Information Ethics: A Reappraisal

Luciano Floridi[1, 2]

[1]Research Chair in Philosophy of Information and GPI, University of Hertfordshire; [2]IEG and St Cross College, Oxford University.

Address for correspondence: School of Humanities, University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK; l.floridi@herts.ac.uk

Introduction

One of the highest honours that a research and scholarly community can bestow on its members is to pay sustained and careful attention to their work. In philosophy, this attention ultimately translates into constructive criticism. So, when the project for this special issue finally took shape, I was very flattered but I also expected a robust dose of disagreement. I have not been disappointed.

Criticism is in the very nature of any philosophical investigation. With its intrinsically open questions (Floridi [2004b]) philosophy invites dialogue in the form of objections and replies, suggestions for revisions and proposals for further improvements. The scientist may find this process of creative destruction unfamiliar: she might dislike it as at best fruitless, at worst counterproductive, in any case suspiciously symptomatic of a lack of clear criteria – through which progress might be assessed – and hard data, by which the same progress might be anchored and constrained. That scientist may not be in error about the facts, but she would certainly be mistaken about their interpretation, for she would be confusing the different directions in which philosophy and science move. Scientists build, whereas philosophers dig.

In building, one cannot help but establish every higher step on a lower step. It is trivial to remark that there is no second floor without a first and that the solidity of the construction heavily depends on the reliability of every layer. Trust and team-work is everything, getting things right vital. That is why scientific revolutions happen rarely but, when they do, they are as dangerous as major earthquakes.

In digging, on the other hand, every independent shovelful helps. So philosophers are more akin to individual explorers of the depth, and are more likely to proceed by removing rather than augmenting, reminding one of Michelangelo's definition of sculpture (the art of "taking away", as opposed to the art of "adding on" characteristic of painting). As we all

know, the higher one wishes to build, the more deeply (or better: profoundly) one needs to explore. Yet philosophers not only search for the deepest and hardest grounds on which our understanding may rest more safely; they also – or perhaps I should say mainly, in these anti-foundationalist days – seek to extract precious conceptual resources that, once unearthed, purified and carefully processed, may help humanity to make sense of an ever changing reality. We do not pass slabs, like collaborative Wittgensteinian constructors, we go back into the darkness of Plato's cave to help ourselves and others. Our hands are hardened and dirty. Forget about Athena, our god is Hephaestus.

The result is that, in the philosophical underground, where everything is so dimly lit and hidden, one often hears other explorers shouting curses, advices and warnings. These are not sterile invitations; these are signs of passion and interest in the work being done.

It is with this general picture of a collaborative enterprise in mind that I have tried to address the criticisms and suggestions elaborated in this collection of essays on my work on Information Ethics (IE). I believe IE to provide both a reliable foundation for a wider ethical discourse and to offer valuable resources to make sense of our time and the information revolution that characterises it. But I am also aware that it is a field that we have only started to explore (Floridi [2008b]).

Charles Ess has done an admirable work in summarising and framing the essays, and I am deeply indebted to his masterly handling of so many threads and to his insightful advices. I doubt his synthesis could be improved. I owe him more bottles of wine than I can remember. So if you see us tipsy at the next conference, you know why.

My contribution has been to address, in each essay, what I thought was the most salient points been made. But I have also tried to avoid repetitions, so the reader may find that some recurrent issues are addressed only once, where they play a more central role. I have also pointed out in a few cases possible internal references to other articles, but only when I thought it was really useful.

I am aware that, in many cases, I failed to do justice to the value of the theoretical analyses offered by so many colleagues from whom I have learnt so much. To my justification, I may point out that IE is a very rich and still largely unexploited mine. So the reader who may find some of the essays or my replies less than satisfactory is very welcome to join us. The advice to the graduate student is that there is plenty of rewarding work to be done.

Reply to Stahl

In his article, Bernd Carsten Stahl states that his "main concern is the question of universality of IE", while his "main contention is that Floridi's universality claim [concerning IE] is not always clear and possibly even contradictory". Although he discusses several other important topics and touches upon many interesting issues, I agree with Stahl that the aforementioned contention and concern are the most important and serious. Luckily, one can also address them together. For Stahl's concern and contention go hand in hand, or at least Stahl offers no reason to believe that they do not, which means that it will be sufficient to dispose of the latter (the contention) to appease the former (the concern) as well.

The contention appears to be based on at least two points. Both are in need of clarification. The first regards the misconception that treating all entities as informational objects, including human beings, means somehow diminishing our "human dignity". This is not an unusual complain, but it is misaddressed, somewhat outdated and definitely unproductive. I still recall one conference in the nineties when a famous computer ethicist compared me to a sort of Nazi, who wished to reduce humans to numbers, pointing out that Nazi used to tattoo six-digit identity tags on the left arms of the prisoners in their Lager. This is rhetorical nonsense. Suggesting that "surely we are not just...", is an old line that was already flawed when used against Darwin. Of course we are not just animals, but treating human beings as animals or, if you find the expression infelicitous, adopting the Level of Abstraction (LoA) at which one analyses humans as bipedal primates belonging to the mammalian species *Homo sapiens*, does not detract one iota from our nature. On the contrary, such *naturalization* makes it more likely that we might understand and appreciate a side of ourselves otherwise easily neglected. Now, in the same way as we are biological organisms, which share with their natural environment a long history of co-evolution, likewise, we can look at ourselves (change the LoA) as informational organisms, or *inforgs*. And this *informalization* also helps us to appreciate our nature and our relation to reality, now understood as the infosphere. IE adopts this informational ontology (or better: the corresponding LoA) as a minimal common denominator that unifies all entities. We are not just *inforgs*, but we should not fear to consider ourselves as inforgs. This is the first sense in which IE is universal. It is the inclusive sense that the logician will immediately recognise as part of the extensional meaning of a universal quantification: all entities are informational in nature, and IE seeks to address the ethical issues that pertain to all of them.

The second point concerns a mistaken view about how one may choose between different LoAs. It is not just a matter of whimsical preference, personal taste, or subjective inclination of the moment. The reader working in computer science knows already too well that one should never underestimate a crucial component in any use of a LoA, namely its goal or the "what for?" question. There is a perfectly reasonable LoA, say in terms of shape and topology, at which dad's shoes can be observed and even used as ships; but when Columbus grows up, he will find that ludic LoA useless for the purpose of reaching America. LoAs are teleological, or goal-oriented. Thus, when observing a building, which LoA one should adopt – architectural, emotional, financial, historical, legal, and so forth – depends on the goal of the analysis. There is not "right" LoA independently of the purpose for which it is adopted, in the same sense in which there is no right tool independently of the job that needs to be done. So, the position held by IE is that, when it comes to talk about ethical issues in an ontocentric and more inclusive, non-anthropocentric way, an informational LoA does a good job. This is the real thesis that one may wish to criticise. Unfortunately, it is overlooked in the article. The two alternatives proposed there fail to grasp the point. In EI, it is not a matter of showing that "the choice of the right LoA is an ethical imperative in itself", for goals can often be reached equally well in different ways: a hammer and a shoe can both be used to nail a painting to the wall. Nor is "the infosphere the highest LoA", for LoAs rarely come ordered in hierarchies, as one is easily reminded by the previous example about the house (asking which LoA is the highest would be missing the point). Indeed, Plato, Berkeley, Spinoza, or perhaps, I dare say, even Heidegger or Buddhist philosophy, all adopt equally abstract LoAs, just to mention a wide selection of different positions. So this is the second, non-relativistic sense in which IE claims to be universal: its analysis is based on the reasonable choice of a plausible and fruitful approach to the sort of new ethical problems emerging in the information society. Of course, one may disagree on the value of the approach. But the charges of relativism (any LoA is a good LoA) and absolutism (there is only one right LoA, the highest) could not be more misplaced.

At this point, it is worth asking what exactly is being universalised when we request an ethics to be universal. Here is what one may mean.

1) Universality as *universal applicability* of ethics to all entities concerned by the moral discourse. This is one of the senses in which Stahl speaks of universality as a matter of "validity" or "scope, range or applicability". We have seen not only that IE satisfies it but also that it does so better than many other theories, since it endorses a wider scope of ethical concerns.

2) Universality as *universal impartiality* of ethics. This second sense is strictly related to (1), insofar as it qualifies the sort of universal applicability in question (the applicability is *impartial*). In this case too, IE can only be said to be more impartial than many other theories which, for example, discriminate between living and non-living beings, or human/rational and non-human/non-rational agents.

3) Universality as *universal acceptability* of ethics by everyone involved, who shares it without coercion. This might be what Stahl means when he refers to "universal values", shared by all, and to "the ethical theory's strategy of convincing agents to accept or to follow it". This third sense requires some disambiguation. Insofar as it is a matter of *empirical description*, the most universal ethics in the first (applicability) and second (impartiality) sense above may still fail to be universally acceptable in the third (acceptability) sense if, for example, there exists a group of agents who is determined to reject it. In this sense, the existence of the Ku Klux Klan would undermine the universality of human rights. True, so this is not what can be at stake here. Insofar as it is a matter of *normative prescription* – an ethics ought to be universally acceptable by anyone without coercion – then (3) reduces to a combination of (1) and (2). In other words, it is because a theory is universally applicable and impartial that it might rightly aspire to gain uncoerced acceptance. But then it follows that any theory that satisfies (1) and (2) is strategically well-positioned (or at least as well positioned as any other) to satisfy (3) as well. And we have seen that IE does satisfy (1) and (2), so IE is not challenged by (3), or at least no more than other macroethics.

4) Universality as *universal inclusivity* of ethics. This holistic sense is different from (1) or (2) insofar as it refers to the capacity of a theory not only of applying to the specific agents and patients involved in a particular moral action, but also of widening its consideration to ever larger circles of interested parties as stakeholders, who may be taken into account when evaluating a moral action. Now, if any theory seeks to be inclusive, this is certainly IE, which I have often described as an extension of environmental ethics that looks at a wider and more inclusive environment. Indeed, critics have moved the objection that IE runs the risk of being too *inclusive*.

5) Universality as *anti-relativism* of ethics. This last sense has already been discussed above, where we have seen how IE succeeds in being non-relativistic without falling into the trap of being authoritarian or absolute.

To summarise, in any of the five senses in which an ethics might be requested to be universal, IE turns out to be in a rather satisfactory position. Indeed, one should acknowledge that it performs better than many other theories. Stahl's contention is ungrounded:

universality claims concerning IE are neither unclear nor contradictory. And without contention there is no ground for concern either.

Let me now add two corrections before concluding. First, when defining the Philosophy of Information (PI, Floridi [2003]), the term "information" should *not* be used only in a *semantic* sense (Floridi [2005]), contrary to what is stated in the article. PI and IE are *also* concerned with ontological issues that refer to informational entities and the infosphere, not to semantic content only. This misunderstanding is particularly damaging insofar as the article seems to be based on the idea that the concept of informational entity is used in the sense that "information is a description of something". Second, it is imprecise and confusing to say that a Level of Abstraction "requires the agent to follow ontological commitments", since a LoA rather expresses (often only implicitly) a specific ontological commitment, which the agent may or may not endorse.

The article by Bernd Carsten Stahl is a detailed and insightful analysis of IE, and the comparison between IE and Discourse Ethics is both innovative and remarkably enlightening. It is my conviction that more work along this comparative line might easily prove very fruitful in the future.

Reply to Brey

Phil Brey's article concerns IE's defence of the intrinsic value of informational entities or, more precisely, of Being in all its varieties and manifestations, understood in terms of an informational ontology. Brey acknowledges that many features of IE are interesting and even shareable. This positive assessment enables him to grasp IE better than other, less insightful, interpreters. Indeed, his solid understanding of IE is such that it leads him close to endorsing the whole approach. Only "close", however, because, as Brey himself acknowledges, four difficulties still prevent him from accepting IE in full. This is regrettable, given that such difficulties are based on a few simple confusions and mistakes that are easily fixable, as we shall see. So it is to be hoped that, once they are removed, Brey's acceptance of IE will no longer be conditional or qualified.

First difficulty. According to Brey "the first problem with the argument lies in Floridi's inference from the fact that some objects, like rocks and objects of cultural heritage deserve respect that they therefore have intrinsic value". Brey is right, yet not for the reason he believes himself to be right. True, the reasoning he sketches is blatantly fallacious, so anyone adopting it cannot but fail to be convincing. The trouble is that I would never endorse that reasoning in the first place. Brey confuses a *causal* with an *inferential* reasoning. A quick analogy will help. Suppose you tell me that your car does not start *because* its battery is flat. Next, I take you as saying that *if* your car does not start *then* its battery is flat. I then proceed to show you that you are wrong, for it takes a second to realise that your car may not start for a thousand other reasons (no petrol, for example). Of course, you are not impressed, but complain that I misconstrued what you meant: you said "because" but I attributed you an "if... then..." explanation. Now, let's go back to IE. The actual argument seeks to establish that entities deserve respect *because* they have intrinsic value, not that *if* entities deserve respect *then* they have intrinsic value. The latter inference is simply untenable, as Brey easily shows, but it is also irrelevant, as anyone may appreciate in the analogy of the car with the flat battery. It is the causal explanation that is at stake and that (not the fallacious inference formulated by Brey) leads to the really interesting problem: do non-sentient entities have some minimal, easily overridable but still intrinsic value? Without rehearsing the whole discussion, I agree that the answer here can be difficult to grasp. For it consists in shifting the burden of proof by asking, from a patient-oriented perspective, whether there is anything in the universe that is intrinsically and positively worthless ethically and hence rightly disrespectable in *this* particular sense, i.e., insofar as its intrinsic value is concerned (again,

something might deserve to be disrespected for other reasons, e.g., instrumentally or symbolically, as I have repeatedly clarified in the past). In short, one line of reasoning in favour of IE's position – the only one discussed by Brey, although there are others[1] – is that, because we lack arguments against the intrinsic value of Being in all its manifestations, we are led to expand an environmental approach to all non-sentient beings. The injunction is to treat something as intrinsically valuable and hence worthy of moral respect by default, until "proven guilty".

Brey seems to think that "if we are to start valuing things as intrinsically valuable that we do not already value as such, we need good reasons to do so. Since people do not normally seem to assign intrinsic value to information objects, Floridi needs to provide strong arguments for us to start valuing them as such". This is a dangerous error, for several reasons. First, what people normally assign intrinsic value to is a matter of sociology (*description*) not of ethics (*prescription*) and moving from one to the other means committing an obvious naturalistic fallacy (how the latter is avoided by IE is convincingly explained by Hongladarom's article in this issue). Second, history is full of "people" who failed to assign intrinsic value to at least some human beings (e.g. slaves, women, homosexuals, black, foreigners, Jews, handicapped, children, Indians, immigrants..., the menu is rich in choices), but it would be odd to argue that they are (or even were) right until proved wrong. Third, *Vox populi vox Dei* (literally: the voice of the people is God's voice) has never been a decent argument but, if one really likes to stick to the alleged "wisdom of crowds", why not choose the "right crowd"? For example, several philosophical schools, as well as many Buddhist, Christian, Hindu, Taoist or Shinto cultures, attribute intrinsic value both to sentient and to non-sentient realities. Finally, the logical mistake is with the initial argument itself. For its rationale is a conservative and cautious attitude that might be fine, when talking about potential moral risks of evil, but is out of place when engaging with the morally good. Consider its form: if we start *x*-ing things as intrinsically *x*-able that we do not already *x* as such, we need good reasons to do so. Now, it makes a big and quite obvious difference whether we replace *x* with negative (hate, destroy, despise, discriminate etc.) or positive attitudes (love, admire, cherish, protect, etc.). For example, "if we start hating things as intrinsically hateable that we do not already hate as such, we need good reasons to do so" might sound reasonable. But "if we start loving things as intrinsically lovable that we do not already love as such, we need good reasons to do so" is definitely questionable, for love does

[1] See for example the argument, based on the concept of "ontic trust", in Floridi [2007].

not bear very much accountancy. In other words, we should not fear to respect the universe too much. Rather, as Augustine nicely put it, *dilige, at quod vis fac* (love/respect and do what you wish).[2]

Second difficulty. Brey wonders "why should the correct account of intrinsic value be a general, minimalist, homogenous account [...]?". The "because" is in the pudding. Less metaphorically and more explicitly, we encounter here a twofold confusion. First, there is no "the correct account". This approach belongs to a non-pluralist and hence inevitably intolerant way of doing ethics that IE seeks to overcome. There are, however, "correct accounts" that may complement and reinforce each other, like stones in an arch. The second confusion concerns precisely what makes them "correct". Suppose someone says that he is a good driver. Although one might require him to produce a driving license, nobody would demand a syllogism. If pushed, one would eventually test the person's skills by having him actually driving a car. The reader acquainted with Wittgenstein's distinction between *saying* and *showing* will find this familiar. Now, IE tries to *show in practice* that there is a way of conceptualising Being informationally in such a way as to build a minimalist and homogenous account of all entities (Floridi [2004a], Floridi [2008a], Floridi [forthcoming]). IE also tries to *show in practice* that this is *a* correct account. Brey himself acknowledges this much at the beginning of the article. So he provides all the resources to answer his own question.

Third difficulty. Brey claims that "IE is committed to an untenable egalitarianism in the valuation of information objects. [...] from the point of view of IE, a work of Shakespeare is as valuable as a piece of pulp fiction, and a human being as valuable as a vat of toxic waste. Floridi will no doubt reply that differentiation is possible because some objects have additional worth beyond their status as information objects. But note that any such sources of additional worth lie beyond the scope of IE, because IE only assigns worth to things *qua* information objects. IE tells us that we should be equally protective of human beings and vats of toxic waste, or of any other information object, and that we have an (albeit overridable) duty to contribute to the improvement and flourishing of pieces of lint and human excrement. At best, this suggests that IE gives us very little guidance in making moral choices. At worst, it suggests that IE gives us the wrong kind of guidance."

[2] Augustine, "Homily on the First Epistle of St John", Eng. tr. available in Augustine [1984]. Note that Augustine uses the Latin "diligere", not "amare", a term that precisely refers to love as careful respect. As well, this is in keeping with the emphasis in IE highlighted in this issue by Hongladarom – that IE's informational ontocentrism is a naturalistic philosophy that closely resonates with Spinoza, Plato, Confucius, and Buddhist thought (among others) in its affirmation of the intrinsic moral worth of the *cosmos* as such.

The trouble with this criticism is that it misses a crucial point, often repeated in the literature and well-stressed by other interpreters in this special issue: the position taken by IE is metaphysical. When defending the intrinsic value of all aspects of Being, understood as informational entities, the point at stake is not some daft idea about the intrinsic value of Shakespeare vs. Dan Brown, or chocolate vs. excrement. Brey should know better, for the summary of IE that he provides at the beginning of the article states this rather clearly. The actual issue is whether Goodness and Being (capitals meant) might be two sides of the same concept, as Evil and Non-being might be. Again, the reader sufficiently acquainted with the history of Western philosophy need not be told about the classic thinkers, including Plato, Aristotle, Plotinus, Augustine, Aquinas and Spinoza, who have elaborated and defended in various ways this fundamental equation. For Plato, for example, Goodness and Being are intimately connected. Plato's universe is value-ridden at its very roots: value is there from the start, not imposed upon it by a rather late-coming, new mammalian species of animals, as if before evolution had the chance of hitting upon *homo sapiens* the universe were a value-neutral reality, devoid of any moral worth. What the article fails to grasp is that, by and large, IE proposes the same line of reasoning, by updating it in terms of an informational ontology, whereby Being is understood informationally and Non-being in terms of entropy. Note that this is not a defence of IE but an explanation of why Brey's objection fails to apply. Although being in the company of Plato or Spinoza, for example, might be reassuring, it is not an insurance against being mistaken. What I am arguing is that, having missed the key point regarding the informational ontology supported by IE, the rest of the article runs into problems of its own making. Above all, it misses the minimalist approach defended by IE, which invites one to consider every entity ethically valuable in itself and deserving some moral respect *to begin with*, exactly in the same way as environmental ethics invites us to approach any form of life as worth preserving, *if possible*. Nobody would ever argue that this is equivalent to saying that a spider's and a human life are equally worthy of respect. Culling, for example, is an ethical duty in environmental ethics. And even in Buddhism, killing animals is a minor offence (*pāyantika*) compared to the much more serious offence (*pārājika*) represented by killing a human being. Likewise, in IE the destruction of entities might easily be not only inevitable but also mandatory. Again, IE is not about respecting a single grain of sand as much as one respects the whole earth full of life or other human beings. It is about placing the threshold below which something should be morally disrespectable in itself and rightly so. With a Cartesian analogy, Brey's mistake lies in thinking that, if one argues that all physical things are extended, then one is arguing that they are all of the same size. Of

course they are not, and nobody could reasonably argue that they are. To revert to IE, the view that all entities are *at least minimally and overridably* valuable in themselves should not be confused with the view that they all share the same value. Contrary to what Brey seems to think, the adverbs play a crucial role there. As for IE offering little guidance, Brey is contradicting himself. This would be equivalent to saying that, since environmental ethics is based on the value of life and of the absence of suffering, then it offers little help with real-world issues. The truth is exactly the opposite (see for example Tavani's and Burk's articles in this issue). Having some general, basic and robust principles in place helps enormously when it comes to dealing with complex, practical matters. So much so that, at the beginning of the article, Brey himself reports on the growing number of applications that are taking advantage of the conceptual frame provided by IE.

Fourth and last difficulty. Brey complains that "if information objects are to possess intrinsic value, they cannot be observer dependent, because for an object to possess intrinsic value it must possess one or more properties that bestow intrinsic value upon it, such as the property of being rational, being capable of suffering, or being an information object. Such properties have to be objective and inalienable properties of the object in question, not subjective or contingent ones, because otherwise the assigned value is (at best) extrinsic, that is, resulting from the attribution of contingent roles or subjective meanings to objects." What lies behind this is a conceptual confusion. When we adopt a Level of Abstraction at which we observe things in terms of their chemical composition, for example, this does not make water contingently and subjectively H_2O. The same applies to the adoption of an information-theoretical LoA. A LoA is an interface that takes advantage of the constraints and affordances offered by the system under observation, in view of a specific goal.[3] Once this is grasped, the second step is to realise that IE tries to move from a materialist ontology to an informational one. As stressed above, this is a matter of metaphysics, not science, so anyone suggesting that information science, as a science for describing the universe, is not on a par with chemistry or physics, commits a category mistake, using Ryle's terminology. Once the ground is cleared of all these confusions and errors, it is obvious that Object-Oriented Programming (OOP) is provided only as a means to help make sense of an informational ontology. Compare this to Plato's or Descartes' use of geometry to make sense of their metaphysical views. Nobody ever complained that the development of, say, non-Euclidean geometry or topology undermined their ontologies, for this is merely irrelevant.

[3] [Editor's note: the reader is encouraged to compare this point with previous discussion of LoA in Floridi's response to Stahl.]

There are many other questions raised by Brey's article that would be interesting to discuss. For example, his proposal – which, he claims, "could solve many of the problems with IE that were discussed" – would set us back several decades, by making us embrace a conservative, anthropocentric and strongly Western-oriented perspective that new trends in applied ethics, from bioethics to environmental ethics, from medical ethics to IE, have been trying to overcome for some time (see Adam's article in this collection). Or consider the distinction between moral and instrumental reasons for doing "the right thing", which is ignored in the article. In the US, for example, the death penalty is increasingly controversial not because people have come to grips with its immorality, but because they have come to realise that it is uneconomical, since in the American system it is far more expensive to execute someone than to jail that person for life (*The Economist* [Aug 30th 2007]). Yet, any increase in the unpopularity of the death penalty is certainly welcome; no matter how disappointing the explanation can be, ethically speaking. Likewise, one could motivate agents to act morally towards the whole infosphere instrumentally (perhaps using egoistic or economic arguments) without, for this reason, having to drop the ethical stance. But there is no more space to explore these issues.

In conclusion, I am afraid that Brey's article is a missed opportunity. Having started describing IE quite accurately, it then fails to show how the difficulties it raises could be resolved within the framework it just laid down. It reminds one of someone who, having well crafted the handle of what might be an excellent knife then forgets to attach it to a blade and ends up complaining that it does not cut. The handle is a good first step, but it is not its fault if things are not working according to plan. Without completing the job properly there can be no chances of success.

Reply to Grodzinsky, Miller and Wolf

The shortest way of summarising my comment on Grodzinsky's, Miller's and Wolf's article is that I agree entirely with it. In the rest of this reply, I only wish to qualify this agreement.

Let me first stress an aspect that is well highlighted in the article, but that it is often missed by other interpreters: "Designers have an increased burden of care in programming artificial agents that exhibit learning* and intentionality*". I find this absolutely correct, as I have been arguing for some time (Floridi and Sanders [2005]). We should come to terms with the fact that technologies in general (think for example of biochemical engineering, genetic technologies or nanotechnologies), and ICT in particular, place god-like (*demiurgic*, in Plato's terminology, see Floridi and Sanders [2005]) responsibilities on our shoulders, both positively, in terms of what we create and do, and negatively, in terms of what we fail to create and do (Floridi [2006]). The more likely it is that we may unleash extremely powerful artificial agents in our natural and synthetic environments, the more demanding our moral duties become to exercise care, foresight, prevention and even restraint. Consider the pace at which unmanned military weapons are being developed nowadays or software agents are becoming autonomous and ubiquitous. Perhaps certain kind of artefacts should never be built. In Septermber 2008, for example, *The Wall Street Journal* reported that Google News picked up an obscure reprint of a 2002 article about United Airlines' risk of bankruptcy. Although United Airlines had since recovered, there was no dateline, so Google News ran the story as current news. It was then distributed widely by other news aggregators and eventually became a headline on Bloomberg. This triggered automated trading programs and a devaluation of the airlines' stock from $12 to $3, evaporating 1.14 billion dollars in shareholder wealth, close to United's total market value. Later n the day, the stock recovered, but not entirely, and at the end of the day was trading at $9.62, a market cap of $300M less than before Google ran the story. Any more doubts about whether artificial agents may do evil?

This leads to a second important point of agreement. Many artificial agents (AA) are not moral agents, of course. But some are or rather could be, if built and allowed to develop. This is not sci-fi and that is why we need to be very careful. I am not sure that if one could identify a Level of Abstraction (LoA) at which the behaviour of an agent could be ascribed explicitly to its designer then one should conclude that that agent would not be a moral agent. For, if this were the case, one day neuroscience may force us to conclude that there are no moral agents at all, not even human, but only agents whose actions are morally loaded. Any

sort of Turing test is based on the assumption of a particular phenomenon as given (e.g. *human* intelligence, moral agency, or creativity) and then a *comparison* with another, possibly indistinguishable, unknown phenomenon (e.g. *artificial* intelligence, moral agency or creativity). It is not based on an attempt to define what something is in itself (e.g. intelligence, moral agency, creativity). But let us not be distracted by these nuances concerning the debate about determinism. Grodzinsky, Miller, Wolf and I all agree that there are clear and uncontroversial cases in which an AA may qualify as a moral agent (in the article, this is the one with modifiable T that exhibits learning* and intentionality*). They also argue that this "does *not* [emphasis in the original] relieve the designer of responsibility" and seem to believe that I would disagree. *Not* at all. Not only do I share the same view, but, as a matter of record, I have *never* attributed moral *responsibility* to AAs, having always been very careful to specify that, when moral AAs are in question, what counts is their moral *accountability*. This is not philosophical hair-splitting. Parents, for example, may still be *responsible* for the way in which their adult children behave, but they are certainly not *accountable*. They might be bitterly blamed, but they will not go to prison if their son, now in his thirties, turns out to be a serial killer. So I agree with the conclusions reached in the paper. As I have been arguing for some time, engineers will be *responsible* for what and how they design AA, even if they may not be *accountable*. The sooner we take this on board the better.

The paper provides a valuable exploration of some essential questions and more complex scenarios that arise when discussing artificial agents at a variety of LoAs. This is very welcome. Equally welcome, and indeed refreshing, is the firm grasp and fruitful use of the method of LoAs carefully exhibited throughout the article. Two aspects that perhaps could have received more attention, however, concern non-software based AAs, like companies, or hybrid agents, to which I have often tried to call attention in my writings; and the way in which our reflection on agency is being improved and expanded thanks to the presence of non-fictional artificial agents. Ethics is obsessed with religious beliefs, psychologistic introspection and a reliance on often ungrounded, idiosyncratic intuitions or very foggy ideas (intentionality being one of them). It is still largely centred on a stand-alone, Cartesian-like, ratiocinating, human individual, when the world has in fact moved towards hybrid (think of a driver + car + GPS), distributed, and multiagent systems (there is probably more "moral agency" occurring at the level of governments, NGOs, parties, groups, companies and so forth than in any individual life). A pinch of serious computer science and *rigorous* philosophy can provide a great counterbalance. After all, artificial agents tell us as much about ourselves as about our artefacts.

Reply to Johnson and Miller

It is never easy to comment on one's own caricature. One may recognise some loosely familiar traits and wish to be a sport, but the distortions are so many and coarse that the temptation to dismiss the sketch as merely amusing is great. So let me hasten to say that, in the case of Johnson and Miller's article, such temptation should be resisted. True caricatures are not supposed to engage with the real features of their subject either accurately or reliably. They are rather expected to poke some fun at them, by exaggerating or fabricating their shortcomings and peculiarities. They are drawn in black and white and admit no nuances. But the fact remains that they can also serve a serious purpose. Like magnifying glasses, they may succeed in calling our attention to what might be otherwise less noticeable details. It is in this constructive way that I suggest one may approach the article in question. But first, a warning. The reader interested in understanding what I actually hold and argue about artificial agents (and let me repeat here: these are not just computers or software artefacts, but also synthetic and hybrid agents such as a multinational, a government, or a tank with its crew and weapons) and their possible moral roles, as well as the true nature of the philosophical debate at stake, will do better by skipping this reply, forget about its target and read instead the insightful article by Alison Adam in this collection. And now, for something different.

Johnson and Miller tell us a simple story. Some bad guys, the notorious band of computational-modelling rogues, would have us believe that artificial agents can be moral agents. Fortunately, the league of good guys, the computers-in-society group, comes to our rescue and saves the day by reconceptualising these agents for what they really are, mere machines under the full control of humans. Although the plot is not gripping, I quite like it. It is easy to follow and its happy ending is heartening. Imagine then my disappointment when I realised that, in one of the first episodes, I was enlisted as one of the bad guys, then soon got a few idiotic ideas slapped on my back, and ended up hanged for them, rather too hastily, by the end of the script.

One foolish thesis I am attributed is that I, as one of "the Computational Modellers see computational modeling as the ultimate in the philosophical endeavour to capture reality". I strongly doubt that anyone might be so silly to believe that "computational models represent reality [sic]". I certainly do not. I am too much of a Kantian (Floridi [2008a] and Floridi [2008c] provide plenty of evidence, but see Hongladarom's paper in this issue) to

consider even the possibility of "capturing reality" a serious philosophical project, let alone to believe that it might be pursued by the very limited means of a computational approach.

The article continues claiming that I "seem to have confused, on the one hand, something being autonomous within a level of abstraction and, on the other hand, something being autonomous writ large". Not at all, but I might have failed to make the point clear, or the authors would not be so mistaken. The method of abstraction may seem difficult to grasp, but its main lesson is simple enough (Floridi [2008c]). When applied to moral agents, it can be fruitfully used in order to understand autonomy in terms of self-regulating agency. The interested reader might check how, in the incriminated paper, artificial agents are shown to be autonomous moral agents exactly and precisely at the established (and indeed required) level of abstraction. The real issue is whether a better analysis, i.e. a better level of abstraction, may be provided. Hand-waving or feet-stomping are not an alternative, while trying to explain a concept (autonomous agency) by referring to more obscure concepts (freedom, intentionality and so forth) is simply falling into the classic fallacy of *obscurum per obscurius*.

Further "important flaws", which the authors fail to realise are actually in the eye of the beholders, relate to the basic thesis, which I am supposed to hold, that computer systems model human behaviour in the way scientific models model natural systems. It gets worse. "Their [mine included] logic assumes that because operational systems have certain outcomes, the behavior is equivalent to human behavior producing the same outcomes. They fail to see that tasks can be achieved by very different means [...]". Honestly, not even the bad guys could be so obtuse. As for myself, trust me, I know that a dish-washer and I perform the same task and achieve the same goal through rather different means. I actually explained this simple distinction as textbook material in Floridi [1999], so allow me to declare myself "not guilty".

The list of accusations unravels without mercy. It includes "fail[ing] to see that whether or not the processes count as moral action is a matter of human convention. Moreover, they fail to recognize that the meanings that humans give to particular operations of a computer system are contingent". This is a much more dangerous form of relativism. I do not believe that we can simply agree about whether any process (such as raping a child) counts as a moral action. Ethical discourse normally starts from *the fact* that it is. I like that.

To put it shortly, the silly views I am attributed by Johnson and Miller are patently untenable, nobody could seriously defend them, there are no Computational Modellers waiting in the dark, and so the article has no trouble in refuting them. Its criticisms are

correct; they just fail to be relevant or interesting. Of all sins that an interpreter may commit, that of *ignoratio elenchi* (the fallacy of offering an argument that may in itself be valid, but that is irrelevant in that it fails to address the issue in question) is among the most embarrassing, for Aristotle quite rightly equated it to logical ignorance. Unfortunately, the article manages to commit several times. There is an Italian phrase that says that sinning is human, but persisting is diabolical. Let us now look at the more constructive side of the article.

On 12 October 2007, during a shooting exercise at the South African Army's Combat Training Centre, at Lohatlha, in the Northern Cape, a computerised Oerlikon 35mm MK5 anti-aircraft gun went out of control and killed nine soldiers, injuring another eleven. Initially, the National Defence Force suspected it might have been a software glitch, but in January 2008 the verdict was in favour of a mechanical problem combined with human negligence. I doubt this might be a question of "interpretive flexibility". Johnson and Miller like to argue that similar issues are not a matter of truth, that there are no right answers to the question whether computer systems are (or ever could be considered) moral agents: "there is no truth to be uncovered, no test that involves identifying whether a system meets or does not meet a set of criteria". This is postmodernism at its worst, a relativistic game that is fine to play in the ivory tower of academia, but should not be exported to real life. Because it is obviously untenable, it ends up fostering bad faith since we have seen that in the article, when convenient, references to the real and true nature of the artefacts in question (as mere machines) appear to play a decisive role. It is contradictory, for why worry so much about engineers building artificial agents that could be moral agents if it is really just a matter of interpretation? But ultimately, the real danger is that it seeks to block what Peirce defined as the road of enquiry, by means of a stumbling block of rhetoric and socio-political agendas. Claims such as "this group of scholars [the computers-in-society group] [...] importantly [...] has a stake in *not* [italics in the original] establishing computer systems as moral agents" not only are unreasonable (if artificial agents are moral agents then they are, whatever the agenda of a lobby of scholars might wish them to be) but also promote the sort of head-in-the-sand strategy (if I refuse to see it for long enough it will disappear) that has never worked in the past, and has always increased the amount of troubles left to resolve. Again, some bit of realism might help. Since August 2007, the US army has deployed to Iraq three armed robots known as SWORDS (Special Weapons Observation Remote reconnaissance Direct action system). These agents are armed with M249 machine guns. I would argue that we should consider very carefully whether they are or can ever become sources of actions which are

morally loaded (i.e. good or evil) and *accountable* (mind, not *responsible*) for foreseeable disasters. We should look at their design, set clear criteria that they should satisfy, investigate the *truth* about their actual specifications and safety measures and so forth, and perhaps conclude that they should have never been built in the first place, or deployed to that context, or that they ought to be dismantled as soon as possible. This and similar policies would help us build a better and safer environment. What I strongly doubt is that engaging in some "interpretive flexibility" exercise might be useful at all. Other phrases like "this group [the computers-in-society group] has a two-part agenda: show that technology is an important component of morality but also show that technology is under human control" may be good for military propaganda but rather hard to believe in real life. They foster a false sense of security, especially when you realise that, for example, US military robots ran 30,000 missions in 2006 in Iraq and that the so called "surge" involved the deployment of 3000 new robots.[4]

There are many important distinctions that the article fails to acknowledge or understand: between *responsibility* (human) and *accountability* (of whatever agent might be involved); between *pluralism* (there are many ways of analysing a phenomenon, and using the methodology of levels of abstraction helps to make them clear) and *relativism* (these ways are comparable and assessable as better or worse); between *critical and informed discussion* (which hopes to reach a rational consensus with the interlocutor) and *agenda-driven axe-grinding* (which seeks to convince the interlocutor at all costs, irrespective of any facts and regulative ideal of actual truth). The list is longer, the blurring of essential, philosophical distinctions is unhelpful.

The world is undergoing the fastest and most profound transformations ever experienced by humanity. Technology is driving them at a neck-breaking pace and it is constantly outpacing our reflection and deliberation. This is not a matter of techno-determinism – a mythical *bête noire* long exorcised and dead. It is a far more interesting and subtle issue. As Shakespeare wonderfully puts it in *Macbeth*, we must do our best to understand "which seeds will grow and which will not". It is important that collaborative work of the kind shown by this special issue may foster a dialogue among all stakeholders because we will reap what we sow, agents included.

[4] Source: *Wired*, http://blog.wired.com/defense/2007/08/3000-more-bomb-.html

Reply to Adam

If I were to synthesise my reply to Adam's paper in only one sentence, I would simply subscribe to the following quote: "de-centring the human may not involve over-centring non-living things". Once this key point is grasped, it becomes difficult to move objections against IE's basic tenets. Since Adam does a very good job in showing how this is the case, I shall limit myself to highlighting here three of the points she makes that I find most interesting.

First, the article convincingly argues that IE can be seen as part of a larger, environmental movement, which is leading ethics and the philosophy of technology away from entrenched anthropocentric dogmas and towards ontocentric, patient-oriented approaches. The latter offer the advantage of being able to include, in their ethical discourses, synthetic environments, distributed and multi-agent systems as well as hybrid agents in which the human component is only part of the overall system. As an ethics of things, understood as either patients or agents, IE can thus be placed in a larger scenario of converging views, which include Actor-Network Theory. I must confess that I had not thought about it, but that I am delighted to be in such a good company.

A second, interesting point, raised by Adam, is that IE might be in for some heated debates that, she suspects, will resemble the philosophical controversies over Artificial Intelligence and its feasibility, which were suffered during the second half of last century. In this case too, I am afraid she might be right. This would be unfortunate though. For we have learnt little from decades of subtle arguments on the Chinese Room, for example, and the fruitlessness of such arm-chair speculations should have taught us a robust lesson. We shall see, but I am not optimistic.

Finally, but equally importantly, the reader will notice that in Adam's article the emphasis is on artificial agents broadly conceived. This is a crucial point, but it often escapes some of the interpreters. An artificial agent need not be a piece of software or something resembling a robot. True, nowadays we are witnessing an explosion of interest in webbots and so called artificial companions. The interactive doll *Primo Puel*, produced by Bandai (interestingly the same producer of the Tamagotchi) has sold more than one million copies since 2000. But, I would insist, with Adam, that the real challenge comes from the symbiotic mixture of biological and artificial, natural and engineered features to be found in complex agents. It might be the simple case of a driver, her car and her GPS on a motorway, or the more complex case of an *M1 Abrams* tank with its mechanical, electronic, computational and weaponry devices and a crew of 4 people, or indeed a government, an international bank or

any of the Fortune 500 American global corporations. More and more commonly, moral actions are the result of complex interactions among distributed systems integrated on a scale vastly larger than the single human being. Globalization is also a measure of the size of the agents involved in moral decisions that crucially affect the life and future of millions of individuals and their environments. What we are discovering is that we need an augmented ethics for a theory of augmented moral agency. This is where some of the challenging problems are arising, in terms of distributed morality. It is to be hoped that IE will be able to contribute to tackle them.

Reply to Tavani

Tavani's analysis of the ontological theory of informational privacy is outstanding. The level of scholarship, clarity, acumen and understanding of even the most subtle implications of the theory is impressive. This does not mean that Tavani agrees on everything I say. On the contrary, in his article he raises two specific challenges. It will be worth addressing both in detail and at some length. Here, I can only try to sketch the sort of approach that seems to be most promising.

According to Tavani "Floridi's theory also needs to respond to two additional challenges before it can be viewed as an adequate privacy theory. For one thing, it must be able to differentiate informational privacy from mental (or psychological) privacy, which Floridi (1999) recognizes as a distinct type of privacy. Additionally, his theory must be able to distinguish between descriptive and normative aspects of informational privacy." Tavani is correct in demanding an explanation, so let me address each challenge separately.

Regarding the distinction between informational privacy and mental/psychological privacy, I agree that there is a sense in which it would be counterintuitive (if not plainly wrong) to collapse one onto the other. As Tavani correctly puts it, even if "I am my information, it should still be possible for me to distinguish between privacy invasions that violate me because of an (unjustified) loss of personal information and those that violate me because of the 'mental harm' that they can cause (irrespective of the kind of information about me that may or may not be gained by others)." Admittedly, in the writings referred to by Tavani I was not interested in discussing this critical distinction. So let me make amends. What Tavani sees as a risk of collapse is real, but I would urge that it can be avoided in the following way. On the one hand, mental/psychological privacy and informational privacy *in the ordinary sense of the word* are simply two distinct and separate phenomena. For example, a laptop belonging to the British Ministry of Defence was recently stolen, causing the loss of the personal details of 600,000 people (including passport, National Insurance numbers and bank details).[5] Here, we are talking about a colossal breach in informational privacy, which does not have an obvious counterpart in mental/psychological privacy. On the other hand, in the ontological sense of informational privacy that I have also defended (in the sense that "to be is to be an informational entity"), there is a continuum between the informational nature of an individual and his or her mental/psychological privacy, insofar as the personal identity of

[5] Source: BBC, http://news.bbc.co.uk/1/hi/uk/7197628.stm

an individual is also constituted by her mental life. Where one decides to draw a threshold, or which conceptual vocabulary might be more apt to capture the relevant nuances, is probably a matter of circumstantial agreement. In February 2008, for example, counter-terrorism officers in Britain secretly and illegally recorded conversations between a politician, Tooting MP Sadiq Khan, and a constituent he was visiting in jail. In this case, the breach of privacy may be qualified as informational, but also mental/psychological, depending on which aspect one sees as more relevant. The fact that an informational ontology may help us to understand an individual as constituted by her information is meant to contribute and be complementary to other approaches to e.g. physical or mental/psychological privacy. In this case too, I see IE not as providing *the* only right theory, but a minimalist, common framework that can support dialogue.

Regarding the second challenge, I agree that the distinction between a naturally private situation and a normatively private situation is not only sound but important to maintain in order to be able to distinguish between loss and violation. And I also agree that we could preserve it by following Tavani's suggestion of "[...] 'contextualizing the infosphere' in ways that would differentiate segments of it into contexts or 'situations' in which a certain kind of information entity (e.g., a 'social agent') may or may not have a reasonable expectation of privacy. In this scheme, some 'disruptions in the infosphere' that would qualify as a loss of privacy for one or more agents, might not also count as violations of privacy for those agents." The work that remains to be done, as in the case of Tavani's RALC framework, is to develop this approach into a full theory. It may not be easy, but it looks perfectly doable.

Reply to Capurro

I admit that Capurro's article leaves me more puzzled than enlightened. I suspect that there might be something to be learnt in his "continental" approach, and that his contribution to IE, in terms of a Heideggerian perspective, might be somehow enriching. But the distance between our conceptual vocabularies, philosophical methods and frames of reference has so far slowed me down. The following examples may clarify my difficulty to the reader.

Capurro believes that "Value is not a property of things, […] humans are *per se* invaluable". Now, as every schoolboy knows, if humans are things, and things have no value, then humans have no value. Is this what Capurro means when he calls humans invaluable? I doubt it, but then, either Capurro is just committing a plain logical fallacy, which seems unlikely, or he must mean something different from what we normally mean by "value", "property", "things", "humans" and "invaluable". For example, at some points, he clearly has in mind "economic value", but at some other points he silently shifts to "moral value". These are very different things indeed, and the ambiguity is confusing. A bit later, Capurro continues by claiming that "We humans, *qua* human, are evaluating beings (*ens aestimans*). We estimate the price of things no less than we esteem each other, learning thus to esteem ourselves." This is fairly intelligible and quite uncontroversial, even if running together economic-price-giving activities and self- or mutual esteem processes is bewildering (this is a bit like someone trying to find a connection between "banks" as institutions and "banks" as shelving elevations in the bed of a river: I am sure some metaphorical sense can be made of it, but I wonder whether it might not be better to rely on some Cartesian "clear and distinct" ideas). However, depending on how accurate and lucid one may wish to be (and you may expect a philosopher to wish to be reasonably demanding here) then, in terms of evaluating, animals too evaluate best strategies, peers, social interactions and group hierarchies, richer sources of food, higher rewards, better partners, weaker foes and so forth. In a way, all animal life is about evaluating and being naturally selected to evaluate well. So, in the intelligible and uncontroversial sense of Capurro's statement, there is nothing especially human about "evaluation", unless here too Capurro means something else; but it is this "else" that escapes me.

Things hardly improve in the course of the article. Suppose we managed to follow Capurro until now. I would expect anyone to agree that the fact that something is given an evaluation does not mean that it is given the *right* evaluation (or life would be much easier), that it should get an evaluation at all (Socrates *docet*: we often value what is worthless), or

that it would have no value without the evaluation, for the evaluation might be just a matter of value-recognition and not value-creation (imagine nobody valuing you, would you have no value?). Yet, Capurro ignores all these distinctions and loses me once again. Speaking of "mirroring the world as the common invaluable horizon that allows us to evaluate things" blatantly fail to help. It makes humans and the horizon equally "invaluable", together with some "invaluable source we bring into the play that is also the source for the determinate value we ascribe to the productive action we call labour", then "our own invaluable evaluating presence" and finally "human beings who function as invaluable mirrors". The trouble is that, among all these invaluable items, I truly have no idea of what an "invaluable horizon" might be, let alone an activity such as "mirroring the world". In the end, we are led to conclude that humans are invaluable mirrors of an invaluable horizon. But what does this mean? "The mutual interplay of the human interplay with the common world as an invaluable horizon" fails to enlighten the logically-minded reader, who will be baffled by the double interplay with an invaluable horizon. As anticipated, I am happy to admit my limitations.

Perhaps the conversation could be improved by a few, simple clarifications. I apologise to the reader, who has been following this issue of *Ethics and Information Technology* up here, for the inevitable repetitions. To bore you least, I shall be brief and schematic.

First, I do not hold a digital ontology or metaphysics.[6] The informational ontology I defend is an ontology of structures, structures are relations and relations are not the sort of things that can be meaningfully qualified as either analogue or digital ("being taller than", for example, is not analogue/continuous or digital/discrete, the distinction simply fails to apply). So, the following statement "Floridi develops a metaphysical foundation of information ethics [...] instead of considering the digital interpretation of beings (i.e., as metaphysically more primary), it aims at fixing the meaning of Being *within* the digital perspective as today's prevailing interpretation of Being" is not only completely off target but also widely so.

Second, the infosphere is Being considered informationally, as simple as that. "Esse est information", where here information is not a semantic but an ontological concept (imagine a structural pattern). The scientifically-minded reader may think of this ontology in terms of energy, matter and informational structures as being all interdefinable. So the citation below is badly mistaken and makes what follows it a wild goose chase: "It becomes clear from his argumentation that the 'infosphere' is conceived as ontologically different

[6] See Floridi [2004a], Floridi [2008a] and Floridi [forthcoming].

from the physical world. If this is the case, he is not arguing within the background of what I call digital metaphysics, i.e., the interpretation of all beings from a digital perspective. But is in another sense still metaphysics, as it clearly distinguishes between a physical or material and an immaterial world without making explicit the *question* of Being itself." Honestly, I most definitely do not conceive of "the 'infosphere' as a 'hyperreality' separated from, as phenomenologists say, the 'life world'". Note that I might still be deluded about something else, yet that specific delusion is not mine.

Third, it is disappointing to see Capurro, who is a fine scholar, failing to get a message I have been broadcasting widely and loudly since 1999 (see, for example, Hongladarom's article in this collection, which captures it very well): entropy in IE is not meant to refer to the thermodynamic concept nor to Shannon[1]s equivalent measure at all. It is a metaphysical term and means Non-Being, or Nothingness (the concepts are related, of course, but also have their distinct meaning and should not be confused). Metaphysical entropy is increased when Being, interpreted informationally, is annihilated or degraded. So, the four basic norms of IE discuss entropy understood as the destruction, impoverishment or vandalizing of Being. They do not "contradict not only the unavoidable phenomenon of, say ICT-generated waste and energy consumption, but also, for instance, deleting viruses, SPAM and all kind of 'non useful' information" because they are not about Shannon's entropy, despite Capurro's stubborn insistence. Frankly, who could be such a fool as to hold the views that Capurro rather uncharitably attributes me? I plead guilty to the use of terms, such as "entropy" and "information", which have many meanings. But, in my defence, I must say that I have clarified what I mean by them on many occasions.

To conclude, I entirely agree with what Capurro writes at the beginning of his article: "shifting from computer ethics to information ethics addresses not just the question of how far computers challenge the morality of our actions, but also the question of how far we – and I think Floridi interprets this 'we' as addressing not primarily computer professionals but all stakeholders – are challenged by what he calls the 'infosphere'". This is a good starting point, on which one can build further understanding. Towards the end of the article, Capurro quotes me again as saying that we need "an open and rational process of discussion about what needs to be done first" in our world. Since he agrees, we also share a common goal. All that remains to be stressed is that "rational" is not an optional qualification but plays a crucial role in how we may bridge the starting point with the goal.

Reply to Hongladarom

Hongladarom has succeeded in clarifying several aspects of IE, which I have only managed to outline in my writings. I could start with his insightful and correct intuition that "Nonetheless, in agreeing with Kant that moral worth is an intrinsic property of some thing (humans included), Floridi's position is much closer to that of Kant than he perhaps admits." Touché. But even more importantly, Hongladarom has captured the essence of IE, namely its informational ontocentrism, impeccably. For the ethical ontocentrism I have been defending is indeed a naturalistic philosophy, in line with some of the best expressions of Western and Eastern thought: "If we look closely at the ancient conceptions of ethics, such as Aristotle's and much of the ethical theories of the East (such as the Confucianist and the Buddhist), we find that these conceptions are more or less naturalist[ic]". Plato, of course, is another great defender of the intrinsic "Goodness of Being", and in *Genesis* we are told that the Biblical God not only creates the universe but also rejoices again and again at the sight of its intrinsic goodness.

As Hongladarom remarks, an ontocentric approach is often threatened with the naturalistic fallacy. This presupposes a value-empty or value-neutral reality, from which then not a single drop of morality could be squeezed, on pain of contradiction. The "no ought from is" principle, with its Humean roots, is perfectly fine. If Being (or reality or nature or indeed the infosphere) is interpreted as being entirely and absolutely devoid of any moral value – if it is simply meaningless to say that "to be is to be good" – then any moral value, any goodness, and the corresponding ethical orientations that we long for, must come from elsewhere. A drained and dry container cannot fill itself. But if the ontic source, from which we seek to draw some moral guidance, is not empty, if, following Plato and Spinoza for example, we acknowledge that Being and Goodness are intrinsically intertwined well before any metaphysical or ethical discourse attempts to rescind them, then trying to extract values and the corresponding moral lessons from Being becomes a very natural process. One may try to find guidance and inspiration in the life of the universe without committing any logical fallacy.

Now, it seems to me that IE has had the merit to revive if not establish this ontocentric perspective. I am happy to concede that perhaps it takes a spiritualistic form of naturalism to find the approach attractive. Any materialistic view of the world, like Hume's, will struggle with the possibility that Being might be morally pregnant and overflowing with Goodness. And any metaphysics *à la* Heidegger will be too anthropocentric, self-referential, nihilistic

and reluctant to de-centralise the human condition to be truly enlightening (to see this, the reader is invited to read Capurro's article in this collection, along with my response). This is why it is very fruitful to read Spinoza as Hongladarom does, in terms of a naturalistic philosopher closer to the Greeks.

I have learnt much from Hongladarom's interpretation of Spinoza as a precursor of IE. It seems to me that, although I had some intuitions about the convergence between Spinoza's and my views on the possibility of grounding the ethical discourse on an ontocentric and patient-oriented perspective, his article does a much better, more comprehensive and far more instructive job than any of my writings. I remain indebted. So let me conclude by endorsing Hongladarom's perfect choice of a wonderful quote from Spinoza: "Only insofar as men live according to the guidance of reason, must they always agree in nature" (Proposition 34, 35, Part IV). This is philosophy at its best.

Reply to Burk

Burk's discussion of the potential application of IE to information law is fascinating. It is also quite inspiring. For it shows how far we have progressed since the nineties, when IE begun to argue for a new, informational approach to agents, their actions and environment. Overall, I have learnt much from Burk's analysis and I agree with his fundamental invitation: more work needs to be done in IE. Although the direction is right, and the path travelled heartening, there is still quite a bit of road ahead. To paraphrase him, it is important that we look at some instances of data as representing exactly the sort of personal data that constitute an informational entity. "Legal doctrine in some instances proves sympathetic to such an assertion, but remains largely inchoate as to *which data might constitute a given information entity in a given instance*. Neither is information ethics, in its current state of development, entirely helpful in answering this critical question. While information ethics *holds some promise to bring coherence* to this area of the law, further work articulating a richer theory of information ethics will be necessary before it can do so [italics added]." The problem, in his view, is that "[...] information ethics leaves open this central and exceptionally difficult question as to what data is or ought to be considered the individual's data". Burk is right, but the trouble is that the question might have to be left at least partly open (and hence up to the court or judge to decide, case by case) for a good reason.

What Burk is demanding is a full ontology that will tell us, with certainty and precision, in a variety of disparate and possibly complex cases, when some specific data are (or fail to be) part of what constitute an individual (and mind that the individual in question need not be a *single* individual, it could easily be a married couple, a company, a team, a social group and so forth). Now, in formal or engineered cases (e.g. in set theory or in the car industry), this is achievable, as one can be fairly sure about what does and what does not constitute the class of rational numbers, or the specific Jaguar parked in the garage, for example. But even in everyday life, when we think we should know better because we are dealing with concrete and very well-known entities, the fuzziness and slippery nature of the boundary between what counts in or out bubbles up everywhere, and it reminds us of our epistemic limits, if not of the ontic vagueness of reality. On 6 September 2001, for example, the European Parliament adopted an initiative called "25 years' application of Community legislation for hill and mountain farming" in which it urged "the Commission to lay down an exact definition [of hill and mountain farming] based on the criteria of height (in metres),

slope, shortened growing seasons, and appropriate combinations of those criteria [...]".[7] That definition is still to be found. But then, if it is so hard to agree on an exact and uncontroversial understanding of what counts, for legislative and economic purposes, as hill and mountain farming, how much harder can it be to define all and only those data that constitute a person? The request for necessary and sufficient conditions is often natural but, equally often, needs to be resisted, if one wishes to be reasonable and accurate rather than precise and inflexible. Wiener famously described human beings as "patterns that perpetuate themselves".[8] I have argued for a very similar view in IE: we are homeostatic information patterns ("persistent information patterns" in the phrase that Burk attributes to Wiener). Yet, patterns tend to lack sharp or clear-cut edges and, being in constant dynamic evolution, they can easily be polymorphic. The waves on the beach are quite clearly individual waves, but any attempt to fix the precise drops of water that constitute each wave would be a pointless exercise. So, does all this mean that, although Burk is right, his demand is bound to remain unsatisfied? This would be overly pessimistic, for two reasons.

On the one hand, we might not have a definition, but we can rely on our intelligent understanding, as in the case of the waves. Certainly, borderline, complex or extreme cases, such as *C.B.C. Distribution and Marketing, Inc. v. Major League Baseball Mass Media*, may test our capacity of discernment, but then, IE should be praised for opening our eyes to these new perspectives and for providing the right approach to tackle them (essentially: humans are, or at least might also be considered and treated as, informational objects), rather than blamed for failing to deliver a secure route to a final verdict, a request which to some may appear supererogatory. Thus, an influential verdict by Germany's highest court in February 2008 on whether the government might have the right to check remotely a citizen's computer has, according to many commentators, established a new "fundamental right" for the 21st century, according to which a person's "private sphere" includes her computer, even when that person is online.[9] This is going in the direction pointed by IE.

On the other hand, and this is the alternative that Burk seems to favour and that I find equally appealing, it might be possible to enrich IE with some guidelines, which could bring some coherence to the law relevant to data representations. This is not the place to provide

[7] http://www.europarl.europa.eu/pv2/pv2?PRG=CALDOC&FILE=010906&LANGUE=EN&TPV=PROV&LASTCHAP=9&SDOCTA=22&TXTLST=1&Type_Doc=FIRST&POS=1

[8] This is the precise quotation from Wiener's *The Human Use of Human Beings*, which can be found on p. 96. The phrase that Burk attributes to Wiener is not to be found in the text and is best used to describe the ontology that I have bene advocating.

[9] See for example the report and comments on the Spiegel International Online website, **http://www.spiegel.de/international/germany/0,1518,538378,00.html** Ralf Bendrath' blog entry is enlightening: **http://bendrath.blogspot.com/2008/02/germany-new-basic-right-to-privacy-of.html**

them, but I suppose an example might be useful to illustrate the sort of research that one may wish to see developed in the close future. In brief, the task is to identify some criteria (recall the example of the definition of hill and mountain farming based on height, slope, and shortened growing seasons) that could help us in determining when some data are (or fail to be) personal data. Here is an initial proposal.

Let me first remind the reader about the concept of inverse function. To do so, a simple example will suffice. If the function is $f(x) = x^2$, and x ranges over the domain of positive natural numbers (that is, $x \geq 0$), then the inverse function is $f^{-1}(y) = \sqrt{y}$, so that, if we have $f(3) = 3^2 = 9$, then the inverse function is $\sqrt{9} = 3$. More precisely, if f is a function whose domain is the set X and whose range is the set Y, then the inverse of f is the function f^{-1} with domain Y and range X, defined by the following rule: if $f(x) = y$ then $f^{-1}(y) = x$. The obvious but powerful property that an inverse function enjoys is that of uniquely identifying the input x of another function based only on its output y, for all $y \in Y$. In plain English, a function leads you from x to y and an inverse function leads you back, from y to x. Not all functions have an inverse function. As we all know, things in life may or may not be reversible: you cannot un-break the eggs you mistakenly broke (no reverse function available here), but you can empty the bottle you mistakenly filled. The precise concept of reversibility, as a reverse relation that leads us back to where we were, is what we need. If some data are personal data, they are linked to the person to which they belong in a way similar to that in which y is related to x through a function f. But then, at least in theory, one might be able to move not only from the person's properties (say, Peter's) to the data (say, some credit card information), but also back, and uniquely identify the properties from the data. This backward route is often blocked on purpose. Peter might be a conservative who voted for a particular presidential candidate, but the electoral system makes sure it should not be possible to guess who the person behind that conservative vote is. Abstract or obliterate sufficient details (some data), and the output becomes irreversible. This is privacy through irreversibility. It follows that a possible way of answering Burk's request might be to develop criteria of personal data identification on the basis of their reversibility. Finger prints certainly seem to qualify rather well, so does DNA. What about the statistical data mentioned by Burk in his article? They might as well, it all depends on whether they can be reverse-engineered, as output, in order to obtain at least some information about the original input. Suppose they could. Then the verdict might have to be revised, with a twist though, as I shall explain in the conclusion of this reply.

There are slightly different versions of the following story, which seems to circulate among different cultures, but the reader will certainly capture its essence. A poor man has only a piece of bread, so he goes to a nearby rich house and holds it over the steaming pot, hoping to capture a bit of the flavour from the good-smelling vapour. Unfortunately, the rich owner of the house sees him, drags him in front of the judge, and asks him to pay for the smell. The judge hears the complaint. After a moment of reflection, he gives to the poor man a coin and asks him to throw it on the marble table as noisily as possible. The poor man obliges him, the judge regains his coin and dismisses the two men. When the rich man complains, the judge explains to him that the poor man has paid for the smell of the soup with the sound of the coin. Let us consider the problem analytically. The smell of the soup may or may not be constitutive of the soup itself. If it is not, then we have the same result reached in *C.B.C. Distribution and Marketing, Inc. v. Major League Baseball Mass Media*: smells are as public as players' statistics and anyone can take advantage of them. If it is, then "taking" it may be worth a reimbursement, even if that smell is "public". But then, the sound that a coin makes when hitting a table should be treated in the same way: it is equally public and may belong to the coin in a way comparable to the way in which the smell is related to the soup. The lesson is simple. Perhaps the U.S. Court might have hold that C.B.C. Distribution and Marketing, Inc. did violate the publicity rights of the players but then decided that the best way to settle the challenge was to allow the players, whose statistics are used, free access to any of the fantasy sports games offered by the company, where they could enjoy their virtual use (http://www.cdmsports.com/games.php).

References

Augustine 1984, *Selected Writings* (New York: Paulist Press).

Floridi, L. 1999, *Philosophy and Computing: An Introduction* (London, New York: Routledge).

Floridi, L. 2003, "What Is the Philosophy of Information?" in *Cyberphilosophy: The Intersection of Philosophy and Computing*, edited by James H. Moor and Terrel Ward Bynum (Oxford – New York: Blackwell),

Floridi, L. 2004a, "Informational Realism" in *Computers and Philosophy 2003 - Selected Papers from the Computer and Philosophy Conference (Cap 2003), Acs - Conferences in Research and Practice in Information Technology* edited by John Weckert and Yeslam Al-Saggaf (7-12).

Floridi, L. 2004b, "Open Problems in the Philosophy of Information", *Metaphilosophy*, 35(4), 554-582.

Floridi, L. 2005, "Information, Semantic Conceptions Of ", *Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = <**http://plato.stanford.edu/entries/information-semantic/>**.

Floridi, L. 2006, "Information Technologies and the Tragedy of the Good Will", *Ethics and Information Technology*, 8(4), 253-262.

Floridi, L. 2007, "Global Information Ethics: The Importance of Being Environmentally Earnest", *International Journal of Technology and Human Interaction*, 3(3), 1-11.

Floridi, L. 2008a, "A Defence of Informational Structural Realism", *Synthese*, 161(2), 219-253.

Floridi, L. 2008b, "Information Ethics: Its Nature and Scope" in *Moral Philosophy and Information Technology*, edited by Jeroen van den Hoven and John Weckert (Cambridge: Cambridge University Press), 40-65.

Floridi, L. 2008c, "The Method of Levels of Abstraction", *Minds and Machines*, 18(3), 303-329.

Floridi, L. forthcoming, "Against Digital Ontology", *Synthese*.

Floridi, L., and Sanders, J. W. 2005, "Internet Ethics: The Constructionist Values of Homo Poieticus" in *The Impact of the Internet on Our Moral Lives*, edited by Robert Cavalier (New York: SUNY),

The Economist Aug 30th 2007, "Revenge Begins to Seem Less Sweet".