Ethics and Information Technology (2004) 6: 279–290
DOI: 10.1007/s10676-005-6710-5

A pragmatic evaluation of the theory of information ethics

Mikko Siponen
Department of Information Processing Science, University of Oulu, P.O. Box 3000, 90014, Oulu, Finland
E-mail: Mikko.T.Siponen@oulu.fi

Abstract. It has been argued that moral problems in relation to Information Technology (IT) require new theories of ethics. In recent years, an interesting new theory to address such concerns has been proposed, namely the theory of Information Ethics (IE). Despite the promise of IE, the theory has not enjoyed public discussion. The aim of this paper is to initiate such discussion by critically evaluating the theory of IE.

Key words: agency, computer ethics, information ethics, moral agency

Introduction

It has been argued that computer technology is a source of revolutionary moral problems (see Gorniak-Kocikowska 1996; Maner 1996; Gotterbarn and Rogerson 1997). Computers are seen as revolutionary, due to their logical malleability, which means that computers are able to carry out any activity that can be written off as a series of logical operations (Moor 1985). Gorniak-Kocikowska (1996) and Floridi (1999) see that the existing, or traditional,[1] theories of ethics, are unable to offer solutions to the computer ethics questions. As a result, a new theory of ethics is needed to solve computer ethics dilemmas. To date, one such theory – the theory of Information Ethics by Floridi – is proposed. Although Floridi's ideas are not totally new (similar views have been presented in the area of environmental ethics), they must be praised, in the spirit of Popper, for being bold and anti-conventional, aimed at challenging the fundamentals of moral thinking, including what constitutes moral agency, and how we should treat entities deserving moral respect. Unfortunately, for whatever reasons, Floridi's work has not attracted much interest, which is odd, given the promising nature of this work. Even though I have reservations about Floridi's theory, I believe it deserves to be discussed and better known. As a step in that direction, this paper critically evaluates the theory of Information Ethics (IE), put forward by Floridi (1998, 1999) and further developed by Floridi and Sanders (2001, 2003). I hope that this evaluation will trigger more discussion on the theory of IE.

The rest of the paper is organized as follows. The second section – "Introduction to the theory of IE" – introduces the theory of Information Ethics by Floridi (1998, 1999). The background for advocating a shift from an anthropocentric view of moral agents to an infocentric view is discussed in section "From anthropocentric to infocentric theory". Section "A new classification of moral agents," scrutinizes IE's idea to extend the class of moral agents to cover non-humans as moral agents. Section "The principle of ontological equality" discusses the principle of ontological equality. The normative aspect of IE is considered in section "Normative facets of the theory of IE". Vandalism and Kant's indirect duty, the supererogatory problem and the mathematical problem are considered in the last section. At the end concluding remarks are presented.

Introduction to the theory of IE

Floridi sees that computer ethics has not attracted the philosophical respect it deserves. In fact, philosophers treat it as a kind of carpentry ethics (Floridi 1999, p. 37). According to Floridi, this attitude stems from the interdisciplinary nature of the computer ethics and conservatism on the part of philosophers (Floridi 1999, p. 37). He believes another reason why

[1] The proponents of the new-theory-is-needed view, namely Gorniak-Kocikowska (1996), Floridi (1998, 1999) and Floridi and Sanders (2002), use the terms "existing ethics theories" and "traditional ethics theories" rather loosely, since they do not state what they exactly mean by the traditional theories of ethics.

computer ethics has not gained philosophical recognition originates from the view, according to which, the computer ethics is a practical decision-making issue which: "...can hardly add anything to already well-developed ethical theories." (Floridi 1999, p. 39).

However, in his view computer ethics is more than a 'carpentry ethics' and merits greater attention:

> "ICT [information communication technology], by transforming in a profound way the context in which some old ethical issues arise, not only adds interesting new dimensions to old problems, but may lead us to rethink, methodologically, the very grounds on which our ethical positions are based. Missing the latter perspective, even people who support the importance of the work done in CE are led to adopt a dismissive attitude towards its philosophical significance, and argue that there is no special category of computer ethics, but just ordinary ethical situations in which computers and digital technology are involved, and therefore that CE is at most a microethics, that is practical, field-dependent, applied and professional ethics." (Floridi 1999, pp. 38–39)

Floridi maintains that computer ethics lacks proper theoretical foundations, and thus what is needed right now are (new) conceptual foundations, which Floridi seeks to propound in his papers (Floridi 1998, 1999). Such a theoretical foundation of computer ethics, as Floridi (1999) calls it, would be the theory of Information Ethics, "*a model of macroethics*"[2] . Also, he sees that existing theories of ethics are not able to handle computer ethics problems: "*when consistently applied, both Consequentialism, Contractualism and Deontologism show themselves unable to accommodate CE-problems easily, and in the end may well be inadequate*" (Floridi 1999, pp. 38–39). Before critically scrutinizing this theory, let us consider why Floridi regards existing theories of ethics as inadequate. By existing theories of ethics, Floridi means "*consequentialism, contractualism and deontology*". He also refers to specific theories of ethics, such as Kant's "*moral imperatives*", "*the golden rule*" and "*the law of universality*" and Rawls' 'veil of ignorance'.

For Floridi, the existing theories of ethics can't accommodate the following types of problems: "*CE-problems not involving human beings*" and "*CE-problems with a ludic nature*" (Floridi 1999, p. 40). To give an example of these problems, consider the following passage. According to Floridi, a computer user:

> "...perceives computer crimes as games or an intellectual challenge...because of the remoteness of the process, the immaterial nature of information and the virtual interaction with faceless individuals, the information environment (the infosphere) is easily conceived of as a magical, political, social, financial dream-like environment, and anything but the real world, so a person may wrongly infer that her actions are as unreal and insignificant as the killing of enemies in a virtual game. The consequence is that not only does the person not feel responsible for her actions (no one has ever been charged with murder for having killed some monsters in a video game), but...". (Floridi 1999, p. 40)

In this passage, Floridi seems to take up the argument that people may regard computing as an amoral area, due to de-personalization and anonymity. For example, a hacker breaking into strangers' e-mail-boxes might argue that reading other's e-mail does not harm those people. The hacker might take the view that, after all, s/he is not changing anything in the system. Broadly speaking, such concerns have been recognized by many scholars (see Siponen and Vartiainen 2002). Notwithstanding, Floridi maintains that while the existing theories of ethics are incapable of addressing such problems, IE would deal with them well. After reviewing the concepts of the theory of IE, I must express some doubts about this claim.

Floridi also argues that a hacker may not understand the real consequences of his/her actions, or that the consequences of the hackers' actions do not feel as real for him/her. Therefore, the hacker may do bad things on the Internet. Furthermore, Floridi sees that the hacker may perceive these things as morally acceptable in the light of the Kantian universality thesis: "*The hacker can be a perfect Kantian because universality without any concern for the actual consequences of an action is ethically powerless in a moral game*." (Floridi 1999, p. 40).

Additionally, Floridi points out general[3] problems related to the use of consequentialism to solve computer ethics cases (1999, p. 40). He argues that actions in a virtual environment may not leave "*perceptible effects behind*" and that the computer creates a moral vacuum by distancing "*the agent from, and hence diminishes his sense of direct responsibility for his computer-mediated, computer-controlled and computer-generated action...*". Floridi also takes the view that the complexity involved "*often makes any reasonable calculation or forecasting of the long-term,*

[2] "*... we may use IE to refer to the philosophical foundation of CE.*" (Floridi 1999, p. 43).

[3] Unrelated to any particular case.

aggregate value of the global consequences of an individual's actions impossible...''. Finally, he states that:

> ''The increasing number and variety of computer crimes committed by perfectly respectable and honest people shows the full limits of an action-oriented approach to CE: computer criminals do not often perceive, or perceive in distorted way, the nature of their actions because they have been educated to conceive as potentially immoral only human interactions in real life, or actions involving physical and tangible objects. A cursory analysis of the justification that hackers usually offer for their actions, for example, is sufficient to clarify immediately that they often do not understand the real implications of their behaviour, independently of their technical competence. We have already seen how this problem affects a Deontological approach as well (the ludic problem)''.

Like the aforementioned general objections to Kant's ethics raised by Floridi, these general objections targeted at consequentialism, relate to problems where people do not see the consequences of their actions.

However, I see these problems as a part of the general computer ethics problematic. That is, these problems are not particularly related to specific ethics theories (e.g., consequentialism). For one thing, the theory of IE may as well face these or similar problems. For example, IE may also ''*diminish ethical sense*'' by widening the scope of moral responsibility to cover non-humans in the domain of morally responsible agents (see later). For another example, if a Kantian hacker does not understand the implications of his/her actions (cf., citation above), then how can the theory of IE, or any theory of ethics, help?[4] For another thing, are these objections to utilitarianism and Kant's ethics more a question of psychological factors, i.e., matters of awareness and knowledge (e.g., the hacker does not understand the implications of his/her actions), and therefore issues that can be tackled by proper education strategies, than particular problems of consequentialism? Indeed, when it comes to the Kantian hacker objection by Floridi, the hacker may well understand, after applying the universality thesis (''but what if other people treated me this way'') that his/her actions have certain consequences, and s/he the hacker personally would not like be on the receiving end of such actions. Thus, traditional theories of ethics (the universality thesis at least) may in fact be able to handle the situation.

Of the problems raised by Floridi, the calculation problem is perhaps the most consequentialism-oriented, and it may be true that utilitarian calculations are difficult to accomplish in the Internet-environment, for example. This is the case, since it may be difficult to estimate what kind of feelings of ''happiness'' and ''pain'' an action results in.

From anthropocentric to infocentric theory

Floridi differentiates theories of ethics in the light of an agent–patient relationship (Floridi 1999, p. 41). He argues that traditional theories of ethics, from virtue ethics to Kant, are anthropocentric and primarily focused on the agent committing the actions, not the object (or patient) receiving those actions:

> ''...any classic ethics is inevitably egocentric and logo-centric-all theorising concerns conscious and self-assessing agents whose behaviour must be supposed sufficiently free, reasonable and informed, for an ethical evaluation to be possible on the basis of his responsibility – whereas non-classic ethics, being bio-centric and patient-oriented, are epistemologically allocentric – i.e. they are centred on, and interested in, the entity itself that receives the action, rather than its relation to or relevance to the agent – and morally altruistic, and now include any form of life and all vulnerable human beings within the ethical sphere, not just foetuses, new-born babies and senile persons, but above all physical or mentally ill, disabled or disadvantaged people. This is an option that simply lies beyond the immediate scope of any classic ethics, from Athens to Könisberg.'' (Floridi 1999, p. 42)

In addition to classical ethics, using Floridi's concept, Land and Environment Ethics are also inadequate, according to Floridi, since these systems only consider live things as worthy of moral claims. As a result, ''*...a whole universe escapes their attention.*'' IE aims to remedy these defects by lowering:

> ''... the condition that needs to be satisfied, in order to qualify as a centre of a moral concern, to the minimal common factor shared by any entity, namely its information state [...]. The fundamental difference, which sets it [the theory of IE] apart

[4] For example, entropy is a bad thing, according to the theory of IE. Then we could argue following Floridi's logic that if the hacker does not understand what entropy means and what causes entropy (or what are the consequences of his actions: what actions cause entropy), how the theory of IE can address the problem faced by the hacker. However, we can hardly rule out any theory by the fact that we find a hacker who does not see the consequences of his her action (and therefore argue that theories paying attention to the consequences are flawed).

> from all other members of the same class of theories, is that CE raises information as such, rather than just life in general, to the role of the true and universal patient of any action, thus presenting itself as an infocentric and object-oriented, rather than just a biocentric and patient-oriented ethics. Without information there is no moral action, but information now moves from being a necessary prerequisite for any morally responsible action to being its primary object." (Floridi 1999, p. 43)

Thus, the shift from an anthropocentric to biocentric view is not enough for Floridi. What is needed is an infocentric view, which Floridi sets out to propose. So, "*IE is an Environmental Macroethics based on the concept of data entity rather than life*." (Floridi and Sanders 2003, p. 55). By 'infocentric' Floridi means that a key element of IE is the concept of information entity. Every existing entity, which is a consistent packet of information, and does not contain a contradiction in itself,[5] is an information entity (Floridi 1999, p. 43):

> "From an IE perspective, the ethical discourse now comes to concern information as such, that is not just all persons, their cultivation, well-being and social interactions, not just animals, plants and their proper natural life, but also anything that exists, from paintings to books to stars and stones; anything that may or will exist, like future generations; and anything that was but is no more, like our ancestors."

To overcome the limitation of the standard ethics theories, Floridi proposes an alternative criterion for moral agents.

A new classification of moral agents and levels of abstraction

By proposing a new criterion for moral agents, Floridi aims to capture a pre-theoretical but widely shared intuition, according to which, all life forms have certain properties that deserve to be respected (Floridi 1999, p. 42). According to Floridi:

> "An agent is any entity ... capable of producing information phenomena that can affect the infosphere. The minimal level of agency is the mere presence of an implemented information entity, in Heideggerian terms, the Dasein – the therebeinghood- of an information entity implemented in the infosphere." (Floridi 1999, p. 44)

He also distinguishes between responsible agent and normal (non-responsible) agents. The former refers to agents that are "*able to acquire knowledge-awareness of the situation and capable of planning, withholding and implementing their actions with some freedom and according to their evaluation*." (Floridi 1999, p. 44). So, humans, at least most of us, are responsible agents, while for instance, cats and dogs are not responsible agents, according to this view. Moreover, even though all information entities should be respected in some degree, all information entities do not deserve equal respect. Responsible agents (according to Floridi, responsible agents include God, humans, angels, gods, full-AI robots) deserve more moral respect than non-responsible agents. In recent article co-authored by Sanders, Floridi and Sander further modifies this view. According to Floridi and Sanders (2003), the criterion for a thing – or an information entity – to be a responsible agent, does not (anymore) require free will:

> "...the concept of moral agents not necessary exhibiting free will, emotions or mental states. That approach complements the more traditional one, common at least since Montaigne and Descartes, which considers whether or not (artificial) agents have mental states, feelings, emotions and so on. By focussing directly on 'mind-less' morality we are able to avoid that question and also many of the concerns of Artificial Intelligence. (Floridi and Sanders 2003).

As a result, artificial agents, animals and organizations alike can be considered as moral agents, i.e., morality is 'mind-less': "*artificial agents, particularly those in Cyberspace, extend the class of entities that can be involved in a moral situation*." (Floridi and Sanders 2003).

Floridi and Sanders (2003) restate their claim that existing ethics theories fail to incorporate non-human entities into the realm of moral agents. To clarify their point, let us first examine the terms used by Floridi and Sanders: 'an agent' refers to an entity that qualifies "*as the source of moral actions*" and "*can perform actions, again for good or evil*") and 'patients' are entities that qualify "*as receivers of the moral actions*" and "*can be acted upon for good or evil*".[6] Simply, Floridi and Sanders take the view that traditional theories of ethics, including Christian Ethics, hold that a moral agent must qualify as a moral

[5] Interestingly, Floridi does not reveal what such contradicting entities might be.

[6] "*Any action, whether morally loaded or not, has the logical structure of a variably interactive process, which relates a set of one or more sources...the agent a, which initiates the process, with a set of (one or more) destinations, the patient p, which reacts to the process*." (Floridi and Sanders 2003, p. 56).

patient and vice versa. They call this as 'the standard position'.[7] They maintain that there are some non-standard ethics where entities that qualify as moral agents also qualify as moral patients, but not vice versa. For example, according to this position, animals are regarded as moral patients, but not as agents. Floridi and Sanders seem to accept the widening of the definition of a moral patient, as suggested by 'non-standard ethics': "*...non-standard macroethics have been discussing the scope of P* [patient] *quite extensively*.", while "*Comparatively little work has been done in reconsidering the nature of moral agenthood and hence the extension of A* [agent]."

So, they want to extend the concept of a moral agent, as well to cover animals – "*...there is nothing wrong with identifying a dog as the moral agent that is the source of a morally good action*." (Floridi and Sanders 2003) – and certain software:

> "Secularism has contracted (some would say deflated) A [moral agent], while enviromentalism has justifiably expanded only P, so the gap between A and P has been widening... Limiting the ethical discourse to individual agents hinders the development of a satisfactory investigation of distributed morality, a macroscopic and growing phenomenon of global moral actions and collective responsibilities resulting from the 'invisible hand' of systemic interactions among several agents at the local level. Insisting on the necessary human-based nature of the agent means undermining the possibility of understanding another major transformation in the ethical field, the appearance of artificial agents (Aas) sufficiently informed, 'smart', autonomous and able to perform morally relevant actions independently of the human engineers who created them, causing 'artificial good' and artificial evil. Both constraints can be eliminated by fully revising the concept of 'moral agent'." (Floridi and Sanders 2003)

As this extract shows, another reason for extending the scope of moral agents is due to 'artificial agents', which are "*sufficiently informed, 'smart', autonomous and able to perform moral actions independently of the human engineers who created them...*". This statement raises a few questions. First, it is highly debatable whether software agents, or computer viruses, are autonomous in terms of Floridi and Sanders. They are, as Floridi and Sanders recognize, programmed by humans to "act" in a certain manner. And simply because software is programmed to act a certain way, it is odd to claim that any computer software is ultimately autonomous. Perhaps Floridi replies to this by arguing that it all comes down to the concept of the 'level of abstraction' (LoA, for short):

> "Indeed [level of] abstraction acts as a 'hidden parameter' behind exact definitions, making a crucial difference. Thus each definiens comes preformatted by an implicit Level of Abstraction (LoA, on which more shortly); it is stabilised, as it were, to allow a proper definition. An x is defined as y never absolutely (i.e. LoA-independently), as a Kantian 'thing-in-itself', but always contextually, as a function of a given LoA, whether it be in the realm of Euclidean geometry or quantum physics. When a LoA is sufficiently common, important, dominating or in fact is the very frame that constructs the definiendum, it becomes 'transparent', and one has the pleasant impression that x can be subject to an adequate definition in a sort of conceptual vacuum. Glass is not a solid but a liquid, tomatoes are not vegetables but berries and whales are mammals not fish. Unintuitive as such views can be initially, they are all accepted without further complaint because one silently bows to the uncontroversial predominance of the corresponding LoA. ...the trick does not lie in fiddling with the definiens or blaming the defiendum, but in deciding the adequate LoA...". (Floridi and Sanders 2003)

Here Floridi and Sanders offer up the concept of LoA. For example, shown in the extract, they see that "*glass is not a solid but a liquid*" depending on the LoA.[8] Yet, "*Turing solved the problem of 'defining' intelligence by first fixing LoA...*" (Floridi and Sanders 2003). In other words, the concept of LoA, in Floridi and Sanders' paper, as in computer science and information systems discipline in general, means that if, for example, a vehicle is a class, then 'van', 'bus' and 'motorbike' can be regarded as subclasses of the class vehicle. Also, 'van', 'bus', and 'motorbike' are at a lower level of abstraction than 'vehicle'. Consequently, "*...an entity may be described at a range of LoAs and can have a range of models*." (Floridi and Sanders 2003). This idea has been used for a long time in the disciplines of computer science, software engineering and information systems where

[7] In this view, theological ethics (e.g., Christian ethics) is lumped together with philosophical theories of ethics. However, the agent–patient relationship in Christian ethics is more complicated than Floridi and Sanders (2003) suggest. For example, a Liebnitzian Christian ethicist may question whether God is an agent in the sense of Floridi and Sanders (2001), if God does automatically what is morally right.

[8] It is my interpretation that this is also how they would explain the existence of 'conceptual muddles' (cf., Moor 1985), i.e., when we change the LoA in viewing these, they seem different.

an attempt to understand reality is made by using an abstraction mechanism, as Floridi and Sanders hint: "*The concept* [of LoA] *comes from modelling in science where the variables in the model correspond to observables in reality, all other being abstracted.*" (Floridi and Sanders 2003).

Then on the basis of this abstraction thinking, Floridi and Sanders make an interesting claim:

> "Consider what makes a human being (call him Henry) not a moral agent to begin with, but just an agent. Described at this LoA, Henry is an agent if he is a system, situated within and a part of the environment; which initiates a transformation, produces an effect or exerts power on it, as contrasted with a system that is (at least initially) acted on or responds to it, called the patient. At LoA there is no difference between Henry [a human being] and an earthquake. There should not be. Earthquakes, however, can hardly count as moral agents, so the LoA is too high for our purposes: it abstracts too many properties. What needs to be re-instantiated? Our proposal...indicates that the right LoA is probably one which includes the following three criteria: (a) interactivity, (b) autonomy and (c) adaptability." (Floridi and Sanders 2003)

Thus, are they perhaps arguing that in some very abstract sense both Henry and an earthquake are systems and the actions of both can be seen in the light of input and outputs? However, according to Floridi and Sanders, an earthquake, along with rock and tables, are not moral agents (in contrast to Henry), since earthquakes and rocks do not satisfy three properties necessary for things to be labeled as moral agents:

(a) interactivity: an agent must interact with its environment;
(b) autonomy: an agent must be able to change its states, so an agent must have at least two states;
(c) adaptability: an agent must be able to change "*the transition rules by which it changes state*".

They maintain that an agent is a moral agent, only if it is able to produce a morally 'qualifiable action', i.e., an action that causes moral evil or good:

> " (O) An action is said to be morally qualifiable if and only if it can cause moral good or evil. An agent is said to be a moral agent if and only if it is capable of a morally qualifiable action." (Floridi and Sanders 2003)

In other words, an agent is a moral agent, if it engages in morally relevant situations. This is an extension of Floridi and Sanders (2001) claim that artificial agents are able to cause evil (artificial evil).

But recognizing the above criterion (O), are software products capable of producing moral actions? Here is a reply by Floridi and Sanders (2003):

> "..are H and W moral agents? Because of (O) we cannot answer unless H and W become involved in moral action. So suppose that H kills the patient and W cures her. Their actions are moral actions. ..They both acted autonomously: they could have taken different courses of actions, and in fact we may assume that they changed their behaviour several times in the course of the action, on the basis of new available information. They both acted adaptable: they were not simply following orders or predetermined instructions; on the contrary, they both had the possibility of changing the general heuristics that led them to take the decisions they took, and we may assume that they did take advantage of the available opportunities to improve their general behaviour. The answer seems rather straightforward: yes, they are both moral agents. There is only one problem: one is a human being, the other is AA; the LoA adopted allows both cases. So can you tell the difference? If you cannot, you will agree with us that the class of moral agents must include AAs like webbots." (Floridi and Sanders 2003)

To my understanding, at the level of the interface and of a certain level of abstraction – albeit computer software are programmed by humans to "behave" in a certain manner – software from an ordinary users point of view seems to posses an autonomy, as suggested by Floridi and Sanders (2003). For instance, a computer virus that is taught to learn by a programmer, interpreted in that light, seems to behave autonomously, and therefore, is likely be considered as morally accountable (agent), according to Floridi and Sanders' (2003) criteria. Also, according to Floridi and Sanders, an organization is an agent, given that, albeit organizations consist of people, at a certain level of abstraction an organization can be viewed as a whole moral agent that interacts with other entities (cf., property I) in morally relevant situations, has different states (cf., property II), and is able to change its "behaviour" (cf., property III). Floridi and Sanders maintain that 'a futuristic thermostat' balancing the temperature in a hospital autonomously is "*morally charged since the LoA includes patients' well-being. It would be regarded as morally good if and only if its output maintains the actual patient's well-being within an agreed tolerance of their desired well-being.*" (Floridi and Sanders 2003). They also see MENACE, software that "*learns to play noughts and crosses*" as an agent: "*viewed at*

an appropriate LoA, then, MENACE system is an agent... and may also be viewed to have moral accountability." (Floridi and Sanders 2003).

I am not convinced on the role of the LoA. Namely, we can try to modify the LoA, and argue that different sorts of things are moral agents when we move the LoA. I do not believe that there is any real advantage to be gained by regarding some computer programs as moral agents. First, the criterion of morally responsible agents lacks an important feature. Namely, in order for something to be a morally responsible agent, we need to be able discuss and negotiate with it. And as we cannot negotiate with animals or any software, these entities are, at best, moral patients. Second, viewing non-human entities as moral agents rather begets the already existing problems of moral vacuums and moral distance (discussed below). If some sort of artificial life *a la* sci-fi stories were to exist in the future, would it rule out the applicability of traditional theories of ethics? Not necessarily, for example, to destroy a computer virus whether or not it is based on an artificial life *a la* sci-fi, we hardly need a new theory of ethics. Kantians may also note that a moral action requires an internal attitude of mind.

Floridi and Sanders (2001, p. 60) claim that artificial agents are able to cause evil, namely artificial evil. Hence, they want to extend the existing classification of evil, from moral evil (ME) and natural evil (NE) to artificial evil. This extension is based on the claim that the ME/NE classification is not able to explain all kinds of evil (e.g., artificial one), as the following extract illustrates:

> "... people are confronted by visible and salient evils that are neither simply natural nor immediately moral: an innocent dies because the ambulance was delayed by the traffic [example 1]; a computer-based monitor 'reboots' in the middle of surgery because of software not fully compatible with other programs also in use, with the result that the patient is at increased risk during the reboot period [example 2]. These examples could be easily multiplied. What kind of evils are these? 'Bad luck' and 'technical incident' are simply admissions of ignorance. Conceptually, they indicate the shortcomings of the ME vs. NE dichotomy." (Floridi and Sanders 2001, p. 59)

I don't think these examples provide any justification for artificial evil. As for example 1, it is unclear what "delayed by traffic" means? If it is accustomary, everyday, traffic jam, perhaps the ambulance drivers could have been selected another route (or send an ambulance helicopter). Nevertheless, I don't see what traffic problems have to do with artificial evil (are Floridi and Sanders going to blame traffic lights in example 1?). With respect to the example 2 (the software incompatibility case), it seems that the system developer forgot to do enough testing.

Floridi and Sanders offers other examples. They (Floridi and Sanders 2001, p. 61) claim that AAA, referring to artificial and autonomous agents, such as a computer virus or robot, may indeed cause 'artificial evil' (AE). To illustrate their argument, they compare this situation with that of the accusation that parents are responsible for their children "*the sins of the sons will be not passed on to their fathers*" (Floridi and Sanders 2001, p. 61) and pets:

> "Artificial 'creatures' can be compared to pets, agents whose scope of action is very wide, which can cause all imaginable evils, but which cannot be held morally responsible for their behaviour, owing to their insufficient degree of intentionality, intelligence and freedom. It turns out that, like in a universe without God, in cyberspace evil may be utterly gratuitous; there may be evil actions without any causing agent being morally blameable for them. Digital Artificial Agents are becoming sufficiently autonomous to pre-empt the possibility that their creators may be nomologically in charge of, and hence morally accountable for their misbehaviour. And we are still dealing with a generation of agents fairly simple, predictable and controllable. The phenomenon of potential artificial evil will become even more obvious as self-produced generations of AAA evolve." (Floridi and Sanders 2001, p. 61)

Even if we accept that parents are not responsible for the actions of their children, I have serious difficulties in seeing how this illustrates the notion that the developers of software have no responsibility over the software they have created. Clearly, kids in general are likely to have more freedom in choosing their behaviour than software. This is partly owing to the fact that software are always programmed by humans to "behave" in a specific manner, while it is difficult to accept that children can be programmed in the same way. Floridi and Sanders might defend their view of the moral accountability of "autonomous" software by arguing that, although software is developed by humans, software is able to learn and then "change" its behaviour in ways that the original developers are unable to explain:

> "Artificial agents... are able to adapt their behaviour on the basis of experience (in only an indirect sense were the programmers of Deep Blue responsible for its win, since it 'learnt' by being exposed to volumes of game; thus its programmers were quite

> unable to explain, in any of the terms of chess parlance, how Deep Blue played).'' (Floridi and Sanders 2001, p. 62)

My criticism of this thesis of regarding non-human entities (e.g., software) as morally responsible agents is pragmatic. If we go along with the thesis by Floridi and Sanders, it would mean that we can, once again, start blaming computers for our mistakes. In other words, we can claim that ''I didn't do it – it was a computer error'', while ignoring the fact that the software has been programmed by people to ''behave in certain ways'', and thus people may have caused this error either incidentally or intentionally (or users have otherwise contributed to the cause of this error). Ladd (1982, 1989) argues that too much blame is wrongly laid on computers, or computer errors, and by means of this excuse, people avoid moral responsibility altogether. For example, in the case of a tax office sending a wrongful request to pay more taxes, we may be told that this mistake was due to a computer error. It is interesting that Floridi agreed earlier that one problem in computer ethics is depersonalization and not perceiving the consequences of our activities. It seems to me that this broadening of (responsible) moral agents to cover non-human entities does everything but avoid this problem. For one thing, following Ladd, we are now able to blame computer viruses, instead of their developers as well as people who, intentionally or by acting too carelessly, allow viruses to spread. In the same vein, it seems difficult to imagine that we would blame an atomic bomb that goes off unexpectedly, even the bomb would satisfy the three properties of moral agent, suggested by Floridi and Sanders.

For another thing, if the problem is that we have difficulties in seeing how our actions, implemented by computers, affect other human users, how will the introduction of non-human moral agents contribute to the solving of this problem? Would it not be better to stress that ''there are no computer errors *per se*'', but actions we carry out using computers can affect other human beings, who have feelings like us. If people have difficulty respecting other people in for example the Internet environment, how we can assume that they will be able to display moral respect for non-human moral agents, such as computer software.

Floridi and Sanders (2003) offer a position, which contrasts with the argument just made. They argue that dismissing the view that a moral agent can be something other than a human being ''*has led to an enormous increase in individuals' moral responsibility*'' (Floridi and Sanders 2003). To my understanding, are they saying that if we are unable to hold, say, a piece of computer software as morally responsible for an action, we are forced into blaming humans, perhaps the developer or user of the software in question? And would doing this would increase human responsibility (which is a bad thing since we would not be able to blame computers)? I am lost here, since in their earlier works (e.g., Floridi 1999) they claimed that a key problem is that people do not see and understand the real consequences of their actions. And now Floridi and Sanders claim that, indeed, there is too much moral responsibility on individuals' shoulders and it should be cut down by placing it on the shoulders of non-human entities, such as computer software.

Another motive for the new classification is the point that software ''*is largely constructed by teams... working software is the result of maintenance over its lifetime and so not just of its originators... automated tools are employed in the construction of much software...*'' (Floridi and Sanders 2003). This claim entails three problems. First, it may be the case that it is often difficult in practice to identify the individual human being who developed the particular features of the software in question. But is it not impossible (in theory). Moreover, I have difficulty in seeing how this possible practical problem means that we should not try to find responsible human developers, but rather hold software as morally responsible. Second, cars, aeroplanes and other products of engineers follow the same development practice, as described by Floridi and Sanders (2003). In fact, this typical software engineering description of software development (e.g., where requirement capture, design, implementation and testing are carried out by separate teams) stems from other engineering fields. But is this reason (software is developed by many persons) really the relevant reason for extending the scope of the domain of moral agents? I find it odd to even consider that should we regard cars and airplanes as moral agents, given that it may be equally difficult to retrace the particular individual responsible for making the specific part of the plane or car in question. It should be also noted that software development using large teams is not the only possibility. For example, the current trend in agile software development (Abrahamsson et al. 2003) stresses the use of very small teams, or individuals who develop software alone or in very small teams. Computer viruses (even though they would meet the necessary properties of moral agents suggested by Floridi and Sanders), for instance, seem to have been developed by individuals (not teams). Third, even if we accept their proposal, the problem remains of where to draw the line between the moral responsibility of developers and the software (considered herein as an artificial moral agent) they created.

The principle of ontological equality

Floridi sees that every information entity has a certain degree of moral worth *per se*, owing to fact that they exist as information entities[9] in the infosphere.[10] The "ontological equality principle" guides us on how we should respect and regard the information entities. This "ontological equality principle" holds that every information entity has a right to exist "*simply for the fact of being what it is*" (Floridi 1999, p. 45). Thus, "*Any information entity has a 'Spinozian' right to persist in its own status, and a 'Constructionist' right to flourish, i.e. improve and enrich its existence and essence*." (Floridi 1999, p. 45). On the one hand, one might see the appeal of this ontological equality principle in a general sense; i.e., we should respect different things as such. On the other hand, one may ponder whether the "ontological equality principle" is actually too demanding in the final analysis. For example, are fishing, hunting and the felling of trees morally deprecatory actions, according to this "ontological equality principle"? After all, these objects (e.g., fish, animals) are clearly information entities, and we seem to be taking away their rights to exist (by fishing and hunting). Thus, is, for example, fishing always morally wrong? According to the "ontological equality principle, the answers seem to be yes. We should not kill or destroy these information entities, but let these *develop in a way which is appropriate to their nature*" (Floridi 1999, p. 44). This idea seems to be too all-embracing, if the killing of an insect (e.g., a cockroach) is morally wrong in every case. For example, it has been suggested that cockroaches spread SARS in Hong Kong. In this situation, I find it odd that one should let such cockroaches live and to spread SARS, as seems to be prescribed by the ontological equality principle.

This ontological equality principle of IE also becomes interesting from the point of view of anti-virus activity. It seems that a computer virus is regarded as an information entity in terms of IE: it is a consistent packet of information. This granted, in the light of the ontological equality principle, computer viruses (being information entities) should be allowed to "*develop in a way which is appropriate to their nature*" (Floridi 1999, p. 44) and have the "*right to persist in its own status*" (Floridi 1999, p. 44). Clearly, viruses are designed to spread and/or destroy (or cause other harm). Thus, if spreading computer viruses, for example, is an action that is appropriate to their nature, anti-virus activity seems to be morally wrong *per se*, at least from this point of view. No doubt this treatment of computer viruses is infocentric, and indeed, it respects the information entity (the computer virus) *per se*. However, spreading computer viruses definitely omits the moral respect and care the receiver of a computer virus is entitled to deserve, which is also one of Floridi's aims. However, considering IE from the normative aspect, we might conclude that computer viruses should be stopped, even destroyed altogether as this would be a less evil action than letting the viruses do their jobs, even though the destruction of viruses is a morally blameworthy action (in the light of the ontological equality principle).

Normative facets of the theory of IE

To solve such moral problems, Floridi offers four moral laws, which will be discussed next. These moral laws are based on the concept of entropy: (disorder) in the infosphere:

> "...entropy is a quantity specifying the amount of disorder, degradation or randomness in a system bearing energy or information" (Floridi 1999, p. 44). By asking "what is good for an information entity and the infosphere in general?"

Floridi suggests that entropy is a bad thing, and one should always avoid causing entropy:

> "IE suggests that there is something even more elementary and fundamental than life and pain, namely being, understood as information, and entropy. IE holds that being/information has an intrinsic worthiness...". (Floridi 1999, p. 45)

A morally blameworthy act increases entropy, whilst a morally good act extends information quantity, improves information quality and enriches information variety in the infosphere:

> "IE evaluates the duty of any rational being in terms of contribution to the growth of the infosphere, and any process, action or event that negatively affects the whole infosphere – not just an information entity – as an increase in its level of entropy and hence an instance of evil." (Floridi 1999, p. 45)

The wrongness of an action, or of rival actions, should be measured in terms of entropy (e.g., the action producing the least entropy is the least worse one). To be

[9] "*IE holds that every entity, as an expression of being, has a dignity, constituted by its mode of existence and essence..., which deserves to be respected...*" (Floridi 1999, p. 44).

[10] "*The infosphere is the environment constituted by the totality of information entities – including all agents – processes, their properties and mutual relations.*" (Floridi 1999, p. 44).

more specific, the four moral principles (laws), the normative dimension of IE, are (in order of increasing moral value): (1) an action ought not cause entropy in the infosphere[11]; (2) entropy ought be prevented (from existing in the infosphere); (3) entropy ought to be removed (from the infosphere); and (4) "*information welfare ought to be promoted by extending (information quantity), improving (information quality) and enriching (information variety) in the infosphere*" (Floridi 1999, p. 47). With the help of these moral laws, Floridi aims to improve the infosphere:

> "According to its [IE] semi-teleological approach (information processes are goal-driven, but their goals are internal goals of a reflective self-development of the infosphere, they are not heteronomous), the best thing that can happen to the infosphere is to be subject to a process of enrichment, extension and improvement without any loss of information, so on the most commendable courses of action always have a caring and constructionist nature. The moral agent is an agent that looks after the information environment and is able to bring about positive improvements in it, so as to leave the infosphere in a better state than it was in before the agent's intervention." (Floridi 1999, p. 50)

One may question whether entropy really offers a sound basis for a moral theory, given that thinking (Dyson 1979), a deletion of a piece of writing using an eraser, and a heat production in general produce entropy (Gell-Mann 1994).

Vandalism and Kant's indirect duty, the supererogatory problem and the mathematical problem

Floridi puts forward a critique of existing theories of ethics, targeted at Kant's ethics in particular. Floridi argues that different sorts of (harmless) vandalism, such as a boy stoning abandoned cars (Floridi 1999, pp. 53–54), cannot be deemed morally blameworthy by reference to Kantian theory. Floridi offers two examples to illustrate this point. First, Floridi argues that "*its* [Kant's ethics] *ends/means maxim is inapplicable*" (Floridi 1999, p. 54). Second, a possible problem for the universality thesis, if applied to this case, is its possible bias towards subjective decisions, leading to condoning the above mentioned act of vandalism. Well, how about the Kantian indirect duty? Perhaps a Kantian thinker would regard this action (a boy stoning abandoned cars) as morally wrong due to the requirement of an indirect duty. It may be claimed that we have an indirect duty to avoid such stoning of abandoned cars. Namely, such action may bring up viciousness in humans; hence "[...] *any such cruelty for sport cannot be justified*." (Kant in Hursthouse 2000, p. 76).[12] Nevertheless, if the action is really regarded as 'vandalism' in general, we may consider it wrong in terms of utilitarianism or the universalizability thesis. To start with an utilitarian example, presume that people in general (excluding the boys) would regard this as morally wrong; if so, a utilitarian may regard it wrong. And if it is harmless (e.g., to put it in utilitarian sense: it is neutral measured in terms of pain and happiness), then it may be accepted in the light of utilitarianism. But if it is really a harmless action, is it really wrong, even if it causes entropy? For example, a martial artist who hits bricks, or a piece of wood, just for fun (or to maintain his/her skills) may also cause entropy. But is there anything wrong with such actions?

This leads us to "*the supererogatory problem*". Floridi argues that IE handles "*the supererogatory problem*" better than consequentialism:

> "...since goodness is a relative concept – relative to the amount of happiness brought by the consequences of an action – Consequentialism may simply be too demanding, place excessive expectations on the agent and run into the supererogatory problem, asking the agent, who wishes to behave morally, to perform actions that are above and beyond the call of duty or even of his good will. In IE, this does not happen because the morality of a process is assessed on the basis of the state of the infosphere only, i.e. relationally, not relative to other processes. So while Consequentialism is in principle satisfied only by the best action, in principle IE prizes any single action, which improves the infosphere according to the laws specified above, as a morally commendable action, independently of the alternatives. According to IE, the state of the world is always morally depreciable (there is always some entropy), so any process that improves it is already a good process." (Floridi, 1999, p. 51)

However, the point that "*IE prizes any single action*" also works the other way. Namely, IE also condemns an activity that violates the four moral laws. The violation of the null law is always the most morally

[11] This principle is referred to as the null law: "*IE treats evil as monotonic: nothing justifies the infringement of the first moral law (an increase in entropy may often be inevitable, but is never morally justified, let alone approved)*." (Floridi 1999, p. 50).

[12] "[...] *he who is hard in his dealings with animals becomes hard in his dealings with men*." and "*We can judge the heart of a man by his dealings with animals*.' (Kant in Hursthouse 2000, p. 75). This view may be applied to the actions of boys stoning abandoned cars.

blameworthy thing that one can do (cf., above "*nothing justifies the infringement of the first moral law*"). Does this means that IE may be in fact *"too demanding, place excessive expectations on the agent and run into the supererogatory problem, asking the agent, who wishes to behave morally, to perform actions that are above and beyond the call of duty...*" (cf., above), if actions causing any increase in the level of entropy are always morally wrong? How can IE blame consequentialism for being too demanding, if the killing of an insect, or a deletion of a piece of writing using an eraser (cf., Gell-Mann 1994), are always morally wrong in the account of IE?

Floridi (1999) also takes up the mathematical problem to compare consequentialism with his theory. He argues that if any quantitative calculation has a role in moral thinking, then IE triumphs over utilitarianism:

> "If any quantification and calculation is possible at all in the determination of a moral life, then IE is clearly in a much better position than Consequentialism. Consequentialism already treats individuals as units of equal value but relies on a mere arithmetical calculus of aggregate happiness, which in the end is far too simplistic, utterly unsatisfactory and amounts to little more than a metaphorical device, despite its crucial importance within the theory. On the contrary, if required, IE may resort to a highly developed mathematical field (information theory) and try to adapt to its own needs a very refined methodology, statistical means and important theorems, in terms of Sigma logarithms and balanced statistics." (Floridi 1999, p. 51)

Here Floridi argues that utilitarianism calculates 'happiness' and absence of 'pain', while 'entropy' can be traced back to Information Theory, and can be therefore calculated mathematically. I am sceptical with respect to this assumed advantage for practical reasons. Let me first recall that we should offer computer ethics teaching to every computer user (Siponen and Vartiainen 2002). Recognizing this, I am doubtful that the ordinary man in the street is able to make such 'entropy' calculations, while ordinary computer users may very well have some sort of picture of what is 'happiness' and 'absence of pain'. That is, although the utilitarian key concepts, 'happiness' and 'absence of pain', may be less exact concepts in the natural science (or naïve positivistic) sense than 'entropy', the fact remains that all and sundry have a gut feeling on 'happiness' and 'pain' – as opposed to 'entropy'. In this practical respect, utilitarianism clearly outperforms IE.

Concluding remarks

IE argues that traditional theories are anthropocentric, and therefore conflict with our everyday common sense, according to which, non-human forms of life also need to be respected. Although, this starting point of Floridi's theory of IE may well accord with our commonsense, the theory itself hardly satisfies this criterion. Indeed, the theory of IE is less pragmatic than its key competitors (such as utilitarianism and the universalizability theses). That is, that ordinary people may not easily associate entropy with wrongdoing (of course, this critique does not imply that IE is fundamentally flawed because it is less pragmatic than universality thesis, for example). For that reason, it is argued that the theory of IE may be less suitable for dealing with problems of moral motivation and moral distance than the universalizability theses (see Siponen and Vartiainen 2002). What is my rationale for reaching this conclusion? I postulate that ordinary end-users do not connect entropy with wrongdoing: would you, the reader, connect it to wrongdoing? The process of pondering whether entropy will be increased by an action on our part may not awaken our feelings of moral sensitivity: would one be distressed if one heard that one had increased the amount of entropy in the infosphere?[13] On the other hand, if the action is considered in the light of the universalizability thesis, i.e., "What if this were to happen to me?" or "What if other people treated me that way?" the issue touches the individual directly. Perhaps Floridi himself recognizes this possible practical limitation of the theory of IE, when he points out that people behaving in morally blameworthy fashion on the internet, for example, "*do not understand the real implications of their behaviour, independently of their technical competence*." (Floridi 1999, p. 41). Here, Floridi seems to imply that the problem stems from not seeing the consequences of one's actions; it is not caused because we do not connect "entropy" to an act of wrongdoing. It is also interesting to note that when Floridi and Sanders talk about morally loaded actions as a necessary condition for an agent to be a moral agent, they do not stress 'entropy'. Instead, they stress hospital patients' 'well-

[13] It is also noteworthy to mention that Brandt (1978, pp. 166–170) suggests that moral codes should be connected with a sensation of disfavour and culpability, they should be considered important (Brandt's first criterion). Finally, people following the moral code are admired by other people (Brandt's second criterion). On the account of Brandt's first criterion, entropy is hardly associated to a blameworthy action. It is also problematic how the theory of IE can satisfy admiration criterion by Brandt (Brandt's second criterion), provided that basically all our actions are morally blameworthy in the light of the theory of IE. Seldom are people who never mow their lawn or let cockroaches run wild in their house admired by other people.

being', as an example of an ethical situation having moral relevance. Then why do they not mention "avoidance of entropy" – instead of well-being – to illustrate what constitutes a morally relevant action? Perhaps they realize that people simply do not see what the concept of 'entropy' really has to do with moral evil, as discussed above.

Furthermore, I levelled two criticisms against the new classification of moral agents by Floridi and Sanders. First, I pointed out that such a classification seems to make things worse by opening up the possibility of shifting the blame on to computers, and computer errors, instead of the negligence of computer users and software developers, which may have in the final analysis caused these errors. Second, I noted that only when we are able to discuss and negotiate with artificial agents (e.g., computer software), are we in a position to regard these artificial entities as morally responsible agents?

Acknowledgement

I would like to thank the reviewers for their valuable comments and insights on earlier versions of this article.

References

P. Abrahamsson, J. Warsta, M.T. Siponen and J. Ronkainen. New Directions on Agile Methods: A Comparative Analysis. *Proceedings of International Conference on Software Engineering*, Portland, Oregon, USA, May 3–10, 2003.

R.B. Brandt. *A Theory of The Good and The Right*. Oxford University Press, UK, 1978.

F.J. Dyson. Time Without End: Physics and Biology in an Open Universe. *Reviews of Modern Physics*, 51(3): 447–460, 1979.

L. Floridi. Does Information have a Moral Worth in Itself? Computer Ethics: Philosophical Enquiry (CEPE'98), in Association with the ACM SIG on Computers and Society, London School of Economics and Political Science, London, 14–15 December, 1998.

L. Floridi. Information Ethics: On the Philosophical Foundation of Computer Ethics. *Ethics and Information Technology*, 1(1): 37–56, 1999.

L. Floridi and J.W. Sanders. Artificial Evil and the Foundation of Computer Ethics. *Ethics and Information Technology*, 3: 55–66, 2001.

L. Floridi and J.W. Sanders. Mapping the Foundationalist Debate in Computer Ethics. *Ethics and Information Technology*, 4: 1–9, 2002.

L. Floridi and J.W. Sanders. On the Morality of Artificial Agents (Forthcoming). In A. Marturano and L. Introna editors, Ethics of Virtualities. *Essays on the Limits of the Bio-power Technologies*, to be published for the series Culture Machine, Athlone Press, London, 2003.

M. Gell-Mann, *The Quark and the Jaguar: Adventures in the Simple and the Complex*. W.H. Freeman and Company, New York, NY, USA, 1994.

D. Gotterbarn and S. Rogerson. The Evolution of the Uniqueness Revolution (What's So Special About Moral Problems in IT). In *Proceedings of the Conference on Computer Ethics: Philosophical Enquiry*, pp. 103–109. Erasmus University Press, Rotterdam, The Netherlands, 1997.

K. Gorniak-Kocikowska. The Computer Revolution and the Problem of Global Ethics. The Computer Revolution and the Problem of Global Ethics. *Science and Engineering Ethics*, 2: 177–190, 1996.

R. Hursthouse, *Ethics, Humans and Other Animals: An Introduction with Readings*. Routledge, London, 2000.

J. Ladd. Collective and Individual Moral Responsibility in Engineering: Some Questions. *IEEE Technology and Society Magazine*, 1(2): 3–10, 1982.

J. Ladd. Computers and Moral Responsibility: A Framework for an Ethical Analysis. In: C. Gould, editor, *The Information Web: Ethical and Social Implications of Computer Networking*, pp. 207–227, 1989.

W. Maner. Unique Ethical Problems in Information Technology. *Science and Engineering Ethics*, 2(2): 137–154, 1996.

J.H. Moor. What is Computer Ethics? *Metaphilosophy*, 16(4): 266–275, 1985.

M.T. Siponen and T. Vartiainen. Teaching End-User Ethics: Issues and a Solution Based on Universalizability. *Communications of the Association for Information Systems*, 8(29): 422–443, 2002.