking Responsibility for our Emotions. In E.F. Paul, F.D. Miller and J. Paul, editors, *Responsibility*, pp. 294–323. Cambridge University Press, Cambridge, 1999.

B.C. Stahl. Accountability and Reflective Responsibility in Information Systems. In C. Zielinski, P. Duquenoy and K. Kimppa, editors, *The Information Society: Emerging Landscapes*. pp. 51–68, (IFIP WG 9.2 proceedings) Springer, New York, 2006.

B.C. Stahl, *Responsible Management of Information Systems*. Idea Group Publishing, Hershey, 2004a.

B.C. Stahl. Information, Ethics, and Computers: The Problem of Autonomous Moral Agents. *Minds and Machines*, 14: 67–83, 2004b.

J.M. Trigeaud, *L*'homme coupable — Critique d'une philosophie de la responsabilité. Editions Bière, Bordeaux, 1999.

M. Velasquez. Why Corporations Are Not Morally Responsible for Anything They Do. In L. May and S. Hoffman, editors, *Collective Responsibility: Fife Decades of Debate in Theoretical and Applied Ethics*, pp. 111–131. Rowman & Littlefield Publishers Inc, Savage, Maryland, 1991.

R.J. Wallace. Responsibility and the Moral Sentiment. Harvard University Press, Cambridge, Massachusetts/ London, England (1996).

J. Weizenbaum, *Computer Power and Human Reason*. W. H. Freeman and Company, San Francisco, 1976.