

# The Philosophy of Information as a Conceptual Framework

Luciano Floridi

**Abstract** The article contains the replies to the collection of contributions discussing my research on the philosophy of information.

**Keywords** Information and communication technologies (ICT) · Information ethics (IE) · Level of abstraction (LoA) · Philosophy of information (PI) · Semantic information

## 1 Introduction

I benefitted much from studying this collection of essays. In many cases, they made me appreciate implications and consequences of my own work which I had not seen at all, or only partially. The whole collection shows that, despite the criticisms, the philosophy of information provides a very fruitful conceptual framework within which new philosophical issues can be addressed and old ones revived. As the reader will see, in many cases fruitful disagreements take place against the background of considerable convergences on what count as interesting problems and valuable methods to tackle them. I am grateful to all contributors for their time and efforts and

---

L. Floridi  
Research Chair in Philosophy of Information and GPI, University of Hertfordshire,  
de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK

L. Floridi  
Faculty of Philosophy and IEG, University of Oxford, Oxford, UK

L. Floridi  
UNESCO Chair in Information and Computer Ethics, University of Hertfordshire,  
de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK

L. Floridi (✉)  
Department of Philosophy, University of Hertfordshire, de Havilland Campus, Hatfield,  
Hertfordshire AL10 9AB, UK  
e-mail: l.floridi@herts.ac.uk

for the rare privilege of seeing my research evaluated so thoroughly. Criticism, sometimes even more than praise, is often a clear sign of intellectual interest. In my replies, I have tried to address what seemed to me the main issues or contentions contained in each contribution, while providing an overall narrative. I hope I have not completely failed in both tasks. I am much indebted to Hilmi Demir, who patiently and carefully dealt with the editing of this special issue and my delays. Without his encouragement and commitment, I would not be writing these lines.

## 2 Reply to Gillies

I have gained many insights from Gillies in the past, especially about the philosophy of artificial intelligence (AI), and I can see that I still have much to learn. His elegant and thoughtful article opens a very interesting perspective on the variety of structuralist philosophy of mathematics that might be promoted following the sort of philosophy of information that I have been supporting (see now Floridi 2010b). Gillies' interpretation of my position as friendly towards a Platonist approach is correct, and his view that my implicit Platonism has a Popperian strand is indeed most perceptive. There is no reason why he should have browsed through an old book of mine, but in *Scepticism and the Foundation of Epistemology* (Floridi 1996, see especially Chapter 7), I discussed at length Popper's World 3 hypothesis as well as his arguments in favour of an "epistemology without a knowing subject". The latter especially was very influential in my understanding of what we now call a philosophy of information. It is easy to see this in my admittedly qualified support for Mark Notturmo's interpretation, according to which "Popper's concept of subjectless knowledge is a legitimate use of the word 'knowledge'. It corresponds to the sense of 'knowledge' as *information* or a branch of learning [my emphasis]" (Notturmo 1985, p. 153). In short, Gillies is absolutely spot on in his analysis of my work. There is more that we share. Both of us are mathematical realists, like Frege was, even if our Greek alliances differ. As he writes

I am completely in agreement with Floridi that we should regard information as real, and that philosophers should try to elaborate a theory of informational realism. However, I differ from him in preferring a different version of informational realism. This version, as we shall see, is more Aristotelian than Platonic.

I shall not try to convince the reader that Plato is better than Aristotle, though you might guess that I prefer to sit with Whitehead when it comes to interpreting Western philosophy. What I shall try to do instead is to highlight two concepts—*structure* and *interaction*—which I believe should play a key role in an informational-theoretic philosophy of mathematical structuralism, independently of whether you prefer Gillies' Aristotelianism to my Platonism.

Structures have been increasingly important in the history and the philosophy of mathematics. Understood as systematic patterns of differences—like a black circle on a white surface, the set of all points whose Cartesian coordinates satisfy the equation  $x^2 + y^2 = 2^2$ , thus identifying a circle of radius 2, or the set of natural numbers—structures shift our ontological focus, from the substantial nature of mathematical entities as *things*, to their underlining relational essence as *patterns*.

This is a crucial lesson I believe I learnt from Cassirer and his masterly essay dedicated to the replacement of the concept of substance by that of function in the philosophy of mathematics (1910, now Cassirer 1953). Just a decade before its publication, Hilbert had elegantly and famously explained such structuralism at the beginning of his *Grundlagen der Geometrie* (1899, see now Hilbert 1971):

We think of these points, straight lines, and planes [in Geometry] as having certain mutual relations, which we indicate by means of such words as ‘are situated,’ ‘between,’ ‘parallel,’ ‘congruent,’ ‘continuous,’ etc. The complete and exact description of these relations follows as a consequence of the axioms of geometry. These axioms may be arranged in five groups. Each of these groups expresses, by itself, certain related fundamental facts of our intuition.

Replace in the quotation “as *having* certain mutual relations” with “as *being constituted* by certain mutual relations”, and you obtain a structuralist ontology of geometrical elements. A similar analysis can be provided for arithmetical structures. Mathematical as well as physical *objects* are what we, as specific informational structures (cognitive systems, in an equivalent vocabulary), find easier to handle logically, mentally and empirically. However, there are good philosophical, mathematical and scientific reasons (Florida 2008b, 2010b; Shapiro 2000; French and Ladyman 2003) to commit ourselves to a structuralist/informationalist ontology according to which *mathematical* as well as *physical* objects do not play the role of ultimate *realia*. Like icons on a computer screen, they are user-friendly, but they should not be confused with what lies behind them and constitutes their nature. Recall that we are, to the best of our knowledge, the only semantically structuring structures in the infosphere (see my reply to Durante). As such, we give and make sense of what we experience by objectifying it. The distance from the world that makes a rich cognitive life possible is also the price imposed by the reification of the world. We freeze changes into state- or phase-transitions and modular events and transform patterns and structures into objects and properties, finally privileging a naïve ontology of sufficiently permanent things and qualities. How many times have you heard philosophers using the expression “the furniture of the world”? We deal with Being by forgetting the -ing part. This is fine, as long as the task is to survive and reproduce in a hostile environment, which promotes fast reactions and punishes slow reflection, but it is much less satisfactory if the goal is to understand the ultimate nature of reality philosophically. From this unifying perspective, it is reasonable to presume that we are always dealing with dynamic structures, whether empirically or mathematically reified as objects, so mathematical structures are not second-class citizens of our ontology, with first-class citizens represented by material things, the usual chairs and horses, kicked stones and white snow, co-referential stars, chariots and trolleys, hammers and thalers and so forth.

The sketch I have just provided is in line with what Shapiro (1997) has defined and defended as the *ante rem* approach to mathematical structuralism. The difference is that the *whole ontology* is *ante rem*, not just the ontology of mathematical objects. At the same time, Gillies’ position is consistent with Shapiro’s characterization of Aristotelian, *in re* mathematical structuralism. In both cases, an informational approach to structuralism helps to prevent (rather than solve) the problem of why mathematics is so readily applicable to the empirical world: it is a question of structures all the way down.

Let me now turn to the concept of interaction. Mathematical structures interplay and interlock with each other in various ways. Such *internal* interactions (Hilbert's more static "certain mutual relations") are what an informational structural realist qualifies as primarily necessary, not the mathematical objects that they end up constituting. Yet interactive is also our relationship with such mathematical structures. As Gillies rightly remarks, I have argued that such *external interactions* give us a metaphysical *criterion* of existence (a yardstick to check whether  $x$  exists, not what it means for  $x$  to exist, or what kind of existence  $x$  enjoys). To put it simply, ghosts do not exist not because you cannot touch them, but because there is no Level of Abstraction at which you can interact with them. That is why subatomic particles exist and so do non-Euclidean triangles, even if you cannot touch or smell them. True, *whether* and *how* we can interact with certain mathematical structures depends on the adopted Level of Abstraction (LoA). You may not consider functions of the kind  $x^y$  among your observables, or once they are admitted by your LoA, you may work with them differently, depending on whether you make the  $x$  and the  $y$  range only on the natural numbers (so that, for example, you obtain  $3^2=9$ ) or on the integers as well (so that  $3^{-2}=1/3^2$ ). This is where Gillies and I are still in agreement. But once the observables are admitted and specified, their dynamics is a matter of acknowledgement and discovery, not construction or invention. This is where I take a more Platonist route.

Let me close my comment with an anecdote. One day, when I was a boy, my mother found me distressed because I had just read about the death of Porthos, one of the four musketeers, in Dumas' *The Man in the Iron Mask* (1847–1850, now Dumas 2008). There was not much she could do, but she tried to console me by pointing out that I was not alone: Dumas himself had been deeply upset by the death of his character. She recounted the story according to which Dumas cried for days because Porthos was dead. This seemed to me perfectly reasonable, but what struck me at the time and made me forget my sadness was the alleged explanation that Dumas offered when asked why he, the omnipotent author, could not change the story of a literary character. He said that Porthos *had* to die. That seemed to me to express better than anything else I had experienced at the time the force of a theorem, the strength of logical coherence, the inescapable, constraining grip of structures, which have features and interact in ways that are utterly independent of our wills and desires. Dumas was powerless. Porthos' death was the inexorable conclusion, given the development of the plot. That evening I thanked my mother for the lesson in logical thinking and felt a bit less upset: necessity is somewhat soothing. Many years later, while writing this reply, I went to check the episode. It still saddens me. But I found a line that seems very appropriate to conclude this short comment. "'Parbleu!' said Porthos again, with laughter that he did not even attempt to restrain, 'when a thing is explained to me I understand it; begone, and give me the light!'"

### 3 Reply to Allo

The paper by Allo is most insightful and ingenious. Insightful because it correctly interprets the theory of strongly semantic information, the task it is supposed to discharge with respect to the Bar-Hillel-Carnal paradox and its Tarskian root: contradictions pose

an informational problem because of their truth value, not because of their inferential effects. Ingenious, because it convincingly shows how the veridicality thesis (for  $P$  to qualify as semantic information,  $P$  must be true) might be (made) compatible with dialetheism (there are true contradictions, so for some  $P$ , both  $P$  and  $\neg P$  are true). Since I am not a dialetheist myself, I will not need to follow Allo in that direction, but it is interesting to see what would happen if one day I were to change my philosophy of logic. Instead, what I would like to do, in the rest of this reply, is to expand on a specific point made by Allo in his paper.

Suppose we might be allowed to play the role of minor gods. We are not just observers of the world, but also makers of it. Let us call the world we are building the *system* and the output of the observation of such world its *model*. The system is completely transparent to us, who have engineered it, but rather opaque to its users. The world system is made of many other sub-systems. One of them is a vending machine: we, the gods, know exactly what goes on inside it, whereas the user can only rely on a simple interface (Level of Abstraction, or simply LoA) to understand and operate it. Could there be “contradictions” inside the system? If we are poor engineers or rather clumsy gods, the answer is obviously yes. This is not a matter of category mistake: if some parts  $A$  of the vending machine click and clang to the effect that state  $D$  arises (the machine dispenses the soft drink), while some other parts  $B$  of the same system simultaneously move and change to the effect that opposite state  $\neg D$  also arises (the machine does not dispense the soft drink) at the same time as  $D$ , then the physical clash—what conceptually, or in terms of design, is described as a logical inconsistency—between  $D$  and  $\neg D$  will arise and cause a mess: users not getting their drinks when they should or perhaps getting them randomly or getting more than they paid for. Clearly, actual contradictory processes or events or patterns or facts or features of entities (depending on one’s own ontology) may be presupposed to occur in the world. *Metaphysical dialetheism* is perfectly conceivable from a God’s (sorry, gods’) eye perspective. The observer, however, placed on this epistemological side (the mortal one) of the LoA, will experience the *malfunctioning* of the system—e.g. the delivery of no cans, or the delivery of more cans than she is entitled to—not the contradictory double-functioning (both  $A$  and  $B$ ) that is causing the malfunctioning, that is, she will not experience the machine dispensing and not dispensing the purchased drink at the same time and in the same sense. The example should clarify that wondering whether the world might contain contradictions means wondering whether it is feasible and sensible to adopt a LoA at which some of the states of the world-system are described by the resulting model as the upshot of contradictions intrinsic to the system.

Let us return to our vending machine. Suppose that it is working perfectly well but that the user develops an inconsistent model of it and its behaviour. She represents the machine as simultaneously dispensing and not dispensing the right drink for the same correct amount of money, in exactly the plain sense in which you and I understand such contradiction. In other words, her model states that both  $P$  and  $\neg P$  are true of the targeted system. If asked, she might explain why she gets the right drink every time she inserts the right coins as a case of amazing good luck. In this case, the contradiction is in the model, not in the system. Strictly speaking, as Allo correctly remarks: “there is no reason to assume that the constraints imposed by the world are such that only consistent models are possible”. It would be wonderful if the opposite

were the case. Unfortunately, *semantic dialetheism* is not only perfectly possible, but probably the norm, given how difficult it is to spot inconsistencies in our ways of handling the world informationally. This point is painfully driven home when we shift our perspective, and instead of talking of contradictory observing models—which requires some conceptual gymnastic, as we have just seen—we concentrate on contradictory *blueprints*, those models that specify the overall design of the systems to be built. Too often contradictory requirements and specifications lead to malfunctioning and even dangerous systems. We expend too much effort and resources debugging and trying to make sure that blueprints are at least internally consistent to have any doubts about the everyday phenomenon of semantic dialetheism.

Contradictory artefacts resulting from contradictory blueprints need not consist of physical bits or digital bytes. They might be social or intellectual constructs as well. Legislation, political institutions or social practices may easily be contradictory, for example. This not only reinforces the view that metaphysical dialetheism is far from being a mere speculation, but it also shows how metaphysical and semantic dialetheism, both viable as independent options, may be combined into a single form of *absolute dialetheism*. Hegel and Marx were keen on explaining contradictory (social, intellectual, political, economic, engineered and so forth) systems in terms of models capable of handling contradictory descriptions of them.

Does all this mean that, after all, I should change my philosophy of logic and embrace a more contradictory-friendly attitude when it comes to the nature of information? It has taken some elaboration, but explaining why the answer is in the negative was the point I said I wished to highlight in Allo's paper. On several occasions, Allo refers to pragmatic considerations lying behind the assumption of consistency (and hence the rejection of dialetheism), but he also seems to dismiss them as merely practical, while the method of levels of abstraction is interpreted exclusively in epistemological terms (roughly, the models are always descriptions, never blueprints). This is where I am less willing to follow Allo's approach. For example, he is right in highlighting

what the costs of an overall consistent approach are. In brief: we cannot appeal to the consistency of the world, and we should be aware of the overall cost of maintaining consistency.

But the pragmatic dimension of our construction of our models of the world is crucial and cannot be left out or underestimated as merely contextual, because it is there that the previous costs are counterbalanced by overwhelming benefits. We seek and value semantic information not just because we wish to pursue the epistemic goals of *description*, *explanation* and *prediction*, but also, if not probably mainly, because we want to interact with the world and build it or modify it successfully. It is the pragmatic dimension provided by the goals of *interaction*, *construction* and *modification* that grants to the virtue of consistency of our models a *normative*, *overriding* value that would otherwise make consistency appear to be on a par with other virtues (thus consistency and elegance, for example, are not on the same plane). It does so in two ways. *Interactively*, being told that one is and is not a vegetarian will not help pragmatically the chef who needs to provide an appropriate meal, exactly in the same way and sense in which being told that one is or is not a vegetarian leaves things unimproved informatively. Recall the Tarskian point above: in information

transactions, consistency is to be sought because without it there is no useful transaction, and not because any transactions will ensue. *Constructively*, a blueprint of a vending machine does not have to have the virtues of being elegant or simple or parsimonious, but it must be devoid of contradictory features, or the resulting artefact will malfunction. The same applies to a piece of legislation, a computer program or the safety system of a nuclear power station. Consistency is a necessary condition for well-functioning. This pragmatic feature cannot be overvalued. That is why in both cases (i.e. interactively and constructively), we should adopt levels of abstraction that generate consistent models.

To conclude, there are three informational contexts where consistency plays a crucial role: in (what we take it to be) the nature of the world (the system), in how we describe it (the model) and in how we go about engaging with it (the praxis). The difference that information makes is not only epistemic (affecting only the system and its model), it is also, if not primarily, pragmatic (affecting also the praxis) and consistency serves both masters. This is “the separate reason to reject expressive completeness”, as Allo writes, or to put it simply, this is why inconsistent information might make sense descriptively but has no value normatively.

#### **4 Reply to Adriaans**

There is a fundamental disagreement between Adriaans and myself. Allow me to explain it with an analogy. Adriaans seems to think that a football game is best and indeed completely understood in terms of Newtonian physics. In contrast, I argue that Newtonian physics is insufficient and that you also, and more importantly, need to understand the meaning (semantics) of the game as well (strategies and calculations, history and psychology and so forth), or you will make little sense of it. Newtonian physics places some firm and inevitable constraints on any physical game but, from its perspective, there is no difference between Wimbledon and the World Cup. For this reason, one needs to keep the negative constraints satisfied but then move on and exercise semantic acumen in order to understand and explain the phenomenon in question. Following this analogy, you might imagine my astonishment when I saw myself caricatured in the article as if I had ever argued that information theory is neither necessary nor sufficient to develop our philosophical understanding of information and its related phenomena. The double negation is obviously nonsensical, I never held it, and I am not sure indulging in punching such a straw man is worthwhile. The actual view that I have defended is that information theory provides the scientific constraints within which we can develop an interesting philosophy of information. So I have argued for its necessity but against its sufficiency. Without information theory, there is no PI, but PI is much more than information theory on steroids. I foolishly thought the point was not only obvious enough, but also uncontroversial. Adriaans made me realise my mistake.

Given such fundamental disagreement, I shall restrain from commenting on the article's proposal. It belongs to an old-fashioned, perfectly respectable but also bankrupted tradition of attempting to squeeze semantics out of syntax. Using the previous analogy: the hope is that if you add enough physical rules and formulae, you will understand football one day. What I might stress is that the failure of such

an approach in the specific field of artificial intelligence has indeed caused me to take the semantic road that Adriaans finds so unappealing. It is not a road without serious intellectual challenges, but there is a significant difference between knowing that one approach is futile—namely, trying to develop a philosophy of information on the exclusive basis of mathematical results obtained from information theory—and having reason to believe that a different approach might be fruitful, namely, developing a philosophy of information semantically richer. This leads me to the valuable remark on my work that one encounters in the article: “Floridi’s efforts belong in [sic] the transcendental program”. It might not be a deep insight, but it is an absolutely correct one. For I explicitly acknowledged in the introduction of Floridi (2010b) that “[this] is also a German book, written from a post-analytic-continental divide perspective, more Kantian than I ever expected it to be”.

As for the rest of the article, its style and content unfortunately deprive it of most of its potential philosophical value. Imagine if I were to juxtapose the following two sentences lifted from the article: (1) “information theory has succeeded in formulating a mathematically sound solution to the general induction problem” and (2) “The quote above are typical for his style of presentation in which perfectly defensible positions are interleaved with fairly radical statements that are hard to interpret in a sensible way.” Wouldn’t this be just evidence of poor scholarship? My recommendation to the reader is to render unto the scientists the problems which are scientific, namely those which are empirico-mathematically solvable at least in principle, and unto the philosophers the problems that are philosophical, namely those which are intrinsically open, i.e. forever subject to well-informed and rational disagreement, even in principle. A careful blend of information theory and philosophy of information is more than just welcome; it is necessary if we wish to understand the complicated world in which we live. But confusing one with the other and randomly mixing bits of Shannon with bits of Heidegger is only a recipe for disaster.

## 5 Reply to Ganascia

I agree with much that Ganascia argues in his article. In particular, his interpretation of AI as a discipline with two souls is both correct and refreshing. Let me comment on the correctness first.

It is a well-known fact, although sometimes underestimated, that AI research seeks both to *reproduce* the outcome of our intelligent behaviour by non-biological means and to *produce* the non-biological equivalent of our intelligence. On the one hand, as a branch of engineering interested in *intelligent behaviour reproduction*, AI has been incredibly successful, well beyond the rosiest expectations. Nowadays, we increasingly rely on AI-related applications (so-called *smart* artefacts) to perform tasks that would be simply impossible by un-aided or un-augmented human intelligence. *Reproductive AI* regularly outperforms and replaces human intelligence in an ever-larger number of contexts. Next time you experience a bumpy landing recall that that is probably because the pilot was in charge, not the computer. On the other hand, as a branch of cognitive science interested in *intelligence production*, AI has been a dismal disappointment. Current machines have the intelligence of a toaster, and we really haven’t got much of a clue about how to move from there



(Floridi et al. 2009). *Productive AI* does not merely underperform with respect to human intelligence; it has not joined the competition yet.

Edsger Wybe Dijkstra's famous comment that "the question of whether a computer can think is no more interesting than the question of whether a submarine can swim" is indicative of the applied approach shared by reproductive AI. John McCarthy's disappointed comments about Deep Blue's victory against Kasparov are symptomatic of the sort of productive AI which frowns upon reproductive AI. The two souls of AI have often engaged in fratricidal feuds for intellectual predominance and financial resources. That is partly because they both claim common ancestors and a single intellectual inheritance: Turing, his machine with its computational limits, and then his famous test, and the fact that a simulation might be used in order to check both whether the simulated source has been produced and whether the targeted source's behaviour or performance has been reproduced or even surpassed. The two souls have been variously named, and sometimes the distinctions weak vs. strong AI or, as Ganascia reminds us, Good-Old Fashioned vs. New or Nouvelle AI, have been used to capture the difference. I prefer to use the less loaded distinction between light vs. strong AI (Floridi 1999). The misalignment of their goals and results has caused endless and most pointless diatribes. Defenders of AI point to the strong results of reproductive AI, which is really weak or light AI in terms of goals, whereas detractors of AI point to the weak results of productive AI, which is really strong AI in terms of goals. It is here that Ganascia's paper is refreshing, and this is my second comment.

Ganascia seeks to escape the dichotomy outlined above by defending the view that AI cannot be reduced to a "science of nature" nor to a "science of culture", because it is what he calls, following Rickert, an "intermediary domain", a science of the artificial, to put it with Simon (1996). The interactions between such sciences, their purposes and logics are a fascinating topic, which deserves further study. Here, I shall limit myself to calling the reader's attention to one specific issue raised by Ganascia: the relationship between artefacts, their users and the environment in which they interact.

Ganascia points out that the alleged limits of AI are "not caused by the oversimplification of AI models, like many people pretend nowadays, but by their inadequacy to the 'outer' environment". I will not develop the objection that one way of explaining why AI applications fail, when they do, is exactly because they are based on oversimplifications. Although failing to cope successfully with one's environment is one way of defining stupidity, there might be other reasons as well. What interests me here is to highlight a risk that might also be an opportunity.

Consider the following parody. Two people A and H are married, and they really wish to make their relationship work, but A, who does increasingly more in the house, is inflexible, stubborn, intolerant of mistakes and unlikely to change, whereas H is just the opposite but is also becoming progressively lazier and dependent on A. The result is an unbalanced environment, in which A ends up shaping the relationship and distorting H's behaviours, practically, if not purposefully. If the marriage works, that is because it is carefully tailored around A. Now, consider AI technologies as an instance of Information and Communications Technologies. I have argued that one of their philosophically interesting features is that they re-ontologise (i.e. modify the very essence of) the physical and conceptual environments in which they operate (Floridi

2007). They play the role of A in the previous analogy, whereas their human users are clearly H. So the risk we are running is that our technologies might shape our physical and conceptual environments and make us adjust to them because that is the best, or sometimes the only, way to make things work. After all, since AI is the stupid but laborious spouse and humanity the intelligent but lazy one, who is going to adapt to whom, if a divorce is not an option? The reader will probably recall many episodes in real life when something could not be done or had to be done in a very cumbersome or silly way because that was the only way to make the computerised system do what it had to do. Here is a more concrete, trivial example (philosophically, things are way more complex). The risk is that we might end up building houses with round walls and furniture with sufficiently high legs in order to fit the capacities of a Roomba (<http://www.irobot.com/>) much more effectively. I certainly wish our house were more Roomba-friendly. The example is useful to illustrate not only the risk but also the opportunity represented by AI's re-ontologising power.

There are many "roundy" places in which we live, from igloos to medieval towers and from bow windows to public buildings where corners of the rooms are rounded for sanitary reasons. If we spend most of our time inside squarish boxes that is because of another set of technologies related to the mass production of bricks and concrete infrastructures and the ease of straight cuts of building material. It is the mechanical circular saw that, paradoxically, generates a right-angled world. In both cases, squarish and roundy places have been built following the predominant technologies, rather than through the choices of their potential inhabitants. Following this example, it is easy to see how the opportunity represented by AI's re-ontologising power comes in three forms: rejection, critical acceptance and proactive design. By becoming more critically aware of the re-ontologising power of AI and ICT in general, we might be able to avoid the worst forms of distortion (rejection) or at least be consciously tolerant of them (acceptance), especially when it does not matter (consider the Roomba-friendly length of the legs of the furniture) or when this is a temporary solution while waiting for a better design. In the latter case, being able to imagine what the future will be like and what adaptive demands technologies will place on their human users may help to devise technological solutions that can lower their anthropological costs. In short, intelligent design should play a major role in shaping the future of our interactions with forthcoming technological artefacts. After all, it is a sign of intelligence to make stupidity work for you.

## 6 Reply to Piazza

I enjoyed Piazza's article. It is not only well-informed, but also analytically discerning about some of the less obvious features of my epistemological proposal. Of its many valuable contributions to the current debate on the possibility of analysing knowledge as accounted information, one seems to me to be of particular value. I may introduce it by quoting Piazza himself:

No less clearly, however, the considerations above also face Floridi's account with a potential difficulty, as they seem to sustain a reasonable doubt about the very viability of this strategy: if one believes that *knowledge can be acquired*

*through perception, or by testimony, and one also believes that in those cases there is no accounting or explaining information which could explain the epistemic status to which it is upgraded, one could well be tempted to suggest that knowledge could not, at least not in general, be analysed as accounted information; for at least in the cases just envisaged, an explanation of it will have to proceed by taking into account the justificatory role which perception and testimony seem to perform (italics added).*

I believe Piazza to be mostly right but perhaps in a way that may not entirely satisfy him, for his premises can be accepted, indeed strengthened, without accepting his conclusion. Let me explain why.

All the empirical information about the world that we enjoy flows and keeps flowing to us through our senses: directly, through our perception of the world, and indirectly, through our perception of other epistemic agents' perception of the world. We either saw it or read it somewhere, to put it simply, if slightly incorrectly (for we might have heard it or tasted it, and reading after all is also a case of seeing etc., but you get the picture). Thus, Aristotelians and Empiricists of various schools are largely correct in holding that *nihil est in intellectu quod non prius fuerit in sensu*. "Nothing is in the understanding that was not earlier in the senses", if and only if (the biconditional qualifies the "largely" above) what we are talking about is empirical information about the external world. If we then distinguish the direct and the indirect perception of the world by referring to the former as simply perception, the first-hand testimony of our senses, and to the latter as testimony, the second-hand perception by proxy, we see immediately that Piazza's article concerns the only two sources of empirical information available to cognitive agents like us. It is therefore essential to check how far his criticism might be justified.

Let me first clear the ground of a potential misunderstanding. Knowledge and information states and processes are sufficiently similar to be interchangeable in most daily circumstances, without any significant loss either in communication or in pragmatic efficacy. This fact reminds us that some tolerance in our semantics might be accepted sensibly. There is an imprecise but still reasonable sense in which, if you see that such and such is the case, then you hold the information that such and such is the case, and *ipso facto* you know that such and such is the case. Thus, if you see a yellow light flashing, then you know that there is a yellow light flashing in front of you. The same holds true for testimony: if you are told by someone, who saw a yellow light flashing, that there was a yellow light flashing, then you know that there was a yellow light flashing.

The value of such mundane equation—perceiving well-formed, meaningful and truthful data amounting to  $p$  is equal to being informed that  $p$ , which is equal to knowing that  $p$ —is that we gain much simplicity. The cost is that we lose the possibility of drawing some conceptual distinctions, which become essential once we wish to be precise in our epistemology. This is why some philosophers, including myself, resist such deflationism. The reluctance is due not only to the cost to be paid, but to the fact that such cost is philosophically unaffordable once we realise that knowledge is a specific kind of information, the kind enriched by the capacity of answering relevant questions about  $p$ , that about which one is informed.

Perception and testimony may be analysed along the same line because—in the best (i.e. non-Gettierised, scepticism-free, error-free, Floridi (2004)) circumstances—

they both *convey* information about their specific references: they are our information providers. Let us consider perception first.

Epistemologically, our bodies are our interfaces with the world. Their sensory apparatus implements hard-wired levels of abstraction (more technically, we are embodied gradients of abstraction), which determine the range and type of observable data that can be acquired. Perception is then a general term that refers to the process of data input through which epistemic agents like us acquire first-hand data about their environment at the levels of abstraction offered by their bodies. Such process of data input is fallible, but it can be corrected, enhanced (e.g. through a pair of glasses) and augmented (e.g. through a telescope). By itself, such a first-hand, data-gathering process may be considered a case of knowledge acquisition, but then any elementary signal-processing gadget would qualify as a cognitive agent, and this seems to be a bullet not worth biting. Let me explain. In some circumstances, we are not much better off than the aforementioned gadget. You see a yellow light flashing on the panel of your dishwasher at home. Suppose you haven't got a clue about what it might mean. At this stage, all you have acquired, through such perception, is at most the information concerning the light flashing. If you have further background information, e.g. about the covariance between the yellow light flashing and the dishwasher having run out of salt, then, by perceiving the light flashing, you may also acquire that further bit of information about the low level of salt. All this is uncontroversial. What is notoriously open to debate is whether such perception, by itself, may amount to more than information-gathering at best. I hold that, if we wish to be epistemologically accurate, it does not. It is not enough to perceive a yellow light flashing to know that there is a yellow light flashing in front of you, not just because a whole set of complex concepts must already be at play (light, yellow, flashing, the fact that lights can flash, that flashing lights of any colour on the panel of a white good are normally not decorative features but signals, that, as signals, lights off are less indicative than lights being on, that lights might not work properly but a flashing light is normally working well and it is meant to be intermittent, etc.), but equally importantly, because the perceptual data input (to simplify, there is a yellow light flashing there), plus the conceptual framework (the yellow light flashing there means...) required to formulate and make sense of it, further demands an explanation in order to graduate from information to the higher status of knowledge. In other words, unless you are able to answer a whole series of "how come" questions—how come that the light is flashing? How come that it is the yellow light and not another light that it is flashing? How come that the light is yellow? etc.—your status is no better than that of a dishwasher manual, where we can read that "the yellow light flashing indicates that the dishwasher has run out of salt". In other words, we would like knowledge to pass the Phaedrus' test.<sup>1</sup>

---

<sup>1</sup> [Socrates]: Writing, Phaedrus, has this strange quality, and is very like painting; for the creatures of painting stand like living beings, but if one asks them a question, they preserve a solemn silence. And so it is with written words; you might think they spoke as if they had intelligence, but if you question them, wishing to know about their sayings, they always say only one and the same thing [they are unary devices, in our terminology]. And every word, when [275e] once it is written, is bandied about, alike among those who understand and those who have no interest in it, and it knows not to whom to speak or not to speak; when ill-treated or unjustly reviled it always needs its father to help it; for it has no power to protect or help itself.

Let us now turn to testimony. This is the process through which epistemic agents like us *transfer* information to each other. Note that testimony does not generate information: the GIGO (garbage in garbage out) rule applies. If you tell me that  $p$ , e.g. that the dishwasher's yellow light was flashing yesterday, then, at most, I now hold the information that  $p$ . Unless we quietly presuppose that the receiver of  $p$  is doing more than just receiving and registering  $p$ —e.g. that the receiver is also evaluating the reliability of the source of  $p$ , but then this “more” is where a theory of account is hiding—all we have, at the end of a testimony process, is the transfer of some information from the original source to the final target, through a network of senders and receivers. The best that can happen is that the informational baton is passed through the several nodes that are relaying it without being lost. Luckily for us, testimony is not a Boolean process, the network is resilient—nodes can implement information correction procedures (as when a later epistemic agent recovers or reconstruct what was the original information and relays it in its corrected, restored format)—and there is often plenty of redundancy—as when several people act as independent sources, conveying the same information about the same event or repeatedly sending the same information at different times and through different channels (your wife tells you that the yellow light was flashing, and so does your daughter). Still, this is information *transfer*, not *yet* information (let alone knowledge) *generation*. Receiving  $p$  can hardly amount to knowing that  $p$ , for knowledge requires more than true content or, which is equivalent, well-formed, meaningful and truthful data. If this were not the case, any database would be very knowledgeable indeed, and all medieval scribes who copied Greek manuscripts without speaking much Greek at all would have been very learned. In other words, we would like knowledge to pass the parrot test (Descartes' *Discourse on the Method*): given that the yellow light was indeed flashing, being told, correctly, by a well-trained parrot that the yellow light was flashing while we were not in the kitchen does not seem to ensure that we know that the yellow light was flashing. At most, we have acquired that bit of information. If we do not do anything with it, that is all the epistemic dividends we may enjoy.

Let us now put the two threads together. Perception generates data about the world, which need to be interpreted to become information. It does not generate knowledge yet. Testimony transfers information (also but not only) about the world but does not yet generate knowledge. In both cases, what is missing, in order to gain empirical knowledge of the world in a precise epistemological sense, is the explanation of the empirical information acquired. Such explanation is obtained through the intelligent accounting of the available information. This is what I have argued in Floridi (2010b).

Time to return to Piazza's criticism. Piazza is right in stating that (first premise) “knowledge can be acquired through perception, or by testimony”, as long as “acquire” is understood, as it should, as stating necessary but not yet sufficient conditions. Compare this to “ $x$  (a mortgage, a passport, a skill, etc.) can be acquired through  $y$  (a credit evaluation, a full application, the relevant training, etc.)”. Indeed, in this sense, I have argued for a stronger thesis: empirical knowledge can be acquired *only* through perception or by testimony. If one day we will be able to implant Wiki-microchips under the skin, it will still be a case of testimony. Piazza is also right in stating that (second premise) “in those cases [perception and testimony]

there is no accounting or explaining information which could explain the epistemic status to which it is upgraded”, if we understand by this that unaccounted perception or testimony does not qualify yet as knowledge. Where he seems to be mistaken is in drawing the following conclusion from the previous two premises: “one could well be tempted to suggest that knowledge could not, at least not in general, be analysed as accounted information”. Nobody who understands the previous two premises should be tempted to jump to such conclusion. He adds that “for at least in the cases just envisaged, an explanation of it will have to proceed by taking into account the justificatory role which perception and testimony seem to perform”. So perhaps the problem lies with the devilish concept of justification. There are at least two ways in which seeing that such and such is the case justifies the seer to hold that such and such is the case. One is by interpreting the justification in terms of causality. Reliabilist theories used to like this approach. It seems impossible to disagree with this interpretation: it is the visual process of data-input that causally makes possible the acquisition of the relevant bits of information about the yellow light flashing. But causality is not what is being invoked here, since we are not looking for a descriptive account, but for a normative one. So the alternative is to use justification to mean exculpation. This, however, adds nothing to our or the seer’s understanding of the case in question, even if it does add a note on the epistemic conduct of the agent in question. He did not dream it nor imagine it; he did not project it out of fear nor carelessly assumed it: he saw a yellow light flashing, eyes wide-open, double-checking, changing angle and perspective. He really did his best to make sure that he saw a yellow light flashing. He did the right thing. The verdict is causally sound and epistemologically not guilty. Yet all this is irrelevant to the epistemic state of such agent. As I have argued above and much more extensively and in detail in Floridi (2010b), being right about  $p$  and having done everything reasonably possible to avoid being wrong about  $p$  do not yet mean that one knows that  $p$ . For a knower is “the man who knows how to ask and answer questions” (Plato, *Cratylus*, 390c), giving an account, that is.

## 7 Reply to Flavio Soares Correa da Silva

The perceptive article by Flavio Soares Correa da Silva provides original insights into a very interesting area of potential development of the philosophy of information, one to which I hope I may dedicate much more attention in the future, namely political philosophy. In this context, I would like to highlight two interesting contributions offered by Correa da Silva.

First, the article does a remarkable job in interlacing three distinct threads in my research on the philosophy of information, in order to form the original fabric of appropriate guidelines for the development of successful programmes for electronic government: (1) a modal logic for the formalisation of the notion of being informed; (2) Information Ethics and (3) the analysis of current trends in the development of the Web. Second, the article introduces the concept of “public moral agents”, and this is certainly worth all our attention. Correa da Silva defines a “public moral agent” as “a moral agent whose actions are, by definition and construction, fully accountable by third party”. This is already very interesting, but the really important

aspect is that he is able to include, in the same class of public moral agents, a whole variety of digital artefacts, engineered robots, social institutions and human organisations. Given the right level of abstraction, a piece of software, a drone, a state department and a company are human constructs which may (have to) obey the same general rules of design and ethical evaluation. Correa da Silva offers then a convincing case in favour of an evaluation of the construction, development and regulations of such agents in light of Information Ethics, with the goal of sustaining increasingly better forms of electronic government. He and I share the same cautious optimism about such potentialities. The article opens up many interesting lines of research. It is to be hoped that at least some of them will be pursued in the future. It is a practical goal that deserves to be taken as seriously as possible (see Briggie and Mitcham 2009).

## **8 Reply to Brenner**

The article develops an independent line of research, fully articulated in Brenner (2008), whose evaluation would require much more space than I have at my disposal here. So I shall limit myself to highlight, as briefly as possible, two simple points.

First, I am grateful to Brenner for providing a very reliable analysis of the work on Information Ethics that I have developed in the course of the past 10 years or so, with great accuracy and remarkable discernment. His scholarship is exemplary. As an example, I would like to offer the following comment:

If we follow Introna's division of approaches to the ethical implications of information technology into phenomenological, artefact/tool and social constructivist, I place Floridi's views in primarily the artefact/tool category, as his ethics analyzes the impact of technology on practices by applying a new moral theory to that can construct 'guidelines or policies' that may help correct injustices or potential infringements of rights resulting from the use of a technology. This is not to say that Floridi does not believe technology and society co-constitute or co-construct each other, but that one should start by focusing on the underlying (informational) characteristics of the technology.

Second, it is interesting to see how, despite the fact that Brenner comes from a very different perspective and tradition (he seems to be working within the dialetheist framework also discussed by Patrick Allo, see my reply to the latter), his position appears to be fruitfully compatible and synergetic with respect to some of the conclusions I have reached on the philosophy of information (Floridi 2010b). As Brenner acknowledges, much work still needs to be done, and many challenges have not been met yet, but optimism does not seem to be out of place.

## **9 Reply to Byron**

The following quotation seems to summarise well the paper by Byron: "once we grant that Floridi is right about the fourth revolution, the interesting issue becomes how we think about human nature". This is exactly what I have been arguing for some time. However, once this essential step is made, two further points of

significant importance distinguish Byron's position from mine: what the fourth revolution consists in and whether Information Ethics might help us to think more adequately about ethical issues brought about by the fourth revolution.

Regarding the first point, it is evident that in the article Byron attaches much (I fear too much) importance to the phenomenon of Artificial Companions (AC). It might be my fault, since, in the work primarily discussed by Byron, I relied quite extensively on AC as a good example of the sort of novelties that we shall be witnessing within the context of the fourth revolution. However, it was never my intention to associate the fourth revolution with some kind of breakthrough in artificial intelligence or human genetics. It seems a mistake to envision such a discontinuity with the previous three revolutions. For if the fourth revolution hypothesis has any hermeneutic value at all, this lies in the fact that, like the first three, computer science and its applications are making us experience a radical change in how we appreciate our own nature, our place in the universe and our potential responsibilities in it. What we rightly feel to be an extraordinary time in the history of humanity—brought about by amazing computational advancements and corresponding mind-blowing technologies—seems hardly explainable merely in terms of unprecedented types and scopes of interactions with the world and their pervasiveness. Such revolution makes much more sense once it is interpreted as a radical shift in our philosophical anthropology. Thus, the fourth revolution is more about social networks today than about futuristic robots. Yes, the capacity of building smart machines that one day might imitate humans so well as to become largely (and perhaps wilfully, on our side) indistinguishable from their human counterparts is indeed part of the fourth revolution (see my reply to Ganascia), but not, simplistically, because we shall be dealing with humanoid agents whose moral status will pose ethical issues. I strongly doubt that anyone well acquainted with the current and foreseeable status of our computer science and AI may take such concern very seriously (on the discussion of a “Floridi test” for AI see Bringsjord 2010). Even assuming a similar scenario were not science fiction, features such as having a mental life, enjoying semantic capacities and possessing intellectual abilities are already more than sufficient to render any speculation about the moral *responsibilities* of future artificial agents not merely idle, but dangerously distracting from the actual challenges we will effectively encounter. So the fourth revolution is not about future anthropominded robots, although increasingly smart artefacts do cast a new light on our self-understanding. And it is not about the future of human cyborgs either. In this case, too, the possibility of fanciful post-human IT-enabled or IT-enhanced beings, whether genetically or prosthetically, should be appreciated, philosophically, as the symptom of a deeper transformation in the way in which we are reassessing our nature. The very fact that we can think coherently about such future humanity shows that something deeper in our self-conception is being revised silently. We are not immobile, at the centre of the universe (Copernicus), we are not unnaturally detached and diverse from the rest of the animal world (Darwin) and we are not Cartesian subjects entirely transparent to ourselves (Freud). We are currently coming to terms with the possibility that we might not be disconnected and stand-alone material entities, but rather informational organisms, not unlike other biological agents and engineered artefacts, with which we share a global environment ultimately made of information, the infosphere (Turing). This is the fourth revolution.



Regarding the second point, according to Byron the Information Ethics that I have defended fails to provide a fruitful approach to the new issues posed by the fourth revolution. In this respect, Byron raises several specific questions. Some are reasonable. Some others seem to be based on a common misconception of Information Ethics, which confuses informational entities with entities carrying information, when in fact I am talking about informational entities in terms of entities understood informationally, i.e. as informational patterns (see the article by Durante and my reply). The essential problem at the roots of Byron's misrepresentation of Information Ethics is a lack of attention to the crucial role played by the concept of levels of abstraction (see in this issue the article by Flavio Soares Correa da Silva and Tony Doyle). If I do not deal with such questions in any detail here, this is only because they were discussed fully, and I hope answered successfully, in another special issue dedicated to my work on Information Ethics, to which I would like to refer the reader (Floridi 2008c). It is a pity Byron does not take that work into account.

What I find puzzling in Byron's approach, over and above his objections, is not his criticism, but the conservative perspective from which the latter is articulated. It is as if Byron were so keen on stressing the intellectual disruption brought about by the fourth revolution—to be more revolutionary than the revolution—that nothing old could possibly be heuristically helpful and anything new would be suspiciously mysterious. Byron holds that “the fourth revolution must be deeply disruptive of our dominant modes of ethical theorizing”. I fully sympathise. According to him, it will break with past ethical discourses (he lists eudaimonism, deontologism and consequentialism), bypass them entirely or render them obsolete. Although driven by a much stronger sense of continuity and less of a sense of total disruption, I am inclined to agree with him that an overall upgrade of our ethical perspective might be in order. But then, despite all his emphasis on how remarkably revolutionary our new post-fourth-revolution ethics will have to be, when presented with the alternative of Information Ethics, instead of making an effort to consider whether this might actually be the kind of quite new approach that requires a fresh start and an innovative outlook, he entrenches himself in the conservative attitude of considering Information Ethics inadequate because based on a “metaphysically puzzling and mysteriously teleological definition of information as something that strives to be realized”. I am not arguing here against the possibility that Information Ethics might be the wrong way forward. I have defended it elsewhere. I am stressing a more general need to keep an open mind towards alternative ways of approaching the ethical discourse, and the more so the more one thinks that the disruption caused by the fourth revolution is a historical fracture with the past. Radically new problems may call for unorthodox ways of thinking. Approaching the latter with a conservative attitude only ensures that the former will remain unsolved.

## **10 Reply to Doyle**

The article by Doyle provides a summary of some essential tenets of Information Ethics. This is the useful part. It also moves some objections against Information Ethics. Unfortunately, this is the less valuable part, not because it is incorrect, but because it is unacquainted with the literature. Similar objections have been articulated

before and have been the subject of a lively and public debate, and I have dealt with them in detail on several occasions. The interested reader is invited to check

- The special issues of the *APA Newsletter on Computers and Philosophy*, Spring 2008 Volume 07, Number 2 and Fall 2008, Volume 08, Number 1 (both freely available online), and especially my replies to John Barker and Edward Howlett Spence in Floridi (2008f; also freely available online)
- The special issue of *Ethics and Information Technology*, guest edited by Charles Ess (2008), especially my reply to Phil Brey in Floridi (2008d; the latter is freely available online)
- The special issue of *Metaphilosophy*, guest-edited by Patrick Allo (Allo 2010), especially my reply to Richard Volkman in Floridi (2010c; freely available online as well)

Of course, this is *not* to say that, on such occasions, I have countered all possible objections against Information Ethics successfully—the reader is invited to judge. But it does mean that it would have been interesting if Doyle’s article had taken into account the *current* state of the debate and discussed how satisfactorily some of the concerns it merely reiterates have actually been addressed. I am afraid objections fail to acquire more force through mere repetition. Thus, claims like (all italics added)

*the best Floridi can do* [to support the expansion of the moral circle to the whole of reality] is cite an apparent trend in moral philosophy towards even greater inclusiveness; [...]

so far as I can tell, *the only reason* that Floridi offers for IE’s theory of value is the alleged trend itself; [...]

aside from the argument from the expanding circle—which fails—Floridi *never argues* for the claim that the landfill, conceived as a collection of information objects, has intrinsic worth at all, let alone why those objects should be worth more intact than in pieces; [...]

so far as I can tell, *the only case* Floridi makes for his novel theory of value is the argument from the expanding moral circle [...]

are only evidence of poor scholarship, for they merely ignore the state of the current debate and 10 years of discussion. The result is that the reader genuinely interested in the current debate will not benefit much from this article but luckily can find a valuable analysis in Durante’s well-informed and insightful contribution. So, at the risk of boring the reader and repeating myself, in this inevitably short and selective reply, I shall sketch briefly what a quick search on Google or PhilPapers (<http://philpapers.org/>) would have helped Doyle to find. Again, my aim is not to prove that I am right; I only intend to show that the arguments in favour of Information Ethics are there for anyone to assess, if they care to engage with the literature and wish to treat their questions as more than mere rhetorical devices.

Doyle is keen on defending an ethics broadly based on “interests”, where interests are defined in terms of avoiding pain and seeking pleasure. Unquestionably, if one sticks to the pain/pleasure scale, any talk about inanimate things, both biological and artificial, indeed any reference to almost anything in the universe, with the exception

of a few biological species on our planet, will count as nothing. When pain and pleasure are in question, the only players in the ethical game are entities with a nervous system, and not even as groups, or as a species, but as single, healthy, individuals. Not even God, often considered the ultimate source of morality, may count, insofar as any biological attribution of pain or pleasure to such hypothetical entity would be preposterous (for the sake of clarity, let me hasten to add that I am not endorsing here the existence of God, but the validity of the following conditional: if God exists, then God should definitely count as a player in the ethical game, but positions based on mere pain/pleasure cannot account for the correctness of such conditional). Past or future generations, places and practices, stars and valleys, monuments and shrines, companies and governments, flowers and trees, a platoon or a team must also be disregarded. The Epicurean in each of us finds this perfectly understandable, tenable and defensible. I find it not mistaken, but unsatisfactorily limited. Not because it fails to square with the development of our ethical concerns. It does, but, as Doyle repeatedly states in the article, this is plainly not an argument, nor was it ever offered as such, since it was always meant to be just a *clarification* of why we might be wondering about the value of extending our environmental concerns these days. I find it unsatisfactorily limited because of a number of arguments, some of which might be worth briefly rehearsing here, since Doyle apparently failed to identify them (for further, valuable contributions see Adam (2008) and Bynum (2010) and the articles by Durante and Flavio Soares Correa da Silva in this special issue).

1. Why pleasure and pain, instead of existence/non-existence? After all, pleasure and pain are Nature's ways of signalling, to a very limited number of biological species, environmental risks and benefits, existence-enhancing and existence-debilitating or -threatening events. Being over-concerned with them is like being over-concerned with an unpleasant (but of course very useful) fever, rather than the nasty viral infection that causes it. It is like thinking that what is really right or wrong with the dishwasher is the yellow light flashing, rather than the low level of salt it indicates, and hence the issuing quality of its performance. Pain and pleasure are the servants of life and death, of well- and ill-being and of existence and non-existence. We should address their masters, the source of the moral discourse, not the evolutionary contingent symptoms.
2. Why adopt a presumption of guilt (an entity is morally worthless unless proven morally valuable), instead of a presumption of innocence (an entity is morally valuable unless proven morally worthless)? If we were to follow Doyle, we would be starting from the view that the set of entities that have some moral value is empty and then revise such a position by progressively including types of entities on the basis of considerations limited to the presence of a well-functioning nervous system. The point here is not to reject such procedure as unsuccessful. After all, I suspect that that is pretty much the way humanity has been upgrading its ethical perspective for centuries, and my past reconstruction of how our ethical discourse might have developed in Western philosophy was based exactly on such a historical trend. The important point is to ask ourselves what is so dogmatically untouchable about such a procedure. In the legal system, we move exactly in the opposite direction. We presume innocence and

ask for evidence against it. And we do this for excellent reasons: because we do not trust our epistemic capacities nor the powers in charge of the legal system. In the worst possible scenario, we would rather be wrong about treating as innocent a guilty person than doing the opposite. Likewise, by default I would rather (invite everybody to) exercise respect and care towards any part of the non-biological universe, by treating it as morally valuable, than vice versa. And we should rather have the powers in charge constrained by a presumption of innocence, than vice versa. So I would rather have individuals, governments and companies abide by the rule of presumption of moral value of the whole world in which we live, no matter whether we are talking about a human artefact or a natural environment, than vice versa. This is why I have argued for a similar approach in ethics. By default, we should begin by exercising respect towards the whole of reality, limiting it, qualifying it or even dropping it whenever necessary. Let us start from a full set and proceed by impoverishing it, if necessary. Recall that I have also defended a minimal and overridable respect for reality. Of course, having a lite ontology will help, and that is why I suggest we adopt an informational ontology (Ess 2009). But other Western philosophies (e.g. Platonism, Stoicism, Spinoza) and Buddhism (Hongladarom 2008) have been suggesting roughly the same approach for a long time, even without a philosophy of information, so this is an open option, not a must. I have articulated this point about the enlargement by showing that there is really no reason to stop lowering the line below which something deserves to be morally disrespected. You can go all the way down and lose nothing. This is worth emphasising. Since we are talking about fundamental assumptions, it is hard to try to convince someone to change his outlook, but consider the following argument.

3. Why be so scared about the ultimate extension of the circle of entities that in principle, if possible, would deserve some (make it as minimal as you wish, but not zero) degree of respect? If we are mistaken in including in our ethical concerns the non-biological universe, nothing morally wrong will have happened; if we are not mistaken, we will have been better moral agents. A win-win situation.
4. Sometimes we wish to have practical motivations to accept an ethical perspective, not just philosophical reasons. So try to ask yourself (this is not a psychological experiment, it is a philosophical argument, just use Rawls' veil of ignorance together with Moore's two worlds): would you prefer to live in a universe where all human agents make their decisions by caring only about their pains and pleasures, in which they really think that, and coherently behave as if, "non-human animal life has no intrinsic worth. It is literally meaningless"? Or would you rather live in a universe in which such agents take into account the well-being of everything, more like Buddhism and Information Ethics and other forms of Environmentalism suggest? I hope I can guess the answer.

Much more could be and has been said in favour of Information Ethics—e.g. in terms of heuristic and hermeneutic value, when it comes to issues raised by our information society, or fruitful applicability—but I hope the reader might have gained a sufficient overview of the sort of arguments that have been developed in its favour. Doyle will probably not like my conclusion, but after

reading his article, I am even more convinced that we need to move firmly and quickly towards an ethical perspective that drops our human chauvinism and expands our ethical concerns beyond justified but insufficient concerns for our pain and pleasure. Recent natural disasters suggest that the state of our nervous system is hardly the only guideline sufficient to orientate our moral behaviours. We are becoming too powerful, and we are still too accident-prone and epistemically fallible to afford such a human-centric approach. One day, *pace* Doyle, I hope we shall be able to define not only crimes against humanity but also crimes against the earth and reality.

## 11 Reply to Hofkirchner

I cannot claim to have been able to appreciate in full all the points made by Hofkirchner in his article. It seems that our divergences are more a matter of nuances than substance, but as the reader knows, the devil loves details, for that is where it hides more easily, so I might be wrong. I am also unclear about what Hofkirchner's proposal amounts to, over and above the mosaic of quotations and a revision of some of my theses, but perhaps a good reply is that I need to go back to the drawing board and just try harder. Of the many things touched upon in his article, I will concentrate on two, which seem to me in need of some clarification. Perhaps the dialogue can start from there, once the confusion is cleared up.

I am indebted to Hofkirchner for this opportunity to clarify that I had not heard of Mazlish (1993) until I gave a talk at Yale in 2010 on the fourth revolution, when Bonnie Kaplan very kindly called my attention to Mazlish's "fourth discontinuity". Shame on me, since I realised that Mazlish published his essay, entitled "The Fourth Discontinuity", in 1967 (Mazlish 1967, repr. in Taviss 1970). Kaplan, as well as Hofkirchner, spotted the family resemblance between Mazlish's and my view. However, she also saw the significant difference. Let me try to summarise it here as succinctly as possible.

There are plenty of "fourthes". Freud's "three revolutions" is a very attractive springboard, and I am hardly the first to use it to make a hermeneutical move forward. Erik Erikson, for example, listed Einstein as a fourth revolutionary figure (Hoare 2002, p. 133). More recently, it has been argued that the discovery of DNA and the genetic code might also qualify, with a "fifth revolution" being represented by the "neuroscience revolution" (Wired 2006). Examples could easily be multiplied. What matters, obviously, is not who listed what and in which order, but rather the nature of the hypothesis suggested and the reasons offered in its favour. I have argued that the fourth revolution—with Turing as its most plausible reference (Bolter 1984)—is a fruitful way of understanding the extraordinary impact that computer science, ICT, digital technologies and more generally the information society are exercising on our self-understanding. So far, the point does not differ significantly from views such as those listed above. This is just a matter of logic: if you are going to use Freud's three revolutions, you are bound to talk about a change in our self-understanding, brought about by some scientific development. The philosophical effort only begins here, for it concerns the aforementioned devilish details. The fourth revolution, as I described it, is most emphatically *not* about

building artificial agents which might one day be confused with, or overcome, their creators and *hence* change our self-understanding. I love science fiction, but I love truth more and find the former dangerously distracting when dealing with serious issues (see my replies to Ganascia and Byron). As a review in the *New York Times* stated, “ultimately, his [Mazlish’s] book degenerates into [sci-fi] movie criticism”.<sup>2</sup> I have tried to stay away from such pitfalls. The change in our self-understanding concerns the interpretation of ourselves as informational organisms (infor), which have much in common with other biological, artificial and hybrid infor. It has to do with Artificial Companions (Floridi 2008a) not *Terminator*, with GPS, not *Star Trek* teleportation, with the construction of personal identities on *Facebook*, not with the force in *Star Wars*. This is where a firm grasp of the method of abstraction and what it means to use a Level of Abstraction in order to analyse a system, becomes a requirement rather than an option. Unfortunately, there is no mention of any of this in Hofkirchner’s paper. As infor, we inhabit an environment, the infosphere, which is full of other non-single-human agents, including your credit card company. As infor endowed with a mental life and the capacity to make sense of the world and think about the choices and consequences of our actions, we are also the special agents who can and should take care of such environment as a whole.

This leads me to a second point that left me slightly disappointed. The article is rather cavalier towards Information Ethics, to say the least, but the criticisms moved are ungrounded. Sound and informed criticisms are always welcome, but generic comments based on a lack of serious engagement with the literature only indicate a missed opportunity. I would like to ask the reader to compare this article with Durante’s. Take the remark “This looks like the notorious naturalistic fallacy” and the ensuing discussion. If something looks like the notorious naturalist fallacy, one should equally suspect that it looks that way to others, including the proponent. So it might be worth googling “luciano floridi naturalistic fallacy” and quickly discover that I have dealt with the problem that “an ontocentric approach [such as the one I have defended in Information Ethics] is often threatened with the naturalistic fallacy” in Floridi (2008c). The whole text is online, so I shall spare the reader my response. Of course, I may still be utterly wrong, but merely being told that I am fails to convince me that I might be.

As I wrote at the beginning, I have the impression that there may be much that I should learn from Hofkirchner’s article. The confusion might be my fault. I will have to try harder.

## 12 Reply to Vakarelov

The paper by Vakarelov is a very welcome revival of a pragmatic and semiotic approach to the philosophy of information, which had been disregarded for too long. It is both instructive and original. I have the impression that its sophisticated proposal is more complementary than alternative to the semantic approach I have privileged in most of my writings. I specify “most” because, in at least two cases,

---

<sup>2</sup> Available online, see <http://www.nytimes.com/1993/10/24/books/we-are-what-we-make.html>

when discussing the symbol grounding problem and the correctness theory of truth, I have explicitly and extensively relied on a pragmatic approach (see now Floridi 2010b). More specifically, Vakarelov seems to me to be right, and in an interesting way, when he writes that

According to this strategy [his pragmatic approach to semantic information], *s* is an information system not because it operates with meaningful (and truthful) data, i.e. because it operates with information, but conversely, it operates with information because it is an information system. The most important idea is that *what counts as data and what gives the data semantic content is determined by the role it plays in the information system* (italics in the original).

I thoroughly agree. It is in the specific implementation of such an important idea that Vakarelov and I might take different directions. Let me sketch mine and briefly compare the two.

For some complicated reasons too long to explain here, some recent research on the informational interpretation and construction of personal identity (Floridi 2005, 2006) has led me to realise something that is almost embarrassingly obvious: Shannon's classic model, even when stripped of all its complex features and reduced to the simple quadruple <Sender, Channel, Message, Receiver>, presupposes a very rich and developed environment, where entities are already constituted as senders and receivers *before* any information flow takes place. Indeed, they actually make possible the presence of a message-passing system. This is perfectly fine for any engineering purpose, but hardly satisfactory from a philosophical perspective. How did the entities in question come to play the roles of senders and receivers? From a biological view, it is obvious that organisms, understood as natural information systems—I am following Vakarelov's terminology here, although I would call them natural *data* systems myself—postdate the availability of the sort of data that they can exploit in order to survive, flourish and reproduce. With a classic example, also used by Vakarelov: in logical order, first comes the oxygen-rich vs. oxygen-poor environment and then bacteria able to exploit such difference. The biological debt incurred by organisms is one way: they owe their existence to the data in the environment, which owe them nothing. Now, I have remarked in many occasions that data are, ontologically interpreted, differences or points of lack of uniformity. Put together the biological and the ontological suggestions and what you obtain is a thesis that can be phrased in Biblical terms: in the beginning were the data. Natural information (or, better, data) systems came later. The same thesis can be formulated negatively: no data, no differences, no organisms, no natural information systems, no minds. It seems to me that this is the right way of approaching the naturalization of information so well analysed by Vakarelov. Replace "influences" with data in Vakarelov's paper and you will see that we are probably pointing in the same direction. He might disagree, however, because he seems to have a more "semantic" notion of data (e.g. as digits in a spreadsheet). There are other places where the terminology of the paper does not help the dialogue on the issues at stake. For example, goal-oriented behaviour is better understood by specifying the level of abstraction at which, and above all, through which, it is identified. Likewise, in a context where "information" and "informational" are highly technical terms, more care should be exercised in

talking about systems as information systems. Strictly speaking, bacteria, for example, are really data systems, insofar as they lack any semantic capacity. Nothing has any meaning to a bacterium, if not metaphorically. But these are details, and some conceptual negotiations could easily reduce the distance between Vakarelov's approach and mine. However, things seem to be somewhat less reconcilable when our mutual strategies are in question. Vakarelov's paper appears to endorse a reductively naturalistic attitude. It seeks to naturalise semantic information from beginning to end, from thermostats to bacteria to humans. In contrast, I do take naturalism to be a very reasonable (perhaps the only sensible) way of explaining the beginning of the story about semantic information, but I also hold the view that naturalism does not have boundless explanatory scope. For it seems to me that it can hardly account for some of the most interesting aspects of the development of semantic information. It does not tell us much about the plot and the end of the story. Metaphor aside, a semantic approach (what Vakarelov calls the amendment view) can be fully naturalistic about the pre- and post-biotic emergence of semantic information, without being reductionism when it comes to understand what epistemic agents are able to do with semantic information mentally, culturally and normatively. For example, the reader might wish to check Gillies' paper in this collection to see very quickly that there is little hope to provide a satisfactory account of mathematical information from a naturalistic-only perspective. Or she might consider Durante's discussion of the normative aspects of a philosophy of information or the analysis of testimony in Piazza's contribution. Information starts as a natural phenomenon but ends as a mental one. No methodological approach that ignores this fact can provide a satisfactory philosophy of information. Focusing exclusively on one side of the bridge means missing a significant and substantial part of the picture. If we need a term, we might be calling the complementary approach the semanticisation (the process of giving meaning and sense to something) of information. The problems caused by an exclusively naturalistic analysis of semantic information are those typical of any tunnel-vision: an initially fruitful focus soon becomes blindness to other significant aspects of the targeted topic. Thus, Peirce, who knew very well the advantages and disadvantages of (what we have labelled) naturalization, insisted that semantic information requires a mental life and ideation to be explicated. As Vakarelov reminds us, Peirce was unhappy about the qualification of sunflowers as information systems. He was right. Sunflowers are not even data systems, like bacteria, for they are coupled to the environment by co-varying relations that are only misrepresented by an informational analysis. Since Peirce had a rather comprehensive concept of the mental, we might translate his point in terms of cognition. Semantic information requires some cognition, at least in terms of the capacity to interpret something as something else, for example the vibration on the web as a signal by the spider that a prey has been captured. If one keeps refining Nauta's definition of information system until it is fully explicit and adequate, my impression is that one ends up with something that is equivalent to a cognitive system. Unsurprisingly, Vakarelov disagrees with this line of reasoning. He writes: "His [Peirce's] solution was to insist that the interpretant of the sign [semantic information] must be a mind. From a naturalistic standpoint this is unacceptable". Vakarelov is right, it is unacceptable, but so was Peirce, a mind is required, so this is precisely why a naturalistic standpoint is insufficient to make sense of information. A semantic standpoint is inevitable.



### 13 Reply to Yukio-Pegio Gunji, Takayuki Niizato, Hisashi Murakami and Iori Tani

The article by Gunji, Niizato, Murakami and Tani seeks to extend some results obtained in the philosophy of information to the biological sciences and, in particular, to zoology. I am intrigued by such developments. Certainly, their analysis of the concept of flattening seems to deserve both careful study and further development. In order to contribute to such interdisciplinary dialogue, let me offer a clarification concerning the notion of change.

One of the most abstract observations possible of a population is its *size* or cardinality as a set, that is, the number of its members. In the terminology of the method of abstraction (Floridi 2008e, 2010a, b), that is an observable of type WHOLE NUMBER, and it is well-typed provided that the population is finite and well-defined. At this LoA, a population provides no more information than its *size*. There is no distinction between live populations, artificial populations and for that matter, any ordinary set. In order to gain more information, one may then add one more observable, call it *generation number of each member*. Its type is FUNCTION FROM MEMBERS OF THE POPULATION TO WHOLE NUMBERS: the argument of the function is the member of the population, and the result is its generation number. If the original members of the population were observable, then some convention would be necessary for their generation number; a reasonable choice would be 0. More details can easily be added. A more general alternative, always well-typed, is obtained by assuming time (rather than generation number) to be observable and observing the *size of the population at any time*. The result is of type FUNCTION FROM TIMES TO WHOLE NUMBERS. An interface (a LoA) containing such observables, though simple, would be sufficient to support the well-known Fibonacci model of population size per generation. At this LoA, a population provides no more information than its growth. However, this might be already of some interest if we include additional observables such as the rates of birth, mortality, migration or harvesting. It is sufficient, for example, to discuss age distribution across the population and for the management of pest populations, for the harvesting of natural populations, for the modelling of insect outbreaks, of interacting populations and so on. One further step, in revealing information about the population, is to investigate aspects of the state of its members, perhaps the simplest being position. Thus, one may consider *spatial distribution* as a further observable. If the population consists of cellular automata, then the distribution is a subset of automaton positions, namely those positions at which automata are located. If it consists of a population of animals, then the observable's type is some mathematical representation of the domain inhabited by the animals, for example a map with a point on it for each animal. In the case of bird flocks, as well illustrated by the authors of the article, it is of interest to observe *spatial distribution at time t*, whose type is FUNCTION FROM TIMES TO SPATIAL DISTRIBUTIONS. This LoA is sufficient for many studies of population dynamics, both in biological and in artificial contexts. One may, for example, consider the population of "bicycles in Oxford". They arrive in Oxford, move around, sometimes in correlated ways or at specific times (e.g. according to the lecturing schedule) and even leave. The story is that they are stolen and sold in Cambridge, and vice versa, thus giving the impression of a sort of two-way migration pattern. At this LoA, they appear to be equivalent to entities which are born, graze, possibly migrate and die. In

terms of recycling of parts, they might even be seen as reproducing. Alternatively, for a system of cellular automata, *spatial distribution* can be seen as a special case of *individual state*, in which state contains just one bit of information: dead or alive. In all these examples, the population forms a dynamical system whose next state is a function of just the current state, and at least (part of) the current state is observed. This important setting characterises most modelling in applied mathematics and computer science. It implicitly endorses a very strong assumption: the *Markovian nature* of the transfer function. This means that state transitions (both continuous and discrete) are treated as a process in which the probability of the system to be in state  $X$  at time  $t+1$  is dependent only upon the state (and hence the properties)  $Y$  of the system at the immediately preceding time  $t$ . Intuitively, this is why you can join a chess game at any time and still follow what is happening, or become an expert in chess endings: the state of the chessboard at  $t+1$  depends only on the state of the chessboard at  $t$ . Such Markovian approach (the causal version is known as the causal Markov condition), interpreted from the perspective of a philosophy of information and a method of abstraction, seems to be neither a case of discovery—since the world often appears to be more inextricably complex and intertwined than a chess game—nor a case of mere invention—the successful nature of our epistemic efforts based on Markov analyses hardly justifies a complete form of scepticism, as if Markov conditions were artificially and externally imposed on a recalcitrant world. It seems more a case of design features: our method of abstraction, both theoretically and empirically (whenever it makes sense to talk of algorithms, the actual software programs and the resulting simulations), naturally leads to a Markovian approach, and at the same time, it privileges and makes more salient those features and aspects of the systems under observation that are Markov-friendly. And this points in the direction of a more metaphysical lesson. Differential equations, state transitions, Bayesian networks, Markov chains and so forth should be seen more as ways of *re-structuring* the targeted systems (structures), after they have been articulated (de-structured) in figures or states, nodes or events. The reifying nature of our cognitive approach is such that we first of all look at a billiard game (the structure) as divided (de-structured) into sequences of well-formed events, in which well-formed and independent objects, with specific properties, behave in such a way as to generate transformations. We then find such dismembered (de-structured) system in need of re-composition, or reconstruction (re-structuring). As in the case of a vase which we have first shattered into pieces—or like a Humean philosopher thinking that there are actually two events in the world, one consisting of a stone flying against the window and the second consisting of the window breaking—we are then left with the task of gluing together what was intact and in one piece in the first place. Such re-structuring of a de-structured structure cannot restore the pristine state. To put it in Humean terms, causing is not necessitating; there is no necessity linking the de-structured event A and event B, for the simple fact that if there were, we would not be able to split the two events in A and B. The broken glass will not be fully restored, yet it is crucial to realise that this is not a problem at all, for the re-structuring follows normative rules (how the modelling should be done), like the Markov condition, which are not meant to ensure that the re-structured model is indistinguishable from the original un-de-structured system. Their goal is epistemic and pragmatic—understanding, prediction, explanation, control, modification, construction, reproducibility—not

metaphysical (grasping as far as possible the essential nature of the system), and this is what both justifies the de-structuring in the first place and what guides the re-structuring afterwards. Causality is an epistemic glue.

## 14 Reply to Durante

I learnt much from Durante's article. His treatment of issues, which seem to have baffled less sharp colleagues, is both original and enlightening. His analysis of the informational nature of Being, for example, and of the correlated issue concerning the richness of Being understood informationally, is impeccable. The same holds true for his discussion of the essential notions of data, lack of uniformity, difference and relation. Here is a wonderful instance of his interpretation:

When destroying informational objects, not only we erase differences, but we tear also the threads of those relations, with a result of a stratified impoverishment of the infosphere and of pluralism: in effect, pluralism is not only concerned with the elimination of the secondary manifestations of the variety of informational objects (implemented in material objects), but it is embedded in the ontological roots of the infosphere.

If I may issue a recommendation, I would strongly advise any reader interested in grasping the nature of Information Ethics to study his article. But I do not wish to waste this opportunity by only praising Durante's work. I would rather add at least one further reflection, which I hope will contribute to the ongoing debate on the nature and scope of Information Ethics.

The reflection concerns the special nature of human beings in the universe. Not a small topic, I know. That is why I shall follow Plato, who, in the *Phaedrus* (246a), acknowledges that

To tell what it really is [the form of the soul, or for us the special nature of human beings] would be a matter for utterly superhuman and long discourse, but it is within human power to describe it briefly in a figure; let us therefore speak in that way.

In other words, allow me to be metaphorical.

In the informational fabric that we call Being, there are some special nodes. They are unlike any other kinds of nodes. Perhaps they emerge out of an odd evolution of unrepeatable chances. Think of them as Nature's beautiful mistake. They were not meant to be there, like a unique kind of a most unlikely pattern, but they have occurred, and they can now marvel at the most improbable chances that brought them into existence. Or perhaps they have a divine origin, as Plato thought. Either way, these nodes are informational structures like all other nodes, encapsulated packets of differences, relations and processes, which contribute to the value and richness of the whole. Their special nature lies not in what they are—in their physics and biochemistry, to use a different level of abstraction—but in what they can do, for they are *structuring structures*, the ultimate defence against entropy/evil. They are the *loci* where the flow of information reaches its maturity and becomes self-conscious, capable of self-determination and able to decouple itself from the rest of

the fabric and reflecting on its own nature and status, thus shifting from a Darwinian, physical evolution to a Lamarckian, mental development (we have developed incommensurably faster than our bodies). Such nodes, you and I included, have an unclear destiny. They might hope that their moral struggle against entropy is actually a small episode in a divine plan. If so, this can only be reason for rejoicing. Or they might fear that such a struggle is unfortunately only a titanic effort in an unrewarding and lonely universe, a thin red line against the vandalism of time, whose failure can be delayed and mitigated, but not avoided. If so, this should still be reason for some modest rejoicing,<sup>3</sup> for they will have helped reality to die of a more graceful death. Such nodes are the stewards of Being. They may do whatever they like, as long as they are careful.

## References

- Adam, A. (2008). Ethics for things. *Ethics and Information Technology*, 10(2), 149–154.
- Allo, P. (2010). Luciano Floridi and the philosophy of information. *Metaphilosophy*, 41(Special Issue, 3), 247–254.
- Bolter, J. D. (1984). *Turing's man: Western culture in the computer age*. London: Duckworth.
- Brenner, J. E. (2008). *Logic in reality*. Dordrecht: Springer.
- Briggle, A., & Mitcham, C. (2009). From the philosophy of information to the philosophy of information culture. *The Information Society: An International Journal*, 25(3), 169–174.
- Bringsjord, S. (2010). Meeting Floridi's challenge to artificial intelligence from the knowledge-game test for self-consciousness. *Metaphilosophy*, 41(3), 292–312.
- Bynum, T. W. (2010). Philosophy in the information age. *Metaphilosophy*, 41(3), 420–442.
- Cassirer, E. (1953). *Substance and function, and Einstein's Theory of Relativity*. New York: Dover. Translation by William Curtis Swabey and Marie Collins Swabey. The first part *Substanzbegriff und Funktionsbegriff* was published in 1910, while the second part, *Zur Einstein'schen Relativitätstheorie*, appeared in 1921. Unabridged reprint of 1923 edition, published by Open Court Publishing Company, Chicago.
- Dumas, A. (2008). *The man in the iron mask*. Oxford: Oxford University Press. Edited with an introduction and notes by David Coward.
- Ess, C. (2008). Luciano Floridi's philosophy of information and information ethics: Critical reflections and the state of the art. *Ethics and Information Technology*, 10(Special Issue, 2–3), 89–96.
- Ess, C. (2009). Floridi's philosophy of information and information ethics: Current perspectives, future directions. *The Information Society: An International Journal*, 25(3), 159–168.
- Floridi, L. (1996). *Scepticism and the foundation of epistemology: A study in the metalogical fallacies*. Leiden: Brill.
- Floridi, L. (1999). *Philosophy and computing: An introduction*. New York: Routledge.
- Floridi, L. (2004). On the logical unsolvability of the Gettier problem. *Synthese*, 142(1), 61–79.
- Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology*, 7(4), 185–200.
- Floridi, L. (2006). Four challenges for a theory of informational privacy. *Ethics and Information Technology*, 8(3), 109–119.
- Floridi, L. (2007). A look into the future impact of Ict on our lives. *The Information Society*, 23(1), 59–64.
- Floridi, L. (2008a). Artificial intelligence's new frontier: artificial companions and the fourth revolution. *Metaphilosophy*, 39(4/5), 651–655.
- Floridi, L. (2008b). A defence of informational structural realism. *Synthese*, 161(2), 219–253.
- Floridi, L. (2008c). Information ethics: A reappraisal. *Ethics and Information Technology*, 10(2–3), 189–204. Special issue on "Luciano Floridi's Philosophy of Information and Information Ethics: Critical Reflections and the State of the Art", edited by Charles Ess.

---

<sup>3</sup> I am most grateful to Scott Molony for having made me realise this point during one of our meetings on Information Ethics.

- Floridi, L. (2008d). Information ethics: A reappraisal. *Ethics and Information Technology*, 10(2), 189–204.
- Floridi, L. (2008e). The method of levels of abstraction. *Minds and Machines*, 18(3), 303–329.
- Floridi, L. (2008f). Understanding information ethics—replies to commentaries on Floridi. *APA Newsletter on Philosophy and Computers*, 8(2). available online.
- Floridi, L. (2010a). Levels of abstraction and the Turing test. *Kybernetes*, 39(3), 423–440.
- Floridi, L. (2010b). *The philosophy of information*. Oxford: Oxford University Press.
- Floridi, L. (2010c). The philosophy of information: Ten years later. *Metaphilosophy*, 41(3), 402–419.
- Floridi, L., Taddeo, M., & Turilli, M. (2009). Turing's imitation game: Still a challenge for any machine and some judges. *Minds and Machines*, 19(1), 145–150.
- French, S., & Ladyman, J. (2003). The dissolution of objects: Between platonism and phenomenalism. *Synthese*, 136(1), 73–77.
- Hilbert, D. (1971). *Foundations of geometry*, 2nd edn (La Salle, Ill: Open Court). Translated by Leo Unger, from the 10th ed. revised and enlarged by Paul Bernays.
- Hoare, C. H. (2002). *Erikson on development in adulthood: new insights from the unpublished papers*. Oxford: Oxford University Press.
- Hongladarom, S. (2008). Floridi and Spinoza on global information ethics. *Ethics and Information Technology*, 10(2), 175–187.
- Mazlish, B. (1967). The fourth discontinuity. *Technology and Culture*, 8, 1–15.
- Mazlish, B. (1993). *The fourth discontinuity: The co-evolution of humans and machines*. New Haven: Yale University Press.
- Notturmo, M. A. (1985). *Objectivity, rationality, and the third realm: Justification and the grounds of psychologism, a study of Frege and Popper*. Dordrecht: Nijhoff.
- Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology*. New York: Oxford University Press.
- Shapiro, S. (2000). *Thinking about mathematics: The philosophy of mathematics*. Oxford: Oxford University Press.
- Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge: MIT Press.
- Taviss, I. (1970). *The computer impact*. Englewood Cliffs, NJ: Prentice-Hall.
- Wired (2006). What is your dangerous idea? [http://www.edge.org/3rd\\_culture/ramachandran06/ramachandran06\\_index.html](http://www.edge.org/3rd_culture/ramachandran06/ramachandran06_index.html).