Kenneth Einar Himma

Associate Professor

Department of Philosophy

Seattle Pacific University (USA)

<http://home.myuw.net/himma>

FOUNDATIONAL ISSUES IN INFORMATION ETHICS

Forthcoming in a special issue of *Library Hi-Tech* devoted to information ethics

Information ethics, as is well known, has emerged as an independent area of ethical and philosophical inquiry. There are a number of academic journals that are devoted entirely to the numerous ethical issues that arise in connection with the new information communication technologies; these issues include a host of intellectual property, information privacy, and security issues of concern to librarians and other information professionals. In addition, there are a number of major international conferences devoted to information ethics every year. It would hardly be overstating the matter to say that information ethics is as “hot” an area of theoretical inquiry as medical ethics.

There are also issues that are more abstract and basic than the substantive issues with which most information ethics theorizing is concerned. These issues are thought to be “foundational” in the sense that we cannot fully succeed in giving an analysis of the concrete problems of information ethics (e.g., are legal intellectual property rights justifiably protected?) until these issues are adequately addressed. Such foundational issues include (1) whether information ethics is unique among other areas of ethics in the sense that it requires its own foundation; (2) which methodological principles should guide substantive ethical inquiry; and (3) what the moral status of information is. The following provides an overview of some of the research on these and related issues.

Is Information Ethics Theoretically Unique?

A number of theorists have attempted to justify the study of computer ethics as a field by arguing that computer ethics is unique in some theoretically significant sense. On this line of analysis, the use of computing technologies gives rise to unique meta-ethical, ethical, or epistemic difficulties that warrant treating those problems as a theoretically unified class that requires specialization. While a number of authors argue that computer ethics is distinct in some theoretically significant way (henceforth the Uniqueness Thesis), they differ with respect to the sense in which they think it is unique.

A. Interpreting the Uniqueness Thesis

There are a number of different interpretations of the Uniqueness Thesis.¹ First, one might argue that computer ethics is unique in the sense that some acts involving computers possess ethical qualities not possessed by any other type of act. Since the existing concepts of obligatory, permissible, good, and supererogatory (i.e., good involving a sacrifice that is beyond the call of duty) purport to adequately describe all existing ethical qualities of acts, this interpretation can adequately be expressed as follows:

The Meta-Ethical Thesis: There are acts in computer ethics that cannot adequately be characterized by the traditional concepts of obligatory, permissible, good, and supererogatory.

The Meta-Ethical Thesis, then, makes the very strong claim that the very meta-ethical foundation for general and applied ethical thinking is inadequate.²

¹ Much of the subsequent analysis is taken from Tavani (2002a) and Himma (2004d).

² I know of no one who actually holds this view.

Second, one might hold that existing ethical theories (or so-called “first principles”) might be adequate to resolve problems in other areas of applied ethics but are insufficient to resolve certain problems involving computer use.³ Computer ethics, according to this view, is unique in the following sense:

The Normative Thesis: Computer technologies present ethical problems that cannot, as an objective matter, be adequately resolved by recourse to existing ethical theories.

The Normative Thesis, then, states a claim about the objective coverage of the existing set of first-principles – and not about our abilities to apply those principles. In particular, it asserts that not even a perfectly intelligent observer could correctly evaluate certain ethical issues solely by recourse to existing first-principles because those principles do not fully cover those acts.

Third, one might argue that certain types of reasoning useful in other areas of applied ethics are of limited utility in the context of computer ethics. Walter Maner (1996) argues, for example, that we lack the resources to build analogical bridges that would link certain problems involving computer use to problems in other areas in applied ethics and asserts that problems in computer ethics are *epistemically* indeterminate⁴:

The Epistemological Thesis: Computer technologies present ethical problems that resist the analogies that enable us to see how ethical theories and first-principles apply in other areas of applied ethics.

The Epistemological Thesis, then, asserts no more than that the techniques that frequently help us in seeing how existing normative materials apply to specific problems are inadequate to help us with problems in computer ethics. While it might be that existing theories and principles are logically adequate to address these problems, we lack sufficient epistemic resources to determine *how* they apply.

Finally, some writers argue that computing machines instantiate properties that are ethically unique among members of some class of entities. The idea here is that computing machines instantiate ethically significant properties (i.e., properties that are relevant in evaluating acts involving computers) that are instantiated by no other non-living artifact:

The Property Thesis: Computer technologies possess moral properties that are unique among non-living artifacts (though such properties might be possessed by living things).

³ It is important to note that a general normative ethical theory usually consists in a set of first-principles together with some sort of justification of that set. Thus, for example, Mill offers a utilitarian first-principle that is justified by an argument to the effect that the relevant sorts of pleasure are the only things that human beings seek for their own sake – and not for the sake of anything else.

⁴ Walter Maner, “Unique Problems in Information Technology,” *Science and Engineering Ethics*, vol. 2, no. 2 (1996), 137-54, reprinted in D. Micah Hester and Paul J. Ford, *Computers and Ethics in the Cyberage* (Upper Saddle River, NJ: Prentice Hall, 2001), 53.

Thus, for example, one might argue that computers are unique among machines in instantiating some form of moral standing (e.g., moral personhood).⁵

B. Evaluating the Uniqueness Thesis

The Meta-Ethical Thesis can pretty much be rejected at the outset. There *is* little reason to think that existing ethical categories are insufficient in the way described by the Meta-Ethical Thesis. Our set of existing categories can be inadequate only if either there is more than one ethically significant class of wrongful acts or there are more than the three known classes of non-wrongful acts (i.e., permissibility, goodness, and supererogatory). No one has attempted an argument in support of either position.

Indeed, it is not clear how one could make a plausible argument for this claim. To make out a case for a novel category, one would have to argue that at least one of these categories is *qualitatively* overbroad in the sense that it has been mistakenly applied to acts to which none of the complementary concepts is applicable. And it is just not clear what sort of argument could be given for this intuitively implausible claim; for example, it is not clear how one could justify the claim that some act that is wrongly characterized as permissible when neither it nor its negation is obligatory. We simply don't have the intuitive resources to even begin to evaluate such claims because such categories are utterly foreign to our existing judgments and practices.

The problem with the Normative Thesis's claim that existing theories give wrong answers is that there is no way to ground that idea in any existing theory. It is not enough to simply to assert that deontological theories, utilitarian theories, care theories, right-based theories all give the wrong result; one must justify this claim by reference to some reasonably plausible theory or principle. But someone who rejects all existing theories and principles lacks an intuitive theory or principle that could serve as a standard for evaluating the adequacy of existing theories and principles.

The problem with the Epistemological Thesis is that epistemic indeterminacy is far from being unusual in applied ethics. Ethicists disagree on a wide variety of non-computer-related issues, including abortion, the death penalty, economic justice, and so forth, despite agreeing on the sanctity of life and the undesirability of poverty. If, as seems uncontroversial, it is true that a set of principles is epistemically indeterminate with respect to an issue only if there are two conflicting conclusions that can plausibly be grounded in those principles, then epistemic indeterminacy is everywhere in applied ethics.

Maner believes that indeterminacy in computer ethics is a function of certain unique factual properties of computing technologies, but this is problematic. It might be true, for example, the fact that the unique cost-efficiency of computing technologies makes it possible to steal large sums of money without inflicting significant harms,⁶ but this doesn't have any obvious relevance. One might

⁵ Luciano Floridi, I believe, is a proponent of this view. See Floridi 2002..

⁶ If I steal half a cent each month from each of one hundred thousand bank accounts, I will have removed \$6000 from those accounts over the course of a year. Since the cost of each transaction is negligible, I make the entire amount of each theft.

plausibly argue that, other things being equal, a theft of \$6000 that inflicts a significant harm on its victims is morally worse than a theft of \$6000 that does not. Notice, however, that such an argument makes no reference whatsoever to the fact that computer calculations are uniquely cost effective.

The Property Thesis also faces difficulties. While proponents of this thesis usually believe that computers have moral standing (i.e., are owed a moral obligation of respect), none of the distinguishing properties of computers have any obvious ethical relevance with respect to the issue of moral standing because these properties seem to have only instrumental value (i.e., value as a means to an end) and lack the intrinsic value (i.e., value as an end-in-itself) thought to be deserving of moral respect or responsibility. For example, we value the malleability, speed, and complexity of computers, not because we think these properties are valuable for their own sakes, but because they make it possible for us to *use* computers to accomplish a variety of different valuable ends. Likewise, we value the fact that computers are cost effective because it means that computers are cheaper to *use* for these ends than other efficacious means.

This distinguishes such properties from life, sentience, understanding, and free will. Life, as the common intuition goes, has a value independent of the various uses to which it can be put and the instrumental benefits that derive from such uses. Happiness and freedom from suffering are pursued for their own sake – and not for the sake of something else. The distinguishing properties of computers simply don't have the right sort of foundation in intuitive views about what is and isn't intrinsically valuable to support the claim that they are unique among artifacts in having a moral standing.

C. Computer Ethics as a Subarea in Applied Ethics

The idea that there exists a fairly substantial class of problems that uniquely and essentially involve some distinguishing characteristic of computer technology seems so obvious it is hard to imagine how it could plausibly be denied. It is clear, for example, that the intellectual property issues implicated by certain P2P file-sharing technologies arise out of the distinct ability of computing technology to produce and distribute perfect reproductions of digital music files at very little cost to the user. Similarly, the ethical issues involving free speech on the Internet arise because computing technologies are unique in making possible the widespread dissemination of text and graphics files. It is obvious that, as an empirical matter, there is no other existing technology that can do the sorts of things that give rise to such ethical problems.

In this way, the class of ethical problems involving computing technologies closely resembles the class of ethical problems involving medical technologies. It is uncontroversial that many problems in medical ethics arise out of some distinctive aspect of medical technology. The issue, for example, of whether it is permissible to clone human beings obviously depends on a technology that is unique to the medical field. The same is clearly true of the privacy concerns raised by the possibility of testing for genes that put people at risk for various life-threatening diseases like cancer.

But notice that while I have stolen an ethically (and legally) significant amount, the loss to each victim is negligible (six cents).

But no one in medical ethics has ever proposed an analogue for medical ethics of any of the uniqueness claims that are thought necessary to justify study of medical ethics. Medical ethicists have taken it for granted that existing meta-ethical categories and normative ethical theories are logically sufficient to evaluate all the problems of medical ethics. It can, of course, be very difficult to determine what the objectively correct answer to a particular problem; existing categories and theories are hence epistemically indeterminate in those instances. But such indeterminacy, as we have seen, is not unique to any particular class of ethical problems.

Indeed, the justification for treating medical ethics as a distinct field relies on a number of comparatively mundane considerations. First, the problems in medical ethics are problems that are faced by a particular class of professionals – namely, those who work in the health care industry. Second, these issues have an ethically significant impact on human well-being; whether or not euthanasia, cloning, or genetic testing is morally justified can make a tremendous difference in the quality of our lives. Third, it is reasonable to think, as is true in general, that people who specialize in problems that are logically related to one another are, as an empirical matter, likely to produce higher quality work; long gone are the days when it was reasonably easy to publish in a variety of different areas without specializing in any particular area.

But exactly the same sorts of considerations apply to computer ethics. First, the problems in computer ethics pose dilemmas for a particular class of professionals – namely those working to design, develop, and implement computer technologies. Second, the issues that arise from these activities bear vitally on human well-being and quality of life⁷; it matters tremendously to me, for example, what sort of property rights I have in the contents of my hard drive and in the contents of my ideas. Third, it is reasonable to think that people who acquire a wide-ranging knowledge of the technical characteristics and ethical problems of such technologies are, as an empirical matter, considerably more likely to produce quality ethical arguments addressing those problems. Understanding computing technologies will help to produce well-informed ethical views – regardless of how we characterize those technologies.

General Theories of Information Ethics

There have also been a wide variety of substantive theoretical claims about the nature and methodology of information ethics. Some of these theories are limited to describing the nature of solving problems in information ethics and its relationship to solving problems in other areas. Others attempt to articulate methodologies or substantive principles that should be employed in solving problems in information ethics. Yet others take information ethics as a starting point for developing a broader methodology for addressing issues in applied ethics. A brief summary of some of these views follows below.

⁷ As James H. Moor (1998) puts the point, “almost everyone would agree that computing is having a significant, if not a revolutionary, impact on the world” (p. 15).

Bernard Gert (1999) has argued that problems in applied ethics, including problems in information ethics, should be resolved by referring to the “common morality.” The common morality, as Gert defines it, is “the [shared] *moral system* that people use ... in deciding how to act when confronting moral problems and in making their moral judgments” (Gert 58). As Gert points out, there is much more agreement among persons in any given culture on moral issues than there is disagreement; though it is quite natural to focus energy on what we disagree upon, “such controversial matters form only a very small part of those matters about which people make moral decisions and judgments” (Gert 57).⁸ Issues in information ethics (and, for that matter, other areas), on Gert’s view, should be resolved by recourse to these shared moral judgments.

Luciano Floridi (2002) attempts to ground a unified approach to solving problems in information ethics in four basic ethical principles that govern the behavior of agents in the “infosphere”; as he describes the matter:

By suggesting that information objects may require respect even if they do not share human or biological properties, IE provides a general frame for moral evaluation, not a list of commandments or detailed prescriptions (compare this to the “emptiness” of deontological approaches). In Floridi 1999a, 2001a and 2001b this frame has been built in terms of ethical stewardship of the information environment, the infosphere (298).

Floridi posits four “universal laws against information entropy”: (1) information entropy ought not to be caused in the infosphere; (2) information entropy ought to be prevented in the infosphere; (3) information entropy ought to be removed from the infosphere; and (4) the infosphere ought to be protected, extended, improved, enriched, and enhanced.

Jeroen van den Hoven (1997) argues that problems in information ethics are properly addressed by the method of reflective equilibrium developed and made famous by John Rawls. Ethical reasoning, on this view, should strive towards producing the most coherent *system* of beliefs in which general ethical claims (e.g., intentionally causing harm is wrong) and more specific case-based judgments (e.g., hitting Ken Himma in the face because his views make you mad is wrong) help to support, justify, and explain one another. Justifying a position on an issue in information ethics will, thus, require showing that the position belongs to the most coherent system of ethical beliefs.

Some information ethicists have focused primarily on concerns about biases that are implicit in ICT or in information-ethics theorizing. Philip Brey (2001), for example, argues that ethical theorizing should be a multi-disciplinary and multi-level inquiry that, among other things, identifies and discloses “embedded normativity” in ICT. Allison Adam (2001) argues that much theorizing in information ethics has failed to take into account important issues regarding sex and gender and advocates the adoption of a feminist ethic of care as a means of achieving a different perspective on various problems in information ethics.

⁸ It is noteworthy that Gert’s proposal, more than any other discussed below, tends to reflect the methods and arguments of applied ethicists in other areas.

A number of information ethicists have focused on explaining the scope of information ethics relative to other areas of ethics. Krystna Gorniak-Kocikowska (1996) argues that the revolutionary worldwide impact of computers points to a need for a new “global” theory of ethics. Gorniak-Kocikowska argues that just as Bentham’s utilitarianism was developed in response to the development of new ICTs (i.e., the printing press), so too a new ethical theory is likely to emerge in response to the development of the modern computer and associated technologies. Moreover, this ethical theory is likely to have a global impact, unlike utilitarianism’s limited influence in Europe and the U.S., and will replace the ethical traditions in most nations.

Deborah G. Johnson (1999) has argued that computer and information ethics will, in some sense, disappear into a more general area of ethical inquiry. On Johnson’s view, information technologies will become so fully integrated into our lives that we will no longer regard its presence as exceptional or as one requiring a compartmentalized way of thinking. Information technologies and the associated problems will become fully absorbed and incorporated into our day-to-day lives. When this happens, the need for a distinct computer and information ethics will disappear; problems in information and computer ethics will be approached holistically alongside other problems.

The Moral Value of Information

One would quite naturally expect that the various ethical issues that arise in connection with ICTs might ultimately depend for their resolution on more general claims about the moral value of information. It seems reasonable to think that we cannot determine whether intellectual property rights exist and are legitimately protected by the state without knowing something about the value of intellectual content. It might be, for example, that information is so valuable that people have a general fundamental (i.e., not derived from other rights) moral right to information. Alternatively, it might be that information is owed some sort of moral duty of respect that is incompatible with legal protection of intellectual property; perhaps, information “wants” to be free.⁹ The following sections discuss some of the issues that arise in connection with assessing the moral status of information.

A. Moral Standing and Two Kinds of Moral Value

Theoretically prior to any substantive moral issues is a more general issue concerning what beings or entities have “moral standing.” *X* has *moral standing* if and only if (1) some moral agent has at least one duty regarding the treatment of *X* and (2) that duty is owed to *X*. The importance of (2) should not be overlooked. Even if my dog lacks moral standing, you still have a duty not to kick my dog; but that duty is “indirect” because owed to *me*, and not to my dog. If, however, you owe my dog a duty not to kick it, then that is a direct duty because it immediately concerns the being to whom the

⁹ See Tavani 2002. See also John Perry Barlow, “The Economy of Ideas”: available at <http://homes.eff.org/~barlow/EconomyOfIdeas.html>.

duty is owed. To have moral standing and be a moral patient, then, is to be the beneficiary of at least one direct duty.¹⁰

While the notion of having moral standing is traditionally associated with having “intrinsic value,” there are two distinct senses of the notion that figure into moral theorizing. The first is concerned with the sort of ends that are characteristically pursued by practically rational agents. In this sense of the phrase, a thing, state, or entity has intrinsic value if and only if practically rational agents typically value it as an end-in-itself; a thing has merely instrumental value, in contrast, if and only if it conduces as a means to some agent’s end. Mill, for example, famously argues that the only thing that people characteristically pursue for its own sake and not for the sake of something else is happiness. Accordingly, all other things, like money or vacations, are instrumentally valuable as a means to securing the intrinsically valuable end of pleasure.¹¹

There are different views about how intrinsic value in this sense figures into moral theorizing. Mill deduced his utilitarianism from the view that the only thing people characteristically value intrinsically is happiness; if happiness is the only thing that people characteristically value for its own sake, then it is the sole ground of moral value. Much more modestly, it is reasonable to think that people have some sort of morally protected interest in what they characteristically value intrinsically – though this tells us nothing about the strength or nature of such an interest (e.g., it tells us nothing about whether it rises to the level of a right).

The second sense of the phrase “intrinsic value” is concerned to identify a class of objects that are entitled to some measure of moral respect. Entities that have intrinsic value in this sense are moral patients deserving of at least minimal respect from practically rational moral agents. Unlike an entity that has only instrumental value, an entity with intrinsic value may not be treated as just an object to be “used.” Whereas the appropriate manner for deciding how to treat things with only instrumental value is cost-benefit analysis, things with intrinsic value are entitled to some level of moral consideration in deliberations by moral agents about what to do. Things with intrinsic value in this second sense count as moral patients with entitlements that must be satisfied.

Theorists have considered the issue of whether information or information objects have intrinsic value in both of these respects. These areas of inquiry will be discussed below in subsections B and C.

B. Fundamental Right to Information as Grounded in Intrinsic Value

As the importance of information becomes more evident, theorists have begun to suggest there is a general fundamental (i.e., one independent of other logically distinct rights) moral right to information. Though the substance of this right has not been described in much detail, the idea is that

¹⁰ Theorists typically distinguish two general categories of moral standing. A being may, of course, enjoy full-strength moral standing; such a being has a full array of moral rights that are defined by constraints on the outward behavior of agents. But a being may also have a much weaker form of moral standing in the following sense: a being that is merely *morally considerable* has only a right to have its well-being taken into consideration in the deliberations of moral agents.

¹¹ See J.S. Mill, *Utilitarianism*; available from <http://www.utilitarianism.com/mill1.htm>.

we have an interest in information *per se* that rises to the level of a basic moral right that ought to be recognized and protected by the law in every nation. Since, as a perfectly general matter, information has intrinsic value, beings like us have a fundamental general moral right to information.

The argument is not implausible on its face. It is clear, for example, that information can have tremendous value as a means. As Mark Alfino and Linda Pierce (1997) point out, having information can conduce to the end of well-being in a variety of ways. Other things being equal, the more information we have, the better able we are to satisfy our own needs and steer clear of threats. Information about the relationship between exercise and good health, for example, can help me to live a healthier life.

It is also clear that information can have tremendous value as an end-in-itself. Many people pursue areas of inquiry, like mathematical inquiry, for no other reason than it is inherently interesting and valuable – that is, regardless of any instrumental applications it might have. Indeed, if much work in pure mathematics, for example, has been used to create instrumentally valuable technologies, one of the most celebrated achievements in pure mathematics is thought to have no such applications. A few years ago, Andrew Wiles devised a successful proof of Fermat's Last Theorem (i.e., there are no positive integers x , y , z , and $n > 2$ such that the equation $x^n + y^n = z^n$ is true). Despite the fact that this theorem lacks any practical applications, mathematicians devoted countless hours to finding a proof or disproof of this theorem.

Nevertheless, it is fairly easy to find examples of informative propositions that cast doubt on the empirical claim that people characteristically assign intrinsic value to information.¹² Consider, for example, the empirical proposition (call it P) that expresses the number of hairs I have on my head at this moment. P is undeniably informative in virtue of expressing a fact about me and hence a fact about the world, but notice that it is not the sort of fact that is characteristically useful to people. While it might have instrumental value in certain circumstances (perhaps involving a bar bet), these circumstances are rare.

In addition, there are mathematical propositions that even mathematicians do not characteristically regard as intrinsically valuable. Kurt Gödel showed that any first-order formalization of arithmetic was incomplete by showing that there were certain true self-referential sentences of arithmetic that could not be proved. These Gödel sentences, in effect, asserted that "I am a true sentence of arithmetic that cannot be proved." Gödel showed that the incompleteness problem could not be solved by adding a finite set of Gödel sentences to the axioms of arithmetic because there are an infinite number of these sentences. While it is reasonable to hypothesize that some of these sentences have intrinsic value because of what they tell us about mathematics, it is reasonable to hypothesize that most mathematicians would regard the vast majority of them as utterly lacking intrinsic value.

¹² For more detailed discussion of the implications of these examples, see Himma 2004b and 2004c.

Such examples show that not all information has intrinsic value and hence that having an informative nature, by itself, is not sufficient for having intrinsic value. *P* is information and therefore has an informative nature, but is not plausibly characterized on ordinary views as being intrinsically valuable. Each Gödel sentence is information and therefore has an informative nature, but only a small number of them are thought to be intrinsically valuable.

But these examples also show something stronger – namely, that *no proposition* has intrinsic value simply in virtue of having an informative nature. If some entity has value *in virtue* of having an informative nature, then it is the fact that it has *that* nature that confers that value on it. Insofar as *any* entity has intrinsic value in virtue of having an informative nature, *every* entity that has that nature must also have intrinsic value. If any piece of information does not have intrinsic value in virtue of having an informative nature, then no piece of information has intrinsic value in virtue of having an informative nature. Accordingly, if *P* utterly lacks intrinsic value despite having an informative nature, then no proposition can have intrinsic value in virtue of its being information.

This should not be taken to deny that many people would converge in assigning intrinsic value to some informative propositions – or even to kinds of information (e.g., religious or mathematical information), but it does show that we do not have any sort of general fundamental right to information that is grounded in its being intrinsically valuable *qua* information. While it is surely true that we have rights to many kinds of information, the explanation for this cannot be grounded in the claim that all information is intrinsically valuable. The explanation will have to make reference to other properties or qualities of the particular kind of information (e.g., that it is political speech) and hence will depend on other moral interests and rights.

C. “Information Objects” as Having Moral Standing

Luciano Floridi (2002) grounds his conception of Information Ethics (IE) as a general ethical theory in an expansive claim about moral standing. On Floridi’s view, every entity in the universe can be understood as being an “information object” that is intrinsically valuable and hence deserving of at least minimal moral respect.¹³ Every existing entity, whether sentient or non-sentient, living or non-living, natural or artificial, has some minimal moral worth, as he puts it, in virtue of its existence “*qua* information object.”

¹³ Floridi explains the concept of an information object by way of a helpful analogy: “Consider a pawn in a chess game. Its identity is not determined by its contingent properties as a physical body, including its shape and colour. Rather, a pawn is a set of data (properties like white or black and its strategic position on the board) and three behavioural rules: it can move forward only, one square at a time (but with the option of two squares on the first move), it can capture other pieces only by a diagonal, forward move, and it can be promoted to any piece except a king when it reaches the opposite side of the board. For a good player, the actual piece is only a placeholder. The real pawn is an “information object.” It is not a material thing but a mental entity (288).” What constitutes the being of a pawn, then, is defined by the rules of chess that govern what can be done with it. Thus, what Floridi terms the “real pawn” is a “cluster of information,” which includes a description of the physical properties of the relevant subject (e.g., a pawn) along with a description of the rules and functions that define the subject’s behaviors and operations.

Floridi offers a couple of arguments for this view. First, Floridi argues that the clear historical trend is in favor of expanding the set of objects having moral standing. While the claim that only human beings have moral standing has predominated throughout history, the 20th Century has seen arguments in support of the claim that non-human animals, plants, ecosystems and even art objects have moral standing and are owed duties of respect. It is quite natural, on Floridi's view, to further expand the class of objects with moral standing to include information objects.

Second, he argues that we cannot explain certain acts without assuming that all objects deserve respect in virtue of being information objects. Despite lacking the various capacities that we usually think give rise to moral standing, we have the strong intuition that some human being who is born brain-dead – call her Mary – deserves some measure of respect – though less than she would have been entitled to had she been born alive. The only plausible justification for this practice, on Floridi's view, is that such a being has moral standing in virtue of being an information object.

Both arguments are vulnerable to objection.¹⁴ First, the historical argument is simply the wrong kind of argument – even if we assume that the theories ascribing moral standing to animals, plants, etc. are all correct. If we construe the argument as citing the history of such theorizing as a reason for us to expand the moral community to include information objects, it is problematic because the mere fact that theorists have, as an empirical matter of historical fact, expanded the moral community to include animals, plants, and land doesn't give us any reason, by itself, to think that theorists *ought* to further expand the moral community to include information objects. Purely descriptive facts about the behavior of theorists have no normative implications whatsoever with respect to what our theories ought to be.

Second, the claim that Mary has moral standing in virtue of being an information object simply cannot adequately explain the kind of things that we do out of respect for Mary's corpse. A stillborn child, for example, is treated with far more reverence than a miscarried fetus, but both are utterly on the same level *qua* information objects. *Qua* information objects, they consist of a set of propositions that describe their various properties, operations, and functions. Mary's nature *qua* information object, then, cannot explain why we show more respect for her if she were stillborn than if she were miscarried.

Indeed, if Floridi's theory lacks even sufficient resources to explain the different levels of respect we show to a stillborn child and a rock. From the standpoint of the claim that a thing has moral standing in virtue of its nature *qua* information object, there is no difference between Mary *qua* information object and a rock *qua* information object; to put it in roughly Quinean terms, to be *qua* information object is to be a set of propositions that has a certain logical structure. For this reason, Floridi's approach cannot seem to explain even the intuition that Mary's corpse deserves more respect than a rock.

¹⁴ See Himma 2004a.

The Possibility of Artificial Moral Agents

There are two different ways in which human beings are morally special that correspond to two different ways in which information or ICTs might be morally special. The first is that we are moral patients in the sense that others owe moral obligations to us. The second is that we are moral agents who are properly held accountable for our behavior under moral standards. The last section considered, among other things, whether information entities are moral patients; the following considers whether information entities might be moral agents.

A. Expanding the Boundaries

A number of researchers have suggested that traditional views about moral agency might be false. These researchers are inclined to reject the traditional view that human beings are the only moral agents in the universe (apart from God), exploring the possibility that artificial (presumably non-living entities) might be moral agents as well. Some have suggested a continuum of moral agency with human beings on one end and entities as mundane as speed bumps on the other (Latour 1994; Keulartz et.al., 2004).

Such attributions of moral agency have been supported with arguments propounding theoretical criteria for moral agency. Floridi and Sanders (2004), for example, define moral agency as follows:

For all X, X is a moral agent if and only if X has the following properties: (1) X and its environment are capable of acting upon each other (the property of “interactivity”); (2) X is able to change its states internally without the stimuli of interaction with the external world (the property of “autonomy”); (3) X is capable of changing the transition rules by which it changes state (the property of “adaptability”); and (4) X is capable of acting such as to have morally significant effects on the world.

The first three conditions are necessary and sufficient for something to count as an “agent.” The fourth, in effect, adds a requirement that one has the capacity to affect the world in ways that matter from the standpoint of moral evaluation. Accordingly, the notion of a moral agent, as Floridi and Sanders define it, requires a particular kind of agency that includes the ability to impact the world in morally relevant ways.

Although the point of Floridi and Sanders analysis is to suggest the possibility of artificial agents involving ICTs, it is important to note that it implies that human beings are not the only existing moral agents in the world. Rattlesnakes and even poisonous spiders might count as moral agents on this definition. It is clear, for example, that rattlesnakes and spiders have the properties of interactivity and adaptability; both interact with the world and at least occasionally learn. It is also clear that both are capable of acting such as to have morally significant effects on the world since both are capable of killing human beings.

One might doubt that they satisfy the autonomy requirement as formulated by Floridi and Sanders, but that would be wrong. If, as seems reasonable, a spider’s/rattlesnake’s apprehension of hunger (a purely internal state) is capable of causing a desire or intention to eat (a purely internal state), then it follows that spiders and rattlesnakes have the property of autonomy as Floridi and Sanders

define it. While it is true that the causal chain leading to the apprehension of hunger might lead outside the organism to the external world, there is nothing in their analysis of autonomy to preclude this. As long as even one internal state of the organism is sufficient to cause a distinct internal state, the organism is autonomous according to the definition above.

This, however, is problematic from the standpoint of ordinary views. While there is no *a priori* reason to rule out the possibility of artificial agents (maybe we are simply computers made of meat), we simply do not talk as though spiders and rattlesnakes are moral agents; and we do not treat them as such. Indeed, there is something radically counterintuitive (almost comical) about the suggestion that a spider or rattlesnake is “evil” if “evil” is meant literally; spiders and rattlesnakes might be “harmful” and “pests,” but they cannot be “evil” precisely because they are not moral agents whose behavior is subject to moral requirements..

B. The Concept of Moral Agency

It would be helpful to examine the concept of moral agency in more detail. The concept of a moral agent is ultimately a normative notion that picks out the class of beings whose behavior is subject to moral requirements. The idea is that, as a conceptual matter, the behavior of a moral agent is subject to being governed by moral standards, while the behavior of something that is not a moral agent is not governed by moral standards. Adult human beings are, for example, typically thought to be moral agents, while cats and dogs are not.

The idea of moral agency is conceptually associated with the idea of being accountable for one’s behavior. To say that one’s behavior is governed by moral standards and hence that one has moral duties or moral obligations is to say that one’s behavior should be guided by and hence evaluated under those standards. Something subject to moral standards is *accountable* (or morally responsible) for its behavior under those standards.¹⁵

This account is not controversial among philosophers or information ethicists. Indeed, so mainstream is this idea that it is routinely reproduced in reference works in philosophy. For example, the *Routledge Encyclopedia of Philosophy* describes the concept of moral agency as follows:

Moral agents are those agents expected to meet the demands of morality. Not all agents are moral agents. Young children and animals, being capable of performing actions, may be agents in the way that stones plants and cars are not. But though they are agents they are not automatically considered moral agents. For a moral agent must also be capable of conforming to at least some of the demands of morality.¹⁶

¹⁵ Though the notions of moral responsibility and accountability are not synonyms, they are intimately related. To say X is *morally responsible* for y is to say y is causally related to some act of X that violates a moral obligation for which X is appropriately held accountable. As a general matter, the term “moral responsibility” is used to designate acts or events that involve breaches of one’s moral duties.

¹⁶ Shorter Routledge, p. 692.

Similarly, the Stanford Encyclopedia of Philosophy (Eshleman 2001) treats “moral agency” as being logically equivalent to “moral responsibility.”

What this shows is that both ordinary linguistic practice and professional consensus converge on the idea that, as a conceptual matter, moral agents are accountable under moral standards for their behavior. It is not clear in advance what the status of certain high level artificial entities capable of action might be; it is not obvious, for example, that an artificial “agent” with a brain structured isomorphically to ours, which would include billions of parallel processing units connected to each other by billions of pathways, is not properly characterized as a “moral agent.” But this much is clear: spiders and rattlesnakes are not moral agents.

The problem with Floridi’s and Sander’s analysis is not with the analysis of *agents* (as opposed to *moral agents*). While there might be some difficult cases to work out, it seems clear that both spiders and rattlesnakes, like children, are agents in the sense that they autonomously initiate action. Indeed, their analysis of the concept of agent is intended to harmonize with our intuitions that agency is primarily a matter of being able to generate activity internally and hence includes nearly all humans and at least higher non-human animals.

Rather, the problem is that the fourth condition is not sufficient to warrant the application of the term “moral agents” to beings that are merely “agents.” Not all agents are moral agents (e.g., very young children). But being an agent capable of doing something that has effects that are morally relevant is not enough either; if it were, a poisonous spider would be rationally characterized as “evil” and accountable for its behavior.

It is generally thought that there are two capacities necessary and jointly sufficient for an agent to be a moral agent accountable for her behavior. The first is the capacity to freely choose one’s acts.¹⁷ The idea here is that, at the very least, one must be the direct cause of one’s behavior in order to be characterized as freely choosing that behavior; something whose behavior is directly caused by something other than itself has not freely chosen its behavior. If, for example, A injects B with a drug that makes B so uncontrollably angry that B is helpless to resist it, B has not freely chosen her behavior.

The second capacity necessary for moral agency is, as traditionally expressed, “knowledge of right and wrong”; someone who is incapable of understanding and applying moral requirements is not a moral agent and not appropriately held accountable for her behaviors. This does not mean that moral agents always, so to speak, get the moral calculus correct, but rather that they have the ability to understand and apply the basic concepts and principles of morality.

Beings lacking one of these capacities are incapable of conforming their behavior to moral requirements and are hence are not properly held accountable for their behavior. If one does not have a

¹⁷ Not surprisingly, this entails that only agents are moral agents. Agents are distinguished from non-agents in that agents initiate responses to the world that count as *acts*. Only something that is capable of acting counts as an agent and only something that is capable of acting is capable of acting freely.

basic grasp of morality, then one cannot be guided by its principles. If one does not freely choose one's behavior, then one cannot control one's behavior in such a way as to ensure that it conforms to moral principles. It is a meta-ethical principle that one cannot be obligated to do, other things being equal, what one is incapable of doing.¹⁸ As the matter is frequently put, "ought implies can."

C. *The Possibility of Artificial Agents*

This, of course, is not to suggest that artificial entities that satisfy the first three conditions of Floridi's and Sander's analysis (i.e., artificial agents) cannot count as moral agents. It is rather to suggest that more is required than simply an ability to cause effects that are morally relevant, which will be trivially true in a large variety of cases (after all, it is not difficult to make someone frustrated or unhappy – states that are morally relevant). For an artificial agent to be a moral agent, it must have the capacity to choose its actions "freely" and understand the basic concepts and requirements of morality.

It is clear that an artificial agent would have to be a remarkably sophisticated piece of technology to be a moral agent. It seems clear that a great deal of processing power would be needed to enable an artificial entity to be able to (in some relevant sense) "process" moral standards. Artificial free will presents different challenges: it is not entirely clear what sorts of technologies would have to be developed in order to enable an artificial entity to make "free" choices – in part, because it is not entirely clear in what sense *our* choices are free. Free will poses tremendous philosophical difficulties that would have to be worked out before the technology can be worked out; if we don't know what free will *is*, we are not going to be able to model it technologically.

While the necessary conditions for moral agency as I have described them do not explicitly contain any reference to consciousness, it is reasonable to think that each of the necessary capacities presuppose consciousness. The idea of accountability, which is central to the meaning of "moral agency," is sensibly attributed only to conscious beings. It seems irrational to praise or censure something that isn't conscious – no matter how otherwise sophisticated its computational abilities might be. Praise, reward, censure, and punishment are rational responses only to beings capable of experiencing conscious states like pride and shame.

Further, it is hard to make sense of the idea that a non-conscious thing might *freely* choose its behavior. It is reasonable to think that there are only two possible explanations for the behavior of any non-conscious thing: its behavior will either be (1) purely random in the sense of being arbitrary and lacking any causal antecedents or (2) fully determined (and explainable) in terms of the mechanistic interactions of mereological simples.

If consciousness is a conceptual or moral prerequisite for moral agency, then an artificial agent would have to be conscious in order to be a moral agent. This, of course, means that we would have to be in a position to determine whether an artificial agent is conscious – and philosophers of mind disagree about whether it is even possible for an artificial ICT (I suppose we are an example of a

¹⁸ Of course, this doesn't apply to self-imposed capacities that are the result of morally problematic choices. Someone who voluntarily drinks herself into a stupor is accountable for the damage she does while intoxicated.

natural ICT) to be conscious. Some philosophers believe that only beings that are biologically alive are conscious, while others believe that any entity with a brain that is as complex as ours will produce consciousness regardless of the materials of which that brain is composed.¹⁹ In any event, if the foregoing analysis is correct, only conscious artificial agents could be moral agents.

References

Adam, A. (2001) "Gender and Computer Ethics" in Spinello, R.A. and Tavani, H., *Cyberethics*. Sudbury, MA: Jones and Bartlett

Alfino, M. and Pierce, L. (1997) *Information Ethics for Librarians*. Jefferson, NC: McFarland & Co.

Brey, P. (2001) "Disclosive Computer Ethics," in Spinello, R.A. and Tavani, H., *Cyberethics*. Sudbury, MA: Jones and Bartlett

Coleman, K. (2004) "Computing and Moral Responsibility," *Stanford Encyclopedia of Philosophy*; available at <http://plato.stanford.edu/entries/computing-responsibility/>

Eshleman, A. (2001) "Moral Responsibility," *Stanford Encyclopedia of Philosophy*; available at <http://plato.stanford.edu/entries/moral-responsibility/>

Floridi, L. (1999) "Information Ethics: On the Philosophical Foundation of Computer Ethics," *Ethics and Information Technology*, vol. 1, no. 1

Floridi, L.. (2002) "On the Intrinsic Value of Information Objects and the Infosphere," *Ethics and Information Technology*, vol. 4, no. 4

Floridi L.. and Sanders J.

Gert, B. (1999) "Common Morality and Computing," *Ethics and Information Technology*, vol. 1, no. 1

Gorniak-Kocikowska, K. (1996) "The Computer Revolution and the Problem of Global Ethics," in T.W. Bynum and S. Rogerson (eds.), *Science and Engineering Ethics*, vol. 2, no. 2

Himma, K.E. (2004a) "There's Something about Mary: The Moral Value of Things *qua* Information Objects," *Ethics and Information Technology*, vol. 6, no.3

¹⁹ It is worth noting here that there are also serious epistemic difficulties involved in our determining whether something is conscious. Since we have direct access to only our own consciousness, knowledge of other minds is necessarily indirect. We can infer that something is conscious only on the strength of analogical similarities (e.g., it has the same kind of brain I do and behaves the same way in response to certain kind of stimuli). But philosophers of mind have shown that such analogical similarities may not be sufficient to justify thinking someone is conscious. In essence, someone who infers that X is conscious based on X's similarity to him is illegitimately generalizing on the strength of just one observed case – one's own. Again, I can directly observe the consciousness of only one being, myself; and in no other context is an inductive argument sufficiently grounded in one observed case. The further from our own case some entity is, the more difficult it is for us to be justified in thinking it is conscious.

Himma, K.E. (2004b) "The Moral Significance of the Interest in Information: Reflections on a Fundamental Right to Information." *Journal of Information, Communication, and Ethics in Society*, 2

Himma, K.E. (2004c) "The Question at the Foundation of Information Ethics: Does Information have Intrinsic Value?" in *Challenges for the Citizen of the Information Society: Proceedings of the Seventh International Conference on the Social and Ethical Impacts of Information and Communications Technologies* (eds. T.W. Bynum, N. Pouloudi, S. Rogerson, and T. Spyrou), University of the Aegean, Syros, Greece

Himma, K.E. (2004d) "The Relationship between the Uniqueness of Computer Ethics and Its Independence as a Discipline in Applied Ethics," in Francis S. Grodzinsky, Richard A. Spinello, and Herman T. Tavani (eds.), *Proceedings for CEPE 2003 and the Sixth Annual Ethics and Technology Conference* (2003), 128-142; Reprinted with minor revisions in: *Ethics and Information Technology*, vol. 5, no. 4 (2004) (Special Issue: Computer Ethics in a Post-9/11 World), 225-237.

Johnson, D. G. (1999) *Computer Ethics in the 21st Century*, Proceedings of the 4th ETHICOMP 1999

Keulartz et.al (2004) "Pragmatism in Progress," *Techne: Journal of the Society for Philosophy and Technology*, vol. 7, no. 3

Latour, B. (1994) "On Technical Mediation – Philosophy, Sociology, Genealogy," *Common Knowledge*, vol. 3

Maner, W. (1996) "Unique Problems in Information Technology," *Science and Engineering Ethics*, vol. 2, no. 2; reprinted in D. Micah Hester and Paul J. Ford, *Computers and Ethics in the Cyberage* (Upper Saddle River, NJ: Prentice Hall, 2001).

Moor, J. (1998) "Reason, Relativity, and Responsibility in Computer Ethics," *Computers and Society*, vol. 28, no. 1

Tavani, H. (2002a) "The Uniqueness Debate in Computer Ethics: What Exactly is at Issue, and Why Does it Matter?" *Ethics and Information Technology*, vol. 4, no. 1

Tavani, H. (2002b) "Information Wants to Be Shared: An Alternative Framework for Analyzing Intellectual Property Disputes in the Information Age," *Catholic Library World*, vol. 73, no. 2

van den Hoven, J. (1997) "Computer Ethics and Moral Methodology," *Metaphilosophy*, vol. 28, no. 3

Weckert, J. (2000) "What is New or Unique about Internet Activity?" in Langford, D. (ed.), *Internet Ethics*. NY: St Martin's Press