NEWSLETTER | The American Philosophical Association

Philosophy and Computers

SPRING 2016 VOLUME 15 | NUMBER 2

ISSN 2155-9708

APA NEWSLETTER ON

Philosophy and Computers

PETER BOLTUC, EDITOR

VOLUME 15 | NUMBER 2 | SPRING 2016

FROM THE EDITOR

Peter Boltuc
UNIVERSITY OF ILLINOIS–SPRINGFIELD

We feature L. Floridi's article "Moving Back on Top of the Wave." It presents Luciano's vision of philosophy as a framework for tackling multi-solvable problems. Floridi argues in favor of a philosophy that, like Ancient Icarus, tries to avoid, on one hand, the heat of relativism, on the other the sea of pre-Kantian naïve ontology. The paper by M. Fultot builds on Floridi's ontologically based ethics of information. The author uses intriguing pieces of science to show that informational richness is a by-product of a broader process, of nature producing the maximum entropy. Hence, well-organized structures are the most efficient conduits in producing entropy. While Fultot views the idea as a naturalistic grounding of Floridi's information ethics, a reader's critical eye may perceive his view as an ethics based on *nirvana* (viewed as eternal non-being), diametrically opposed to Floridi's attempt at building the ethics based on universal criteria of existence.

The topic whether we should fear or applaud the chance of AI developing human-level intelligence and beyond is tackled—from very different angles—in the featured article by L. Floridi and in the note by T. Powers (in the context of computer ethics). It is the main gist of the commentaries by K. Sotala and F. Gobbo. Let me join in on this: Arguably, the most significant debate in the history of philosophy of technology may have taken place in the 1800s. Marx's critique of Proudhon's anarcho-syndicalist offensive against industrialization may seem unexpected for today's readers since, on this topic, Marx largely followed Smith and Riccardo on the side of progress (though he meant to re-shape its socio-political context). Most twentieth-century socially engaged philosophers followed Proudhon's attitude that views technological progress with fear and some level of trembling, which leads to advocating desperate measures to derail it. It turns progressives into the social rearguard. The point is not that every form of technological progress is always already a good in itself, it is rather that progress is *in principle progressive* and, as such, necessary for development. Attempts to enhance human flourishing, justice, or well-being are more than sheer hopes only insofar as they are situated in the broader framework of economic and technological progress, not in opposition to it;[1] the versions of post-Modernism that try to deny and oppose this point, confuse the base with superstructure, thus putting themselves out of the realm of Marxian socio-economic reality. They also, obviously, leave the realm of modern economic theory, which is predicated on growth. Some authors suggest that there is a continuum between attempts to slow down, or derail, technological development on one hand and constructive steps to slow down, and eventually reverse, global warming (and other civilizational hazards) on the other. This is a false analogy. Attempts to tackle global warming and other civilizational perils are best attained by progress, often based on advanced engineering technologies,[2] as long as guided towards socially responsible causes. Solutions rarely ever come from Proudhonian attempts to turn back the clock.[3] When intellectuals eager-up to battle the forces of technological, scientific, and civilizational progress Copernicus, Bruno, and all the saints of the Enlightenment turn in their graves. Hence, conversation between Floridi and some of his critics—e.g., Sotala—builds on the main theme of progressive tradition: Do we follow Riccardo and Marx, or rather Proudhon and the Luddites?

We also present, in this issue, significant reviews of Floridi's *The Ethics of Information* written by J. Bus and D. Chapman. We close with a call for more philosophy in the philosophy of computer science by R. Hill. This is a nice fit with the recent initiative by people at the APA to revise the mission of this committee. I hope we will have a chance for discussion to continue in the following issue.

NOTES

1. K. Marx and free market economists such as A. Smith would agree.
2. Tzvi Bisk and P. Boltuc, "Sustainability as Growth," in *Technology, Society, Sustainability*, ed. Lech Zacher (Springer, 2016), 175–83.
3. Preservation of species and some of the landscapes may be an exception; however, few of us would want such preservation to go untamed and destroy civilization.

FROM THE CHAIR

Notoriety for Machine Ethics?

Thomas M. Powers
UNIVERSITY OF DELAWARE

It has been said that there is "no such thing as bad publicity." The origin of this claim may give us a clue as to the sense in which it ought to be taken. If P. T. Barnum is the source, we might well draw the lesson that even bad publicity attracts the public's attention—and maybe its money. If Oscar Wilde is the source, perhaps bad publicity is a psychological antidote to the insult of being overlooked or ignored.

Surely, the claim that there is *no* bad publicity must be false, if taken literally, but what exactly is the nugget of wisdom in this falsehood? The field of machine ethics could soon find out, since last year brought a lot of publicity to the field, and much of it seemed to be bad. No fewer than nine articles appeared in just two U.S. daily newspapers—the *New York Times* and the *Washington Post*—on topics that could indicate crises for machine ethics: driverless cars, autonomous weapons, sex robots, artificial general intelligence or even superintelligence, and computerized human enhancement, replacement, or perhaps extinction. Reading these texts gives one the sense that we are riding a technological juggernaut that is just about to veer out of control. And though most ethicists love a good puzzle, the media articles go beyond that; they leave the feeling that machine ethicists are in no way positioned to address the impending problems that these technologies would bring. One searches in vain in these articles for positive mention of approaches to controlling the technological juggernaut through advances in machine learning, logic programming, cognitive science, or formal approaches to ethical theory. Nothing about deontic logic or non-monotonic logic, and no mention of formal work on agency and action.

Now, in some sense, maybe the news was not bad for machine ethics. After all, the field is clearly needed—that is one upshot of reading these articles together. Some philosophers were even featured, or at least quoted, and a light was shown upon issues that researchers in machine ethics have been worrying about for some time. In general, most philosophers would agree, research topics in philosophy get very little attention from the popular media (not to mention the occasional belittling of philosophy from politicians and other shapers of public opinion). Maybe this recent attention might start to make up for years of benign neglect and malign contempt.

Two reasons for the attention may be money and fame. Elon Musk has donated $10 million for research into questions of machine ethics (and related AI issues), and leading AI researchers such as Stuart Russell have enlisted the help of other famous scientists (Stephen Hawking among them) to publicize the aforementioned crises. A new organization—the "Future of Life Institute" led by Max Tegmark—has become a focal point for publicity by publishing two open letters—one for beneficial AI, and the other against autonomous weapons. The second letter itself generated much publicity.

I can think of three reasons why this "bad" publicity might be good for philosophical approaches to machines ethics, at least indirectly. First, it is possible that scientists and engineers will finally recognize that theirs are not the only knowledge-producing disciplines—and in particular, that reasoning about how technology ought to behave is itself an intellectually serious enterprise. Second, there is the possibility that the funders of AI and robotics research (e.g., DARPA and the NSF in the U.S.) will give much more support to philosophy graduate programs where the bases of the study of machine ethics are taught—especially to departments that take an "ecumenical" approach to the ethics of technology by combing philosophy of mind, logic, formal studies in ethics, the philosophy of computer science, and the philosophy of technology. Third, there is the possibility that philosophers, computer scientists, and robotics engineers will begin to undertake collaborative research on a large scale. I don't mean, here, that philosophers will be added, at the last minute, to robotics and AI grant proposals to make it seem like they have "broader impact." No; for serious progress to be made in machine ethics, philosophy will have to be taken seriously as a partner discipline to the sciences and engineering.

Despite these reasons to be optimistic about the sudden "notoriety" of machine ethics, there are some concerns with this development. It is worth noting that, consistent with journalistic intentions, these popular media articles are sometimes fear-mongering and rarely careful about exploring substantive issues. They raise more questions than they answer—perhaps this is a settled trait of the genre. To the extent that AI and robotics researchers are feeding the media interest, one wonders if there might not be an attempt here to play favorites with particular research programs in AI, or to raise issues now so that AI funding isn't hurt later. One result of highly publicized worries about the trajectory of AI may even be an increase in funding for AI, precisely to head off the worse-case-scenarios.

While there are positive and negative aspects to this new notoriety of machine ethics, there is also (as with most crises) an opportunity: to make the case now to academic deans, provosts, and other higher-education administrators, and to program officers of funding agencies, that the time for machine ethics is upon us.

FEATURED ARTICLE

Moving Back on Top of the Wave

Luciano Floridi
OXFORD INTERNET INSTITUTE, UNIVERSITY OF OXFORD,
LUCIANO.FLORIDI@OII.OX.AC.UK

The history of philosophy looks a bit like a sine wave (or a roller coaster, if you prefer). It goes up and then down, up again, and then down. The ups, the crests of the wave, are the innovative periods, when we deal with *philosophical problems*. These are the periods when philosophy is engaged with open and fundamental problems in relation to its own time. Once successful, philosophy falls in love with its own image, which is admittedly beautiful and attractive to any speculative mind. And, like Narcissus, it drowns, unable to leave the beauty of its reflection. These downs, the troughs of the wave, are the scholastic periods, when we deal with *philosophers' problems*. In my research, I try to show that the information revolution is a great opportunity to renovate philosophy and climb up again on a new crest.[1] Academic philosophy is definitely too narcissistic today. It would be very healthy to make it look at the world, instead of itself. And the world itself is in great need of philosophical understanding and design of new ideas. We need philosophy on board while we are creating our information societies, shaping the new

digital environments in which billions of people will spend increasingly amount of time, and as we re-think what I like to call *the human project*. But what kind of philosophy? It seems to me that it should be a philosophy engaged with the profound transformations caused by information and communication technologies (ICTs). This is problematic not least because philosophy has never had a very friendly relation with technology in general. Yet it seems that technology has progressively increased its role in our lives, and philosophy should come to terms with this fact. Today, no aspect of human life is being left untouched by ICTs: education, work, conflicts, social relations and interactions, entertainment, governance, politics, art, literature, mass media, law, health, business, industry, communication, science . . . it is hard to think of anything that is not being deeply transformed or widely redefined by the information revolution.[2] This means that old philosophical problems are being upgraded; think of issues about personal identity, memory, the nature of knowledge, the foundations of science, fundamental rights, and so forth. And new philosophical problems acquire prominence: What is the nature of information? What is the new morphology of power in a mature information society? Can we reconcile human freedom and its predictability with smart machines? What balance can be found between privacy, security, and freedom of speech? These are just a few examples among many. Clearly, the philosophy of information is not a matter of developing a philosophy of the next gadget or new app. It is about engaging with the deep transformations caused by ICTs in how we understand the world, hence in our epistemology and metaphysics; in how we make sense of it, hence in our semantics; in how we conceptualize ourselves, and what we think we can be or become, hence in our theories of education, identity, and in our philosophy of mind and our philosophical anthropology; in how we interact with each other, how we manage and shape collaborative and conflicting relations, and how we may construct the society we want, hence in our ethical, socio-economic, political, and legal thinking. ICTs and the infosphere they are creating are providing the new environments in which we live, think, and interact. Surely, this is what philosophy should try to understand and help to shape properly. So, ultimately, it is a question of ethics or, as I prefer to put it, of *e-nvironmental ethics*. It is time to move back on top of the wave. To do so, we must adopt a post-analytic-continental divide perspective and regain the right balance between *control* and *power*. Allow me to explain this with two caricatures.

On the one hand, analytic philosophy, broadly understood, excels at controlling the philosophical discourse. An exact vocabulary, logic, formal distinctions, scientific information, empirical or thought experiments, mathematical formulations, statistical data, cogent and coherent arguments, a piecemeal and inferential way of discussing problems . . . these are all ways in which analytic philosophy can exercise a high degree of control over a philosophical topic. The "but" is represented by the risk that so much technical control may be exercised over nothing, minutiae and irrelevancies, what I called, above, philosophers' problems. John Locke once remarked that logicians keep sharpening their pens but never write. It seems an apt description of some analytic philosophy, entirely engrossed in its internal discourse. It may get worse, if the degree of what can be controlled ends up determining the scope of what is worth investigating philosophically.

On the other hand, continental philosophy, understood in an equally broad sense, excels at enriching the philosophical discourse with powerful thoughts. An evocative vocabulary, rhetoric, scholarly references, literature, art, poetry, socio-political analyses, historical facts and interpretations, a more narrative style, existential and religious approaches to problems . . . these are all ways in which continental philosophy can add profound, powerful contents to a philosophical topic. The other "but" is represented by the risk that so much rich and powerful content may spill all over the place and be vague, confusing, incoherent, and sometimes downright preposterous. In this case too, it may get worse, if the power of the content ends up promoting irrationality and an irritated impatience towards logic, or anti-scientific views, relativism, obscurantism, and an oracular philosophy.

As a famous slogan of Pirelli (the tire company) reminds us, "power is nothing without control," yet so is control without power. The best philosophy (the one you find on the crests of the sine wave) has always combined a high degree of rational control with very powerful ideas. And this is what I hope a post-analytic-continental divide perspective may regain. It is certainly what we need today. As for the philosophy of information, I can only hope that it will mature into a first philosophy. Anything less and it will have failed in its task of providing us with the powerful and controlled ideas that we need to shape and make sense of the human project today. This is the last remark I wish to make.

In the twenty-first century, we need to approach philosophy from a design perspective.[3] Philosophy deals with open problems, that is, problems that are constrained by facts and figures but ultimately solved by neither.[4] Open problems are such that two people could be informed, rational, and not stubborn about them and still disagree about their acceptable solutions. We move forward when we can design (not invent, not discover) ways in which open problems can be solved satisfactorily. Yet opting for a metaphysical approach means forgetting the Kantian lesson and falling into the illusion that we can talk about reality in itself, without accepting any level of abstraction, that is, any interface through which questions may become sensibly answerable.[5] With an analogy, it would be as if two people disagreed on the value of a secondhand automobile without even being willing to accept that such value must be given within a framework of considerations (financial value, historical value, emotional value, running-cost effectiveness value, and so forth). Philosophical questions, precisely because they are philosophical, are intrinsically open to disagreement, and hence subject to more than one answer. Even in mathematics we are used to equations that have more than one solution, infinitely many solutions, no solutions at all, or solutions that can only be approximated. Philosophical problems are not different. If we wish to find their solutions, we must drop any absolute metaphysics in favor of a reasonable approach to clarify the level of abstraction at which the question that is being

asked is actually answerable, in some cases agree on further constraints, and ultimately accept that there may be many solutions, some preferable to others, depending on the purpose for which a level of abstraction has been privileged. Plenty of philosophers' problems fail to be clear about all this and become sources of endless diatribes, turning into cottage industries and scholastic monopolies.

The appearance of information societies is related to a major shift in our philosophy and the appearance of a philosophy of information, not unlike historical events and philosophical ideas were coupled in the Enlightenment, for example.[6] This macroscopic shift has generated attempts to explain what is happening under our eyes. We sense a deep and widespread transformation. So fashionable ideas, such as "singularity," "posthumanism," "cyberculture," are not necessarily mere philosophical snake oil. In some cases they can be evidence of growing pains: we are confused, in search of new certainties, in need of meaningful frameworks, and so we resort to the ageless practice of telling stories, some reassuring, some scary, all fanciful. What we need to do is to develop a robust, controlled, and rich philosophy of our time for our time. This should not be left to bizarre speculations, but it cannot be delegated to "scientists and IT boffins" either. They usually do not deal with open problems and with the design of the ideas necessary to answer them, with the ultimate goal of making sense and shaping the world. And when they do, they are simply stepping into a philosophical debate, often rather naively. We need experts in conceptual design and multisolvability. In other words, we need philosophers.

NOTES

1. See, in particular, the tetralogy for OUP on the foundations of the philosophy of information. The first and the second volume have already been published: Luciano Floridi, *The Philosophy of Information* (Oxford: Oxford University Press, 2010); Floridi, *The Ethics of Information* (Oxford: Oxford University Press, 2013). The next two volumes are in preparation and it will tame me a few more years to complete them: (Floridi, *The Logic of Information* (to be submitted to Oxford University Press); Floridi, *The Politics of Information* (to be submitted to Oxford University Press). Stay tuned.
2. For a simple overview see Floridi, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality* (Oxford: Oxford University Press, 2014).
3. Floridi, "A Defence of Constructionism: Philosophy as Conceptual Engineering," *Metaphilosophy* 42, no. 3 (2011): 282–304.
4. Floridi, "What Is a Philosophical Question?," *Metaphilosophy* 44, no. 3 (2013): 195–221.
5. Floridi, "The Method of Levels of Abstraction," *Minds and Machines* 18, no. 3 (2008): 303–29.
6. Floridi, "Turing's Three Philosophical Lessons and the Philosophy of Information," *Philosophical Transactions of the Royal Society A* 370 (2012): 3536–42.

ARTICLES

Ethics of Entropy

Martin Flament Fultot
PARIS IV SORBONNE / SND / CNRS

INTRODUCTION

Luciano Floridi reinterprets and re-ontologizes our world informationally. That part of his theory may (or may not) work, but what matters for this paper's topic is that when it comes to defining what the value of Being is, his informational-ontological interpretation is based on order, organization, and structure. Therefore, there is a common ground between his interpretation and the way modern thermodynamics formalizes the concept of order. Floridi proposes to think of Good as a qualitative order and Evil as its absence or entropy. However, the kind of entities that are of importance to us for our judgments and interventions as agents are ordered and thus valuable because they exist far from equilibrium. In this paper I shall attempt to establish that far-from-equilibrium systems attain ever increasing degrees of order at the cost of faster entropy production. Yet, inversely, by promoting an increase in entropy production, more complex and ordered forms emerge on Earth. Entropy production and order are thus complementary; they imply each other reciprocally. By promoting Evil in Floridi's sense, Good happens lawfully because order is nature's favorite way of producing entropy. In short, moving against entropy only creates more entropy.

I. THE VALUE OF ONTOLOGICAL INFORMATION

Floridi's Informational Ethics presents three highly attractive features. The first one is that it develops a theory of macroethics. The second one is that it grounds the origin of value in Being, that is, beyond humans and even living creatures. And the third one, on which I will focus mostly, is that Being is defined in terms of information and *entropy*.

Floridi starts from the observation that the space where ethically relevant human behavior takes place is being completely and irreversibly transformed by the development and diffusion of information technologies. This particular kind of transformation or "re-ontologization," as he conceptualizes it, affects "the whole realm of reality," thus requiring a macroethics approach.[1] It may have been more appropriate to talk about *holo*ethics, rather than "macro," since it is concerned with how information reconfigures human behavior holistically and globally as opposed to locally and individually. In other words, the space of ethical events becomes, in the new "infosphere," completely interconnected. Thus, single—ethically relevant—events need to conform to norms that target value in its totality.

Floridi's macroethics ascribes value not to humans nor living creatures as such but to Being itself. Good corresponds to Being, and Evil corresponds to the suppression or the degradation of Being. As a consequence, Floridi's radical approach makes room for ethical concerns about inanimate things such as rocks. This is understandable since Information

Ethics is, by definition, concerned with information, and the concept of information applies to a lot more than human beings or living creatures. More specifically, however, Floridi makes a move from the common idea of information as, say, a message delivering content such as "tomorrow it will rain" to an *ontological* conception where entities are re-interpreted informationally. The move seems justified by the polarized axiological scale shared by information and Being, where the latter is clearly a value when compared to nothingness, and the former stands as a value when compared to lack of information. But for information to count as a value *qua information*, it needs to be understood semantically, that is, not so much as *in*formation—despite the fact that Floridi's theory revolves around that concept—but more simply as form, order, structure. I thus assume that Floridi's "informational" interpretation of ontological Being is simply a structural interpretation, with order or organization being qualitatively opposed to randomness or "mixed-upness."[2]

In this way, Floridi's macroethics approach establishes a normativity that bestows intrinsic value on Being:

> Information Ethics holds that *being*/information has an intrinsic worthiness. It substantiates this position by recognizing that any informational entity has a *Spinozian* right to persist in its own status, and a *Constructionist* right to flourish, i.e., to improve and enrich its existence and essence.[3]

Hence, according to Information Ethics, protecting and improving Being constitutes the absolute norm. Now, with Being defined in terms of form, its polar opposite, we have mentioned, consists in lack of form or organization. These notions are intuitive and naturally understandable. Yet Floridi establishes another link, through the notion of information, with entropy. Entropy is a thermodynamical concept that was mathematically related to that of information by Claude Shannon, thanks to their both being defined in terms of order and randomness.

The relationship between the mathematical formalisms of entropy and information and Floridi's own ontological or metaphysical take on them is tricky, though.[4] Indeed, Floridi insists "emphatically" that although his own interpretation and the mathematical formalisms are related, they are not the same. The reason for this is that information theory is silent about content or meaning. In thermodynamic terms, that translates into entropy being randomness as opposed to order. However, the *qualitative structure* of an ordered state is unspecified by thermodynamics. This can be problematic as it challenges the possibility of a graded normative axiological scale. For instance, two entities may contain the same *quantity* of information as measured by Shannon's formula, yet differ qualitatively, as in having different shapes. Do they have identical moral value? Do they deserve equal respect? After all, as Schrödinger said, "any calorie is worth as much as any other calorie."[5]

Another difficulty for Floridi's theory of information as constituting the fundamental value comes from the sheer existence of the unilateral arrow of thermodynamic processes. The second law of thermodynamics implies that when there is a potential gradient between two systems, A and B, such that A has a higher level of order, then in time, order will be degraded until A and B are in equilibrium. The typical example is that of heat flowing inevitably from a hotter body (a source) towards a colder body (a sink), thereby dissipating free energy, i.e., reducing the overall amount of order. From the globally encompassing perspective of macroethics, this appears to be problematic since having information on planet Earth comes at the price of degrading the Sun's own informational state. Moreover, as I will show in the next sections, the increase in Earth's information entails an *ever faster* rate of solar informational degradation. The problem for Floridi's theory of ethics is that this implies that the Earth and all its inhabitants as informational entities are actually doing the work of Evil, defined ontologically as the increase in entropy. The Sun embodies more free energy than the Earth; therefore, it should have more value. Protecting the Sun's integrity against the entropic action of the Earth should be the norm.

II. FAR-FROM-EQUILIBRIUM SYSTEMS, ORDER, AND ENTROPY

It is surprising that, even though Floridi is well aware of the second law of thermodynamics and the fact that informational entities in one way or another will generate entropy in order to persist in their Being, his theory lacks a conceptual treatment of the crucial case of systems that exist far from thermodynamic equilibrium. Yet these systems present an important obstacle to his view that Being has intrinsic and fundamental value. To see this, consider the following proverbial far-from-equilibrium example: the Rayleigh-Bénard experiment (henceforth R-B).

R-B consists in heating from below a shallow layer of viscous fluid contained in a recipient (think of oil in a circular frying pan.[6]) This creates a uniform potential energy gradient between its bottom temperature and the surface's temperature at the top of the fluid. Following the second law of thermodynamics, the energy gradient operates as a vector, i.e., a force with a direction, so that the fluid fights the asymmetrical concentration of energy at the bottom (order) by transferring the heat towards the top, thereby restoring thermodynamic equilibrium with the surroundings (entropy). Below a given magnitude for the difference of temperature between the bottom and the top, the transfer occurs by conduction, i.e., stochastic collisions between the moving particles that constitute the fluid. The fluid is thus *disordered* and disorganized under this regime. Under Floridi's account, the fluid has little being and therefore value. However, when the magnitude of the potential exceeds a given threshold, a new regular pattern of organization emerges from the interaction between the particles in the fluid. Typically, in a circular recipient, the patterns are constituted by hexagonal convection cells visible to the naked eye. Each cell consists of hundreds of millions of molecules moving in a coordinate fashion.[7] Now the fluid is in a *dynamically ordered* state and, interestingly, this order or organization constitutes a *pattern*, i.e., it has a shape, a form. How can this qualitative aspect of organization be understood given the quantitative formalism of thermodynamics?

The answer to this question lies in Ilya Prigogine's work on far-from-equilibrium systems.[8] The main insight is that the emergence of ordered patterns is due precisely to the requirement to dissipate the free energy pumped into the system from the outside in conformity to the second law of thermodynamics. Concretely, in R-B, the emergence of the very specific pattern of hexagonal convection cells corresponds to an *optimal* configuration of energy flows within the fluid given the magnitude of the potential and the boundary conditions. The latter include the circular shape of the recipient and constraints such as surface tension. Hexagonal shapes distribute the cells so as to collectively *maximize* their dissipative surface, which translates into higher entropy generation.

So here we can see a *qualitative* form of order responding to thermodynamical quantitative principles. The magnitude of the potential field as well as the rate of entropy production vary continuously as a simple scalar; the force simply becomes stronger and entropy increases. However, the state of the system transitions *qualitatively* from a disordered state to an organized state according to a very specific pattern, which in this case is geometrical. The qualitative aspect serves a quantitative function of maximizing entropy production in response to the asymmetric conditions under which free energy is being pumped into the system. We can see that, in a sense, R-B shows how, in far-from-equilibrium systems, information in Floridi's sense, or simply qualitative order, is not exactly an *intrinsic* value, but rather a *functional* value. Hexagonal cells, as a qualitative ordered Being, have the value of optimizing a natural function, and the function is to conform to the second law of thermodynamics by always creating at least as much entropy as the order is being added.

III. MAXIMUM ENTROPY PRODUCTION RATE

Yet it seems odd to claim that the second law of thermodynamics is responsible for the spontaneous emergence of order. After all, in R-B, order helps maximize a function, yet the second law doesn't predict any such helping. Why, it may be asked, doesn't conduction remain the heat transfer regime although simply at a faster rate, proportional to the increase in the potential gradient? The answer is that the second law is only one part of the principle of maximum entropy, the other being, precisely, the maximization function. Indeed, the second law states that, on the long term and on average, entropy tends to increase. In other words, entropy in a system will become maximal given *enough time*. But it doesn't say anything about how entropy is maximized.[9] However, several observations have led many independent researchers to the conclusion that the law of entropy production should state rather that the system will tend to disorder at the *fastest rate* (given the constraints).[10] With this extension I will refer, following Rod Swenson, to the Law of Maximum Entropy Production (LMEP).[11]

LMEP can be observed even in systems not far from equilibrium. Swenson and Turvey illustrate this by a simple experiment in which an adiabatically sealed chamber is divided by an adiabatic wall into two compartments, each filled with an equal quantity of the same gas although at different temperatures.[12] There is thus a potential field between both compartments with the hottest holding more order or information in Floridi's metaphysical sense. If a hole is opened on the dividing wall, a channel allows heat to proceed from the hotter to the colder chamber until equilibrium is reached and entropy is maximized, as stated by the second law. If a second hole is opened such that it conducts heat at a different rate from the first one, then depending on the constraints and the configuration of the holes, the system will always distribute the flows along the holes in the optimal way. For instance, if hole 2 can drain some heat before it is all drained through hole 1, then some heat will be drained through hole 2 also. In other words, free energy always seeks to exploit the optimal paths to its own dissipation. The same process can be observed in the mundane setting of a cabin in the woods heated from the inside, where heat will drain to the surroundings through the fastest configuration of windows, doors, and other openings.

Back to the R-B case which is far-from-equilibrium, Swenson and Turvey show that the emergence of the convection cells is inevitable due to the opportunistic exploitation of the configurations that tend to optimize the rate of entropy production. The threshold corresponds to the minimal magnitude of force that will sustain the dynamical ordered state. With enough free energy available within the fluid, a new, non-random configuration becomes possible, and because of LMEP, that configuration will be favored and stabilized "as soon as it gets the chance."[13]

The point about "getting the chance" deserves a brief pause. The formation of the ordered regime *takes time*. It is a search the system undergoes, facilitated by the increased amount of kinetic energy produced by the potential field. Akin to a selection process with winner-takes-all rules, the formation of hexagonal cells (1) occurs in time by progressively entraining more and more molecules in the macroscopic motion and (2) is imperfect, perturbed by random fluctuations and many other factors (constraints). These facts allow us to foresee already that what happens relatively quickly and with success in a simple R-B setting, i.e., the establishment of an optimal regime of free energy degradation, will become dramatically more fluctuating, complex, and hence time-consuming in the case of a setting as wide as the Sun-Earth system.[14] This, of course, will have crucial consequences for macroethics.

IV. THE EARTH AS A GLOBAL FAR-FROM-EQUILIBRIUM DISSIPATIVE STRUCTURE

There is life on Earth. Any theory of macroethics worth its salt must have the resources to give a central role to that simple fact. Floridi's axiological scale leaves room for such a role, thanks to the overridability of different levels of informational value. However, this is somewhat unsatisfactory.[15] One would have expected from an informational macroethics, which is based on a technical ontological framework, something like an equation, a formula to *measure* worth, and thus, in some way, to mechanize morals as Alan Turing's famous formalism mechanized intelligence. The problem is that, as stated above, the quantity of information cannot serve as a gauge of value since two very different entities may contain the

same amount of information. For instance, there may be a configuration of some amount of potatoes that is quantitatively equivalent to, say, an innocent child. What is needed is a criterion able to locate qualitative differences on an axiological scale where they can make a difference.

I suggest that the points raised above about far-from-equilibrium systems and LMEP are in a good position to ground such an axiological scale. Consider the question: What *difference* does life make? To begin with, living organisms are far-from-equilibrium systems.[16] All the metabolic and adaptive processes of living things are sustained by a continuous energy flow, and their ordered patterns are self-organizing, i.e., dynamically maintained, as in R-B. Existing far from thermodynamic equilibrium means existing beyond the thresholds in magnitude mentioned above. This non-linearity implies that the rate of entropy production in living creatures must differ from that of a non-living entity under the same conditions. This implication has been developed by Ulanowicz R. E., Hannon B. M., who hypothesized that it could be proven that forests, for instance, produce more entropy than the desert under the same electromagnetic field.[17] Meysman & Bruers recently tested the hypothesis that "living communities augment the rate of entropy production over what would be found in the absence of biota, all other things being equal."[18] Using an ecologically inspired model of entropy production in food webs with predators and preys, they showed that the hypothesis holds every time.

The consequence is that far-from-equilibrium systems such as living creatures on Earth operate according to an *adaptive* principle. In other words, the structures and dynamic patterns that emerge when crossing critical thresholds are such that they tend to optimize entropy production *given the constraints*. This means that, for a given potential gradient *P* and a set of constraints *C*, there is only a restricted set of patterns—perhaps even a singleton—able to optimize the rate of entropy production. In R-B, as we saw, hexagonal cells do the job, but there is a set of geometrical patterns and dynamic organizations different from hexagons that may possess the same amount of information as the fluid yet dissipate the potential at a slower rate under those same conditions. Therefore, one can assume that LMEP working as a thermodynamical selection principle at the planetary level is ensuring that the living forms that emerge in time are coordinated and increasingly evolving towards higher rates of global dissipation of the geo-cosmic potential constituted mainly by the electromagnetic radiation from the Sun in which the Earth is immersed. For instance, chlorophyll is particularly efficient in its capacity to absorb blue and red light, thanks to the structural complementarity between the spatial distribution of its p-orbitals and the wavelengths of blue and red light. In this way, we can see value as depending on the *fit* between the qualitative aspects of living order and the qualitative aspects of the geo-comic potential taken as dynamic patterns. Moreover, visible light corresponds to more than half of the total solar emission, implying a massive free energy influx to be degraded.[19] There is value in the capacity of photosynthetic organisms to contribute drastically to the degradation of this tremendous amount of free energy.

The idea that the Earth is operating holistically as a maximizer of entropy production is increasingly gaining adepts. Special attention is paid to the link between LMEP or similar characterizations of entropy maximization and evolution.[20] Recently, Martyushev and Seleznev have responded to some claims that LMEP doesn't generalize well and that it shouldn't be considered a law of thermodynamics.[21] The authors show that such conclusions are based on the wrong application of LMEP's predictions without properly assessing some key restrictions. One of those restrictions is, I think, of particular importance for the present discussion about macroethics as it concerns the time delays in thermodynamic processes. As I have already mentioned above, the self-organized emergence of new order in a far-from-equilibrium system is a time-dependent process akin to searching. This means that from the onset of a supra-threshold energetical inflow to the actual assembly of an optimal or near optimal dissipative pattern, the system goes through transient heuristic stages analogous to trial and error. During all that time, the system is obviously performing *sub-optimally*. Yet, one could argue that, even during the searching period, the system is still performing optimally, since the very state of the system during the long process of assembly counts as a *constraint* and, thus, given *that* constraint, the system is still "doing its best." Although such a view sounds Panglossian, it may actually present the advantage of reconciling the apparent unlawful normativity of ethics with the lawful determinism of LMEP. Indeed, at some level of analysis and from a local vantage point, the system is performing sub-optimally and, hence, something like a norm may help seek ways to improve the situation, for instance, by removing or changing the constraints that keep the system from producing entropy at higher rates. From another point of view, however, the system is working optimally given the constraints and in perfect agreement with lawful determinism. A theory of macroethics capable of naturalizing normativity would present a very strong advantage over other alternatives.

Another restriction linked to the former that needs to be taken into account concerns *local maxima*. It would seem that value based on the optimization of entropy production should go against all rationality concerning viability and even common sense. After all, as Floridi points out, entropy is metaphysically tantamount to *Evil*. Would this imply that forests have to be burned as fast as possible, for instance? Certainly not, because every kind of direct, local application of entropy production might contribute to trapping the whole Earth in a local maximum. Consider petroleum, for instance. In a relatively small and homogeneous system such as R-B, the time delay between energetical transactions is very short. If, say, some small regions of the fluid are hotter than other regions, because of the fast moving particles and transmissive medium, such small local gradients are very short lived, and the global bottom-up force overpowers them completely, driving the system very quickly to the formation of hexagonal patterns. In a system as vast and complex as the Earth, on the other hand, the local formation and maintenance of gradients is ubiquitous and inevitable. Petroleum represents one such local gradient, which embodies in its chemical structure a significant amount of free energy. This free energy took thousands of years to undergo a transformation from

solar energy to living tissue and then to petroleum, and at every step, entropy was produced. However, until mankind started exploiting petroleum globally and industrially, this source of free energy was sub-optimally unaffected, thus missing a great opportunity for entropy production. Does this mean that we should go ahead and deplete the source at the fastest possible rate as we are currently doing, against all the advice from ecologists? Probably not, but not for the reasons ecologists think. The Earth is still undergoing its transformation towards an optimal regime of geo-cosmic energy degradation. This ongoing transformation has been taking millions of years, and it is not likely to stop anytime soon. However, local potentials such as petroleum and the other so-called fossil fuels may present an opportunity for mankind as part of the Earth system to transition into a higher level of dynamic order, which might improve the rate of *solar* energy degradation. In other words, consuming the local potentials without taking into account the global field might *transiently* increase overall entropy production (and therefore terrestrial order), yet as soon as the local source is depleted, the Earth would go back to its earlier regime, having missed an opportunity to move closer to the optimal form. This would be tantamount to destroying your car in order to increase entropy immediately and locally instead of keeping your car and using it to go every day to the supermarket and deplete the higher energy sources present there.

V. CONCLUSION. IS ENTROPY ETHICS AN ETHICS OF EVIL?

I have tried to argue that although Floridi's looks like a move in the right direction to reconceptualize ethics not only as a holistic foundation of value, but also as encompassing more than just humans or living creatures, it falls short of considering all the aspects related to Being. If I am right, Floridi's appeal to a notion such as entropy and ontological information is fatally incomplete, since, by deciding to "emphatically" detach those notions from their thermodynamical equivalents, he misses an ontologically crucial link between entropy and order. It is crucial in that it shows that, by relocating intrinsic value not on Being but on entropy production, we can still obtain the astonishingly paradoxical result that Being is protected and promoted as Floridi's own Information Ethics requires.

In this way I have challenged Floridi's view, suggesting that contemporary thermodynamical research presents us with ineluctable facts that force us to radically reconsider our axiological principles. Floridi's information/entropy dichotomy doesn't seem to make room for far-from-equilibrium phenomena where both are entangled and complementary. It is not possible to identify entropy with Evil when the value of order happens to be contingent on its capacity to optimize entropy production given the constraints. The case of the whole Earth's evolution as a far-from-equilibrium system makes this point conspicuous.

Considering order as the intrinsic source of value has the disadvantage that we cannot establish a non-arbitrary axiological scale. However, if accelerating (global) entropy production becomes the norm, we can see that those terrestrial forms that our common sense already values most get automatically promoted axiologically because they coincide with the forms that tend to contribute to the production of entropy at optimal rates given the specific context established by the potential field in which the Earth is immersed. If the search for Good is the search for the shortest paths to global energetical degradation, then life and mankind's extremely complex cultures and technological achievements get instantly promoted as optimal media for that end. That is because those kinds of entities and processes happen to fit better the structure of the geo-cosmic potential while satisfying the constraints.

In addition, despite being based on a deterministic physical law, the approach presented here leaves plenty of room for human intervention, normativity, and, therefore, responsibility. Indeed, the search for the optimal forms capable of dissipating the geo-cosmic potential at the fastest rate is extremely long and haunted by local maxima where the Earth can get trapped at every moment. Humans are the only entities in the system that have access to distal, higher-order constraints that modulate the overall rate of entropy production at least at the scale of the Sun-Earth system. Yet, because they are also constantly embedded in local gradients, humans also have a tendency to favor the depletion of those more proximal gradients, and the tendency is becoming exponentially stronger with trends such as technological improvement and overcrowding. For this reason, a macroethics theory is needed more than ever, yet it needs to embrace all aspects of reality, including, ironically, entropy itself.

NOTES

1. L. Floridi, "Understanding Information Ethics," *APA Newsletter on Philosophy and Computers* 7, no. 1 (2007): 8.
2. L. Floridi, *The Ethics of Information* (Oxford: Oxford University Press, 2013), 66.
3. Floridi, "Understanding Information Ethics," 9.
4. Xiaohong Wang, Jian Wang, Kun Zhao, and Chaolin Wang, "Increase or Decrease of Entropy: To Construct a More Universal Macroethics," *APA Newsletter on Philosophy and Computers* 14, no. 2 (2015): 32–36.
5. E. Schrödinger, *What Is Life?* (Cambridge, UK: Cambridge University Press, 1944).
6. J. A. S. Kelso, "Dynamic Patterns: The Self-Organization of Brain and Behavior" (Cambridge, MA: MIT Press, 1995).
7. R. Swenson and M. T. Turvey, "Thermodynamic Reasons for Perception-Action Cycles," *Ecological Psychology* 3, no. 4 (1991): 317–48.
8. I. Prigogine, *Introduction to the Thermodynamics of Irreversible Processes* (New York, NY: John Wiley, 1967).
9. H. T. Odum, *Ecological and General Systems: An Introduction to Systems Ecology* (Colorado University Press, 1994)
10. Swenson and Turvey, "Thermodynamic Reasons for Perception-Action Cycles."
11. R. Swenson, "Emergent Attractors and the Law of Maximum Entropy Production: Foundations to a Theory of General Evolution," *Systems Research* 6 (1989): 187–97. There are several independent developments of the same principle (or very similar) under slightly different names, e.g., Maximum Power principle (H. T. Odum, *Ecological and General Systems: An Introduction to Systems Ecology* [Colorado University Press, 1994]), Maximum Entropy Production principle (L. M. Martyushev and V. D. Seleznev, "The Restrictions of the Maximum Entropy Production Principle," *Physica A: Statistical Mechanics and Its Applications* 410 [2014]: 17–21), and even the Principle of Least Action (V. R. Kaila and

A. Annila, "Natural Selection for Least Action," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 464, no. 2099 (2008): 3055–70).

12. Swenson and Turvey, "Thermodynamic Reasons for Perception-Action Cycles."
13. Ibid., 335.
14. I. van Rooij, "Self-Organization Takes Time Too," *Topics in Cognitive Science* 4, no. 1 (2012): 63–71.
15. See Floridi's object-programming oriented inspired model of moral action. Floridi, *The Ethics of Information*, 103–109.
16. H. J. Morowitz, "Energy Flow in Biology: Biological Organization As a Problem in Thermal Physics" (Woodbridge, CT: Ox Bow Press, 1968).
17. R. E. Ulanowicz and B. M. Hannon, "Life and the Production of Entropy," *Proc. R. Soc. Lond. B* 232 (1987): 181–92.
18. F. J. Meysman and S. Bruers, "Ecosystem Functioning and Maximum Entropy Production: A Quantitative Test of Hypotheses," *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365, no. 1545 (2010): 1405.
19. Swenson and Turvey, "Thermodynamic Reasons for Perception-Action Cycles."
20. E.g., J. S. Kirkaldy, "Thermodynamics of Terrestrial Evolution," *Biophysical Journal* 5, no. 6 (1965): 965; Kaila and Annila, "Natural Selection for Least Action"; Swenson and Turvey, "Thermodynamic Reasons for Perception-Action Cycles"; F. J. Meysman and S. Bruers, "Ecosystem Functioning and Maximum Entropy Production: A Quantitative Test of Hypotheses," *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365, no. 1545 (2010): 1405–16; E. D. Schneider and J. J. Kay, "Life As a Manifestation of the Second Law of Thermodynamics," *Math. Comput. Model.* 19 (1994): 25–48; Martyushev and Seleznev, "The Restrictions of the Maximum Entropy Production Principle; and the references therein.
21. L. M. Martyushev and V. D. Seleznev, "The Restrictions of the Maximum Entropy Production Principle," 17–21.

REFERENCES

Floridi, L. "Understanding Information Ethics." *APA Newsletter on Philosophy and Computers* 7, no. 1 (2007): 3–12.

Floridi, L. *The Ethics of Information*. Oxford: Oxford University Press, 2013.

Kaila, V. R., and A. Annila. "Natural Selection for Least Action." In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 464, no. 2099 (2008): 3055–70. The Royal Society.

Kelso, J. A. S. "Dynamic Patterns: The Self-Organization of Brain and Behavior." Cambridge, MA: MIT Press, 1995.

Kirkaldy, J. S. "Thermodynamics of Terrestrial Evolution." *Biophysical Journal* 5, no. 6 (1965): 965.

Martyushev, L. M., and V. D. Seleznev. "The Restrictions of the Maximum Entropy Production Principle." *Physica A: Statistical Mechanics and Its Applications* 410 (2014): 17–21.

Meysman, F. J., and S. Bruers. "Ecosystem Functioning and Maximum Entropy Production: A Quantitative Test of Hypotheses." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365, no. 1545 (2010): 1405–16.

Morowitz H. J. "Energy Flow in Biology: Biological Organization As a Problem in Thermal Physics." Woodbridge, CT: Ox Bow Press, 1968.

Odum, H. T. *Ecological and General Systems: An Introduction to Systems Ecology*. Colorado University Press, 1994.

Prigogine, I. *Introduction to the Thermodynamics of Irreversible Processes*. New York, NY: John Wiley, 1967.

Schneider, E. D., and J. J. Kay. "Life As a Manifestation of the Second Law of Thermodynamics." *Math. Comput. Model.* 19 (1994): 25–48.

Schrödinger, E. *What Is Life?* Cambridge, UK: Cambridge University Press, 1944.

Swenson, R. "Emergent Attractors and the Law of Maximum Entropy Production: Foundations to a Theory of General Evolution." *Systems Research* 6 (1989): 187–97.

Swenson, R., and M. T. Turvey. "Thermodynamic Reasons for Perception-Action Cycles." *Ecological Psychology* 3, no. 4 (1991): 317–48.

Ulanowicz, R. E., and B. M. Hannon. "Life and the Production of Entropy." *Proc. R. Soc. Lond. B* 232 (1987): 181–92.

van Rooij, I. "Self-Organization Takes Time Too." *Topics in Cognitive Science* 4, no. 1 (2012): 63–71.

Xiaohong Wang, Jian Wang, Kun Zhao, and Chaolin Wang. "Increase or Decrease of Entropy: To Construct a More Universal Macroethics." *APA Newsletter on Philosophy and Computers* 14, no. 2 (2015): 32–36.

Response to Floridi on Dangers from AI

Kaj Sotala

FOUNDATIONAL RESEARCH INSTITUTE

In a recent APA newsletter article, Luciano Floridi took issue with what he called the "Church of Singularitarians," a supposed pseudo-religious group painting apocalyptic visions about an AI disaster.[1] Floridi did not explicitly reference any papers or works, so he might only have been making fun—and rightly so—of the many sensationalist headlines that have been in the popular press recently.

However, Floridi's commentary could also be interpreted as referring to the nascent academic field working on AI safety. Floridi references Elon Musk tweeting, "We need to be super careful with AI. Potentially more dangerous than nukes." In the original tweet, this sentence was preceded by "Worth reading Superintelligence by Bostrom." This is a reference to Nick Bostrom's book *Superintelligence*, a recent academic work on risks from AI, written by an Oxford professor of philosophy. Read as a critique of the academic field, Floridi's piece contains some inaccuracies, which I wish to address.

To provide some brief background, there has been ongoing research on the possible risks from advanced AI systems for at least a decade now.[2] Work in this field includes estimates of when we might expect to have AI, analyses of the extent to which AI might be dangerous, and attempts to identify research which could already be performed to make AI safer. While the field remains speculative, it has received coverage in mainstream academia such as by being discussed in the leading AI textbook, *Artificial Intelligence: A Modern Approach*.[3]

The field also received major support in 2015 when the Future of Life Institute published an open letter, "Research Priorities for Robust and Beneficial Artificial Intelligence."[4] This letter was signed by hundreds of academics and industry professionals and included an associated research priorities document which cited many works from the AI risk field. The open letter called attention to the fact that current AI research is solely focused on increasing the capabilities of AI, with there being much less research that would attempt to ensure that AI remains socially beneficial.

Likely, many of the signatories were more worried about possible near-term consequences of AI, rather than the long-term consequences studied by the field that I have been describing. However, the research priorities document addressed both short- and long-term issues and reflected the consensus of a number of researchers who collaborated on the open letter.

I will now turn back to Floridi's article. Floridi describes what he calls "three dogmas" of the "Church of Singularitarians":

> First, the creation of some form of artificial superintelligence—a so-called technological singularity—is likely to happen in the foreseeable future. Both the nature of such a superintelligence and the exact timeframe of its arrival are left unspecified, although Singularitarians tend to prefer futures that are conveniently close-enough-to-worry-about but far-enough-not-to-be-around-to-be-proved-wrong.[5]

There is some truth to this characterization in that the AI safety field has no commitment to any exact timeframe for the creation of advanced AI. Most publications on the topic either make no claims about AI timeframes or cite surveys of expert opinion. For example, Bostrom in *Superintelligence* references four different surveys conducted among AI researchers which, combined, gave roughly a 10 percent chance for human-level machine intelligence by the year 2022, a 50 percent chance by 2040, and a 90 percent chance by 2075.[6] Additionally, Bostrom remarks that, in his personal opinion, "the median numbers reported in the expert survey do not have enough probability mass on later arrival dates," and says that a 90 percent chance for human-level machine intelligence by even 2100 seems too high.

Floridi's characterization seems to imply that the AI risk field is choosing its predictions for AI timeframes in a way that allows them to maximize their own publicity. It is true that AI researchers who make public predictions about AI timelines tend to prefer predicting AI within 15 to 25 years, possibly due to similar reasons as Floridi suggests.[7]

However, as discussed, major works in the AI risk field typically do not give such early predictions. A paper by the philosopher David Chalmers considers the possibility of AI "within centuries."[8] *Intelligence Explosion – Evidence and Import*, a chapter in an edited Springer volume,[9] discusses "some considerations for and against . . . [the claim that there] is a substantial chance we will create human-level AI before 2100," and a recent review paper, *Responses to Catastrophic AGI Risk*, discusses the possibility of AI "within the next 20 to 100 years" but adds that "[o]ne must not put excess trust in this time frame."[10]

The AI safety field does not significantly differ from other fields concerned with major risks such as asteroid or pandemic safety. These fields also talk about the relative probabilities of a major asteroid strike or a pandemic during a given timeframe without providing any exact dates for when they expect the next risk to manifest itself. The main difference is that for these fields, there is more objective evidence available about the likely probabilities, whereas AI timeline forecasting needs to rely on fuzzier measures, mostly expert opinion.

It is also true that the exact nature of superintelligence is often left somewhat open. However, *Superintelligence* does devote a chapter to various paths to superintelligence (such as AI, whole brain emulation, biological cognition, etc.) as well as another chapter to describing what superintelligence might mean in practice. Previously, there have also been papers discussing the concrete mechanisms by which various kinds of digital minds could develop to have a major advantage over humanity.[11]

The second "dogma" described by Floridi is "humanity runs a major risk of being dominated by such superintelligence." This is mostly an accurate characterization, even if calling it a "dogma" seems unjustified.

One such argument is provided in *Superintelligence*. *Superintelligence* outlines a takeover scenario by which a sufficiently intelligent AI could establish a great deal of control over the planet. It also argues for two other theses. The first is the orthogonality thesis, according to which high intelligence is, in principle, compatible with any kind of goal, including ones indifferent to human well-being. The second thesis is the instrumental convergence thesis, according to which some goals—such as self-preservation and resource acquisition—are useful for the attainment of a wide range of goals and motivations.

Thus, a sufficiently intelligent AI might be capable of taking over the planet, and it could be indifferent to many human values. And even if its intrinsic goals did not directly require it, it would be likely to have instrumental reasons to "acquire an unlimited amount of physical resources and, if possible, to eliminate potential threats to itself and its goal system"—including humans. While this is not an ironclad argument, it would seem plausible enough to be worth considering, rather than simply dismissed as unjustified dogma.

The third "dogma" given by Floridi is that "a primary responsibility of the current generation is to ensure that the Singularity either does not happen or, if it does, it is benign and will benefit humanity." It is true that many works in the field consider the prevention of AI risk a very important priority—for example, *Superintelligence* describes "the reduction of existential risk" as "our principal moral priority." "Existential risks"—threats that could cause our extinction or destroy the potential of Earth-originating intelligent life—also include other possible threats.[12] As can be seen in some of the popular sentiments involved in stopping global warming or nuclear proliferation, the notion of humanity's survival as a moral priority is not restricted to AI safety advocates. Arguably, people and organizations involved in campaigning against global warming are making a much stronger claim of the current generation having a moral priority to act than are the AI safety advocates, with their looser timeframes for the development of AI.

Floridi goes on to consider AI risk scenarios "ludicrously implausible," with little supporting arguments besides a general appeal to incredulity and a criticism of whether

Moore's law can continue. It is true that Moore's law is occasionally invoked as an additional reason for why AI might become dangerous, but major works in the field do not assume that it would necessarily continue. *Intelligence Explosion – Evidence and Import* explicitly notes that it does not assume "the continuation of Moore's Law, nor that hardware trajectories determine software progress, nor that faster computer speeds necessarily imply faster 'thought' [. . .] nor indeed that AI progress will accelerate rather than decelerate." When *Superintelligence* mentions Moore's Law, it notes that "one cannot bank on this rate of improvement continuing up to the development of human-level machine intelligence." Finally, *Responses to Catastrophic AGI Risk* does not mention Moore's Law at all, other than to note that its continuation "depends on the existence of a small number of expensive and centralized chip factories, making them easy targets for regulation."[13]

Finally, Floridi suggests that the main risk is not the appearance of superintelligence, but the misuse of more conventional digital technologies. While I disagree with him on the need to worry about superintelligence, I agree with him on conventional digital technologies certainly posing their own dangers as well. Work on avoiding the risks from superintelligence and more conventional technologies need not be mutually exclusive. There is currently only a very small number of people working full time on the risks from superintelligence, far fewer than there are people working full time on other risks such as pandemics. Effort put into protecting humanity from pandemics has not prevented other people from working on various issues of the digital era. Similarly, work focused on the implications of advanced AI can proceed without impacting the work done on other worthy causes.

NOTES

1. Floridi, "Singularitarians, Altheists, and Why the Problem with Artificial Intelligence is H.A.L. (Humanity At Large), not HAL."
2. Some early works in the recent research tradition are Yudkowsky, "Artificial Intelligence As a Positive and Negative Factor in Global Risk," and Bostrom, "Ethical Issues in Advanced Artificial Intelligence." Some isolated papers have also discussed the issue as far back as in the 1960s.
3. Russell and Norvig, *Artificial Intelligence*, Section 26.3 in the third edition.
4. http://futureoflife.org/ai-open-letter/
5. Floridi, "Singularitarians, Altheists, and Why the Problem with Artificial Intelligence is H.A.L. (Humanity At Large), not HAL," 8.
6. Müller and Bostrom, "Future Progress in Artificial Intelligence: A Survey of Expert Opinion."
7. Armstrong and Sotala, "How We're Predicting AI – or Failing To."
8. Chalmers, "The Singularity: A Philosophical Analysis."
9. Muehlhauser and Salamon, "Intelligence Explosion: Evidence and Import."
10. Sotala and Yampolskiy, "Responses to Catastrophic AGI Risk: A Survey."
11. Sotala, "Advantages of Artificial Intelligences, Uploads, and Digital Minds"; Muehlhauser and Salamon, "Intelligence Explosion: Evidence and Import."
12. Bostrom, "Existential Risks."
13. Referencing Branwen, "Slowing Moore's Law: How It Could Happen"

BIBLIOGRAPHY

Armstrong, Stuart, and Kaj Sotala. "How We're Predicting AI – or Failing To." In *Beyond AI: Artificial Dreams*, Pilsen, November 5-6, 2012, 52–75. Pilsen: University of West Bohemia, 2012.

Bostrom, Nick. "Ethical Issues in Advanced Artificial Intelligence." In *Cognitive, Emotive, and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, vol. 2. International Institute of Advanced Studies in Systems Research and Cybernetics, 2003.

Bostrom, Nick. "Existential Risks." *Journal of Evolution and Technology* 9, no. 1 (2002).

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Italy: Oxford University Press, 2014.

Branwen, Gwern. "Slowing Moore's Law: How It Could Happen." 2012/2015. http://www.gwern.net/Slowing%20Moore%27s%20Law

Chalmers, David. "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies* 17, no. 9-10 (2010): 7–65.

Floridi, Luciano. "Singularitarians, Altheists, and Why the Problem with Artificial Intelligence is H.A.L. (Humanity At Large), not HAL." *APA Newsletter on Philosophy and Computers* 14, no. 2 (2015): 8–11.

Muehlhauser, Luke, and Anna Salamon. "Intelligence Explosion: Evidence and Import." In *Singularity Hypotheses*, 15–42. Springer Berlin Heidelberg, 2012.

Müller, Vincent and Nick Bostrom. "Future Progress in Artificial Intelligence: A Survey of Expert Opinion." In *Fundamental Issues of Artificial Intelligence*, edited by Vincent Müller. Berlin: Springer, 2014.

Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*, 3rd edition. Pearson, 2009.

Sotala, Kaj. "Advantages of Artificial Intelligences, Uploads, and Digital Minds." *International Journal of Machine Consciousness* 4, no. 01 (2012): 275–91.

Sotala, Kaj, and Roman V. Yampolskiy. "Responses to Catastrophic AGI Risk: A Survey." *Physica Scripta* 90, no. 1 (2015): 018001.

Yudkowsky, Eliezer. "Artificial Intelligence as a Positive and Negative Factor in Global Risk." In *Global Catastrophic Risks*. Oxford: Oxford University Press, 2007.

The Unavoidable Charm of the Superintelligence, and Its Risk

Federico Gobbo

UNIVERSITY OF AMSTERDAM

Readers of the APA newsletter are used to speculation and theoretical debates, being philosophers. The last one is the fierce attack of Floridi and the defense of Sotala to the debate about the future of AI and the theoretical possibility of Singularity, Superintelligence, or AI+ (mainly Chalmers), according to the different authors. Is a truly autonomous, morally independent, (bio)mechanical being that can control our digital technologies against us plausible? In short, Floridi argues that it is theoretically possible, but so implausible that it is not worth spending a word on it—of course, he has to spend some words in order to say it, with is somehow paradoxical. And his text calls for reactions, as the advocates of singularity are treated as if they were members of a sect. Sotala adheres to the wording used in Bostrom's book, who—not by chance—uses the word "Superintelligence" instead of "Singularity."

I invite the reader to take a step backwards and look at this debate with more distance. Let us try to recall what we have learned from the history of ideas in AI. Unfortunately, the tradition of AI is sometimes forgotten in such debates

because scholars are urged to quote recent papers and recent authors. We lose our past; we lose our memory. Floridi underlines the proximity between the Singularitarians and Hollywood. I want to extend his metaphor telling that, in my view, this debate is like a new movie with an old plot, like a reboot of a classic of science fiction. In the old days, the debate was about the plausibility of Good Old Fashioned Artificial Intelligence (GOFAI). I tried to read the main positions in this debate, but I failed to find something new. As in any good reboot, some details are different, but the core message is not. What is the concrete result of the debate about GOFAI? Essentially, AI has lost credits because of this speculation. The concrete, operative results of research came from the so called "weak AI," which, in short, rejects all the theoretical problems of true AI as uninteresting or pointless (as Floridi says), adopting an *a posteriori* perspective: an artificial agent which shows intelligent *behavior* can be considered intelligent, regardless if the *process* behind its behavior is really intelligent.

I argue that the point is that the risk we are facing now is a new discredit of AI. But (weak) AI is more and more present in our daily lives than before. That is why I signed the open letter published in 2015 within the charity Future of Life about the research priorities for "robust" (an internal feature with epistemological consequences) and "beneficial" (a moral concern, as it addresses humankind) artificial intelligence. And I can guarantee to the readers that I do not adhere to any church, Singularitarians and Atheists—to use Floridi's terms—included. Sotala mentions that letter as if the whole debate about the plausibility of GOFAI/Singularity were supported by that. Well, it is not. It suffices to quote the opening of the letter itself:

> Artificial intelligence (AI) research has explored a variety of problems and approaches since its inception, but for the last 20 years or so has been focused on the problems surrounding the construction of intelligent agents—systems that perceive and act in some environment. In this context, "intelligence" is related to statistical and economic notions of rationality—colloquially, the ability to make good decisions, plans, or inferences.

This definition of "intelligence" comes from the tradition of weak AI, and it *a priori* excludes the debate of GOFAI/Singularity as completely irrelevant. We desperately need moral philosophers collaborating with hard science researchers in order to achieve the goal of beneficial AI. Now. Possibly, short-termed. It is completely irrelevant the speculations of researchers in the field in the long-term, mentioned by Sotala: experience shows that even great minds playing with the game of futurology ultimately proved to be completely wrong. But there is a more urgent consideration to be made in this sense. As Keynes said, in the long run we are all dead. The risks we are facing are today, not tomorrow: a badly designed multi-agent system can be a disaster when applied to a large scale, interacting with human beings in an unpredicted manner.

I think that the main risk inside the Superintelligence is the risk of losing the focus on the real problems. But then, why are so many people worried? What is the explanation for it? I have my own opinion on that. The computational turn tremendously complexified our lives. We, human beings, fear complexity because we feel that we are losing our control on reality. The reaction is to look for a single reference point where all relevant causes can be addressed. And here it is: Superintelligence, an Orwellian Big Brother that controls everything. A *single* artificial mind. After all, many among us still did not learn the lesson of the Internet, which is *a network with no central point* that controls everything.

I invite all researchers, especially the younger, to devote their energies to the real problems of artificial intelligence in our contemporary world, letting speculation into the realm of science-fiction literature and Hollywood movies.

Some comments on Luciano Floridi's The Ethics of Information

Jacques Bus
SG DIGITAL ENLIGHTENMENT FORUM

Many of us ask ourselves how we have to understand and live in a world with an increasing number of autonomous technical information systems and a society that through digitization reaches levels of complexity that seem to make our democratic and ethical rules and institutions unfit for their tasks.

Luciano Floridi (LF) has done an impressive job addressing these problems in his book *The Ethics of Information*. His philosophical approach is, for me (mathematician of origin), refreshing in the sense that his thinking is built up in highly analytical terms. He explains mathematical concepts like "level of abstraction," "complex and self-emergent systems," and the concept of "entropy" from thermodynamics, which is also used (but differently) in classic information theory.

The introduction of the term "metaphysical entropy" and how this is used to define four ethical principles of Information Ethics (IE) did raise questions for me. As I was reading, I sometimes asked myself where this could lead. The problem of ethics in general does not particularly lend itself to a mathematical or quantifying approach. However, a clear and satisfying answer follows on page 315 as a response to some of the criticisms. Floridi states there:

> IE is equally reasonable: fighting the decaying of *Being* (metaphysical entropy) is the general approach to be followed, not an impossible and ridiculous struggle against thermodynamics, or the ultimate benchmark for any moral evaluation, as if human beings had to be treated as mere numbers.

So good, ethical behavior is fighting the decaying of being or, in LF's terms, fighting the decrease of metaphysical entropy in the overall system.

The concept of Global Information-Ethics, developed throughout the book, is not simply defined in a few sentences without risking wrong interpretations. The interested person will have to read to the end. An important aspect is that the actors in the ethical space are not restricted to human agents and patients, but include all information entities; hence, also non-intelligent objects and creatures, autonomous technical systems, organizations or communities, etc.

A second important issue is the proposal to develop a global informational ontology for a global digital world.

I am not an ethicist, nor a philosopher. Hence I cannot judge the book on its value for those scientific communities. My interests lie in the interaction between digitization and society (with the individuals living in it) and the policy consequences and requirements. What interests me most is the effort to come to a theory of ethics that might form a practical basis for policy development concerning this interaction to the benefit of humankind.

In the last chapter (chapter 15) of his theoretical development process, Floridi addresses the concepts "Physis" and "Techne" and argues the necessity to develop a successful ecological relationship ("marriage") between the two. In LF's view, such "marriage" is vital, and failing to negotiate a global, fruitful, synthetic relationship between technology and nature is not an option. I do not think this is a new view. LF himself refers to techno-philosophers promoting similar views. Nevertheless, his conclusion that for this negotiation philosophy and Information Ethics as he develops it can help, might be worth further exploration. As Secretary General of Digital Enlightenment Forum (DEF) (see www.digitalenlightenment.org), I see the synergetic relation between nature (human) and technology as the reason of existence of DEF. Exploring and developing this relation in a balanced way is essential in my view for humankind.

Finally, I cannot neglect his reference in chapter 15 to the concept of "Social Contract" (a topic that is also in the middle of DEF's discussions) and the analogy he makes with ontic trust. As LF argues, a social contract may be an implicit or merely hypothetical agreement between parties constituting a society, but in general it tends to be highly anthropocentric. However, we might need to include the role of Artificial Intelligence and Information Systems. LF concludes:

> In the case of ontic trust, it is transformed into a primeval, entirely hypothetical *pact*, logically predating the social contract, that all human ( ) agents cannot but sign when they come into existence, and that is constantly renewed in successive generations.

Understanding and exploiting this effectively might be our most important task to ensure sustainability of the societies we live in.

I will not cover the many other aspects LF covers in his book and which are worth contemplating. It is highly recommended reading for anybody who has a keen interest in how in a global environment our society and its ethics may evolve *unisono* with technology.

Comment on Floridi's The Ethics of Information

David Chapman
THE OPEN UNIVERSITY

Concern about the nature of national identity is very much to the fore in the United Kingdom at present, with debates about the membership of the European Union, the rise of nationalism across the whole of Europe, and an independence referendum in Scotland. Identity is inherently an informational issue, so after reading *The Ethics of Information* I found myself wondering whether the framework that it puts forward, Information Ethics (IE), offers any insights into the moral dilemmas around statehood.

For example, in a satirical radio program entitled "How to Define Oneself in Terms of Regional, Cultural, and Geopolitical Identity Without Tears" (BBC, 2014) the comedian Jeremy Hardy said (talking about Israel at this point):

> [O]ne of the frequent demands of [Israel's] government is that others recognise its right to exist. I'm not sure any state has rights. Whether a person has rights is a moral question . . . but at least ethical judgements apply more sensibly to human beings than to geopolitical entities. We'd all say a person has a right to a home. We wouldn't say the home has rights. No one thinks a house has a right to exist.[1]

Information Ethics (IE) as presented by Luciano Floridi explicitly disagrees with Jeremy Hardy. Both houses and states, as Informational Entities, do have rights.

In IE:

> all entities are informational entities, some informational entities are agents, some agents are artificial, some artificial agents are moral, and moral artificial agents are accountable but not necessarily responsible.

Furthermore:

> all entities *qua* informational entities have an intrinsic moral value, although possibly quite minimal and overridable, and hence that they qualify as moral patients subject to some (possibly equally minimal) degree of moral respect.[2]

In the language of IE, both states and houses qualify as moral patients (receivers of moral actions) and as such are entitled to some degree of moral respect. A state (but not a house) is also a moral agent (an entity which can do things with ethical significance) which is accountable though not necessarily responsible. The motivation for IE lies in the need to address ethical issues emerging as a result of technological advances, especially (but not exclusively) computing and ICT, but IE is a new universal macroethic applicable to everything—including states and houses. Floridi's contention is that rather than addressing the ethical dilemmas arising from new technologies piecemeal with existing ethics, a new universal macroethic is needed.

The ethics of information is the second in Floridi's planned series of five books on the philosophy of information, and this volume comprehensively develops and explores Information Ethics. Information Ethics is based on perceiving the whole of reality in terms of informational entities with certain rights and responsibilities. The fundamental good in IE is existence, being, as an informational entity, so morality is assessed in terms of creation, destruction, or harm to informational entities. The book is a formidable piece of work, and the only thing that this brief comment can do is to pick out a few features that caught my attention for one reason or another.

The idea of a house having rights may be surprising, but when Floridi says that "all entities are informational entities," it is important to appreciate the significance of this claim. Quite how universal IE is cannot be overstated: people, animals, plants, books, paintings, even stones are informational entities and within the scope of Information Ethics. Such an all-encompassing ethic is presented by Floridi as the logical conclusion of a progression which has previously got as far as the environment in environmental ethics, but in IE finally extends to all informational entities in what he refers to as e-nvironmental ethics (Floridi uses a lot of neologisms).

Informational entities are entities that exist in what he calls the infosphere (another of his neologisms), and the infosphere is created through the method of (levels of) abstraction. Floridi insists on understanding the nature of IE ontologically, rather than epistemologically, but this is something I feel needs more discussion.

Floridi says that we should "shift our perspective on information from one that is exclusively epistemological and semantic (information about something or for something) to one which is *also* ontological (information as something)."[3] If we start talking of information as something, however, we have to be very careful not to be lulled into thinking of it as some alternative 'substance' and expecting it to behave in a way analogous to other substances like matter or energy. Floridi has argued elsewhere for the "genetic neutrality" of information, meaning that information does not need an informee.[4] He illustrates the idea by reference to the Rosetta Stone, arguing that writing on the stone was known to be information even before it could be understood. However, the writing was believed to be information because it was assumed that there was a time when it could read by someone, even if only by the writer. An informee was needed when the information was created, and anyway by the action of trying to decode the writing, the audience projects meaning on to the symbols. An alternative perspective on information is, therefore, that it only exists within a network or within a communication. Information is "meaningful data" (a definition supported by Floridi when he uses the General Definition of Information, GDI[5]), but meaning is inherently semiotic: It has to be meaning for someone or something in a context. (This is not necessarily a person. It could be meaning for an artificial agent.) It may be significant that a scan of the references in both *The Ethics of Information* and *The Philosophy of Information* reveal an impressive range of sources including technical papers from computing journals (such *Communications of the ACM*) but few from journals in the field of communications engineering, other than the foundational work of Claude Shannon. Similarly, he draws from the work of cybernetics in the writing of Norbert Wiener, but does not consider the work of Gregory Bateson who emphasized information as about difference and relationships.[6]

The perception of information as substance appears in the chapter 1 discussion of "The Zettabyte Era," in which Floridi presents superlatives about the amount of data/information around today. He includes, for example, a quote saying that "in 2011, the amount of information created and replicated will surpass 1.8 zettabytes."[7] Although there are extraordinary things happening with the volume of text, sounds, and images appearing in digital media on- or offline, we need to be careful about assuming that bits of information can be counted as though you were counting coins. Money provides a good example, because a government could print more notes and mint coins and then say that there is more money in the country, but we know from the experience of the Weimar Republic (among others) that, in reality, inflation would devalue the currency with the result that there would not in any meaningful sense be more money around. This is because money is information, and the value of the money comes not from the number of notes or coins in circulation, but from much more complicated networks of the nation's (and the world's) economy. In his own writing on the zettabyte era (as opposed to the Gantz and Reisel quote), Floridi refers to data rather than information, so it may be that he is talking about the equivalent of the notes and coins rather than the money. If that is the case, however, it rather spoils the point of the discussion, which is about information.

This does not mean that information is not ontological, but that our understanding of its ontological nature has to include communication and semiotics, and this has an impact on the nature and application of IE. For example, one of Floridi's claims about the merits of IE is that, unlike standard ethics, it can be used to address the moral claims of a dead body. He demonstrates human respect for a dead body by the story in the *Iliad* of Achilles' treatment of Hector's body. Another story that sheds light on the status of a corpse is Gabriel Garcia Marquez's *The Handsomest Drowned Man in the World* in which the corpse of an unknown man washed up on a beach transforms a village through the imagination of the villagers.[8] In both stories the information associated with the corpse exists in the network of other actors in the narrative. In the case of the

Iliad, this is the meaning that was generated in the life of Hector. In *The Handsomest Drowned Man*, there is no pre-existing knowledge so the information is created only when the dead body is discovered.

Whatever the ontological nature of information, the value of IE as a universal macroethic in addressing issues of the age is impressive, and to finish, I would like to give a flavor of what can be done with IE by mentioning three examples. The first, to do with privacy, is about the rights of moral patients; the second, to do with artificial agents, is about moral accountability and responsibility; and the third, to do with constructionism, is about the proactive responsibilities of moral agents.

IE is not merely descriptive, but also normative and offering a prescription for behavior that is morally right or wrong. Evil in IE is the damage to informational entities through what Floridi calls metaphysical entropy, which he emphasizes is different from, though analogous to, thermodynamic entropy. I was struck by the approach to informational privacy, which argues that personal information is a constitutive part of someone's personal identity and individuality so that trading in some kinds of personal information should be seen as equivalent to trading in human organs. This is an example of how ontologizing information helps, because it emphasizes the reality of personal information which might otherwise be perceived to be of less importance than physical parts of the body.

In addition to addressing the responsibilities towards informational entities as moral patients, the book covers the moral responsibilities of informational agents. As technological agents become more advanced and more ubiquitous, it becomes increasingly important to have a framework for dealing with situations in which an artificial agent causes evil. Floridi argues that artificial agents should be considered as moral agents and can be considered to be morally accountable for their actions. There is an important distinction, though, between accountability and responsibility, and the responsibility still falls on humans. In keeping with the universality of IE, artificial agents are not restricted to technological agents, but can also be social agents such as companies, states, or hybrid systems formed by humans and machines, or technologically augmented humans.

The opposite of entropy is construction (poiesis). To be good agents in the infosphere, therefore, we should be combatting metaphysical entropy but also fostering poiesis. Floridi argues that constructionism, the urge to create physical and conceptual objects is one of *homo sapiens*' secondary needs (i.e., coming after the primary needs for food, shelter, security, and reproduction are met), and he coins another neologism: *homo poieticus*.

A 1993 opinion piece in *Physics Today* presented a list of things a physicist "really needs to know," and one that stuck in my mind was "Go for the big problems. No one cares about publishable petty results."[9] Perhaps it is in the nature of the field that philosophers will be tackling fundamental problems, but Luciano Floridi has certainly gone for the big ones: reontologizing the whole of reality and the creation of a new universal macroethic addressing the nature of good and evil. Personally, I applaud the work. As I write this review, the news reports drone attacks in Iraq, yet more stories of online harassment, continuing concerns about privacy, debates about the power of Internet corporations, and more; the list is endless. An ethical framework that helps us navigate through this treacherous and unknown territory is urgently required.

NOTES

1. "How to Define Oneself in Terms of Regional, Cultural, and Geopolitical Identity Without Tears," BBC Radio, London, September 24, 2014.
2. Luciano Floridi, "The Ethics of Information" (Oxford: Oxford University Press, 2013), 109.
3. Ibid., 28.
4. Luciano Floridi, "The Philosophy of Information" (Oxford: Oxford University Press, 2011), 91.
5. Ibid.
6. See, for example, Bateson's *Steps to an Ecology of Mind* (London: Intertext Books, 1972), which also contains a detailed exploration of the cybernetics of self with relevance to Floridi's discussion of *The Informational Nature of Selves* in chapter 11.
7. John Gantz and David Reisel, "Extracting Value from Chaos," White paper, sponsored by EMC-IDC, 2011, http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf Quoted in Floridi, "The Ethics of Information," 5.
8. Gabriel Garcia Marquez, "The Handsomest Drowned Man in the World," no date, http://www.utdallas.edu/~aargyros/hansomest.htm
9. David P. Stern, "All I Really Need to Know," *Physics Today* 46, no. 5 (1993): 63.

BIBLIOGRAPHY

Bateson, Gregory. "Steps to an Ecology of Mind." London: Intertext Books, 1972.

BBC. "How to Define Oneself in Terms of Regional, Cultural, and Geopolitical Identity Without Tears." Radio programme, BBC Radio 4, London, September 24, 2014.

Floridi, Luciano. "The Philosophy of Information." Oxford: Oxford University Press, 2011.

Floridi, Luciano. "The Ethics of Information." Oxford: Oxford University Press, 2013.

Gantz, John, and David Reisel. "Extracting Value from Chaos." White paper, sponsored by EMC-IDC, 2011. http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf Accessed January 10, 2014.

Garcia Marquez, Gabriel. "The Handsomest Drowned Man in the World." Undated. http://www.utdallas.edu/~aargyros/hansomest.htm Accessed 28 September 2014

Stern, David P. "All I Really Need to Know." *Physics Today* 46, no. 5 (1993): 63.

A Call for More Philosophy in the Philosophy of Computer Science

Robin K. Hill
UNIVERSITY OF WYOMING

When I was a programmer, and even when I was a teacher of programming, I would wonder about where data

structures and algorithms came from, and to what extent people use pass-by-value, pass-by-reference, and pass-by-name for communication in daily life, and whether we use those techniques for activities besides communication, and whether the universe somehow uses those techniques, and many other speculations on the matters of computer science as they might be viewed by philosophy. In the workplace and in the university, opportunities to articulate and consider such questions are rare. While the philosophies of mathematics, science, and logic are well developed, the philosophy of computer science *per se* is only emerging, and the work that does appear commonly applies the computational paradigm to the questions of philosophy. Let's also apply the philosophical paradigm to the objects of computer science.

In the current philosophy of computer science, many (insightful) investigations are rooted in the formal realm, the realm of discrete objects, of definitions and theorems, of formulae and logic. Topics such as programming language paradigms, software specification and implementation (and syntax and semantics), and the mechanisms of artificial intelligence and knowledge representation underlie much of the philosophical offerings, as delineated in the "Philosophy of Computer Science" entry in the Stanford Encyclopedia of Philosophy [Turner 2013]. The current philosophy of computer science, in other words, is cultivated largely within artificial structures and languages formulated by the researcher, and largely focused on programs, with some papers relying on the tacit assumption (perhaps a holdover from the early days of artificial intelligence) that, once we have rendered something as a formal system, we have explained it. Researchers such as Bill Rapaport [Rapaport 2005] and Luciano Floridi [Floridi 2011] both reach farther, extending questions of knowledge, information, and computability into the realms of epistemology and metaphysics, while sometimes deploying formal theories for their solutions. In each of the traditional areas of philosophy—metaphysics, ethics, epistemology, and aesthetics—questions about computer science, often in terms of its relationship with aspects of society and humanity, are already under active consideration by a growing community, embodied in organizations abroad such as the International Association for Computing and Philosophy (IACAP), the American Philosophical Association Committee on Philosophy and Computers, and the Commission on the History and Philosophy of Computing (HaPoC), along with the periodicals that publish related work. This is good stuff. But we can do even more.

Rather than confining ourselves to the interpretation of philosophical material in terms of computing, let's also interpret the material of computing in terms of philosophy. I myself have contributed an analysis of algorithms that looks outside of their implementation as digital programs [Hill 2015]. There is much more to investigate: What about data structures? Do they exist in the world? I myself am working on this question. Do arrays, for instance, exist in nature? If not, what is the closest thing; what are the natural phenomena that correspond to arrays, or linked lists, or the other abstract data types that we find useful? How far do we have to generalize to find a structural and functional equivalence? And what is the pragmatic function of a given abstract data type—organization, control, homogeneity? How is it that these things serve us?

We can apply philosophy to the things exposed in a computer lab via such questions, just as we apply philosophy to phenomena exposed by other social and scientific developments (phenomena such as choice; law versus justice; the scientific method; feminism). Whereas researchers outside the field like to treat computing instrumentally, as a tool in the investigation of traditional philosophical concepts, computer scientists can contribute the questions that are salient in practice. A database designer, trying to write down the entities and attributes for some enterprise, may wonder, "Why is this so hard?" A data scientist might take a moment to ponder how much data we should produce; how much can we stand? How much makes sense? What does it mean to make sense? If data is a resource, as modern management views would have it, should it undergo the same oversight as other resources? Can it be recycled or repurposed? Conserved? Or is it rather a consumer of cognitive resources that call for wise allocation, a liability rather than an asset? These are questions of metaphysics and values.

To identify more questions, we can turn the computer instrumentality inside out. For example, the Indiana Philosophy Ontology Project offers its organization of philosophy research subjects via a thing beloved of computer science, an Application Program Interface (API), that allows programmatic processing of its data. While we make good use of that, looking through the hierarchy of subjects, let us also ask: What is the epistemology of a search tree? Now that we are exploring another traditional area of philosophy, epistemology, let's consider the modern knowledge acquisition mechanism of web search. What is search, anyway? What type of epistemic restructuring does a search result engender? How does finding an answer compare to learning a fact in some other way? How does web search differ from library search? We're already wondering about the implications of the use of big data rather than semantics as it is traditionally conceived. How does that play out in terms of different phases or aspects of knowledge? What kind of thing is the Web itself [Monnin 2012]? Peter Boltuc, editor of the *APA Newsletter on Philosophy and Computers*, raises "the question of what the ontological status of web-based objects is" [Boltuc 2008]. Answers rooted in the contexts of virtual environments and biomedical domains appear in a later issue [Andersen 2009, Arp 2009]. Let's have more; there are many contexts.

We can even follow this into aesthetics. What is the nature of the satisfaction that comes from solving a symbolic problem, as in a game or a software design, and how does it relate to the appreciation of other arts (or sciences)? Why is programming fun? Why are there no elegant algorithms for calendar work (or are there, and what do we mean by "elegant"?)? Is it because our calendar is irredeemably ugly, impervious to the clean patterning that we admire? And is that because it was developed incrementally, ad hoc, or because of nature itself? We view the irregularity of nature—with its spots, wrinkles, colors, shreds, and other details that resist clean patterns—as beautiful. Do

we adopt a different standard when we view the pure and precise algorithm as beautiful? What other kinds of things, like the calendar, resist abstraction, which is essential in computer science [Colburn 2007], and what can we learn from data ontologies [Smith 2003]? Are our abstractions more like abstractions in art or abstractions in science (or even some other quotidian realm), and is there some point of confluence?

Here is a provocative question that could involve all of the classical areas of philosophy: Does learning theoretical computer science make us better people? When we learn about different values of infinity, are we also learning about the limits of perspectives—that context makes a local fact look universal—and thereby something about diversity? When we grasp Godel's theorems, do we grasp something profound about the boundaries of intellectual effort, and therefore something profound about humanity? Can formal studies and digital expertise be normative? And how can we pursue such questions (especially one as self-serving as this) without falling into absurdity or incoherence? This is not a bid for fatuous speculation! Standards need not be lowered. Analysis via formal systems is fruitful, certainly, but rigor can be found outside of formalisms in the close and disciplined reasoning on complex questions that is regularly practiced by philosophy.

To anyone already addressing the questions I suggest, I commend you and apologize for overlooking your work. To the many successful computer scientists eager to continue their work without probing into its philosophy, I salute you as well. Keep it up. This is not a demand, not even a plea, but only an invitation.

REFERENCES

Andersen, David Leech. "A Semantics for Virtual Environments and the Ontological Status of Virtual Objects." *APA Newsletter on Philosophy and Computers* 9, no. 1 (2009): 15–19.

Arp, Robert. "Realism and Antirealism in Informatics Ontologies." *APA Newsletter on Philosophy and Computers* 9, no. 1 (2009): 19–22.

Boltuc, Peter. "From the Editor." *APA Newsletter on Computers and Philosophy* 7, no. 2 (2008): 1.

Colburn, Timothy, and G. Shute. "Abstraction in Computer Science." *Minds and Machines* 17, no. 2 (2007): 169–84.

Floridi, Luciano. "The Philosophy of Information." Oxford University Press, 2011.

Hill, Robin K. "What an Algorithm Is." *Philosophy and Technology* 2015, DOI:10.1007/s13347-014-0184-5.

HaPoC website, maintained by the DHST Commission on the History and Philosophy of Computing. hapoc.org. Accessed May 13, 2015.

IACAP, the International Association for Computing and Philosophy. http://www.iacap.org/. Accessed May 13, 2015.

Indiana University Cognitive Science Program, "Indiana Philosophy Ontology Project." https://inpho.cogs.indiana.edu/, accessed February 5, 2016.

Monnin, Alexadre, and Harry Halpin. "Toward a Philosophy of the Web: Foundations and Open Problems." Metaphilosophy 43, no. 4 (2012): 361–79.

Rapaport, William J. "Philosophy of Computer Science: An Introductory Course." *Teaching Philosophy* 28, no. (2005): 319–41. Available at http://www.cse.buffalo.edu/~rapaport/Papers/rapaport_phics.pdf.

Smith, Barry, "Ontology." In Blackwell Guide to the Philosophy of Computing and Information, ed. Luciano Floridi. 2003.

Turner, Raymond. "The Philosophy of Computer Science." The Stanford Encyclopedia of Philosophy (Winter 2014 Edition), ed. Edward N. Zalta. http://plato.stanford.edu/archives/win2014/entries/computer-science/. Accessed May 13, 2015.

ANNOUNCEMENT

Award for Ongoing Doctoral Dissertation Research in the Philosophy of Information

1. NATURE OF THE AWARD

1.1 The award shall consist of $1,000 given annually to a graduate student who is working on a dissertation on the philosophy of information (broadly construed). As we see it, the range of philosophical questions relating to information is broad and approachable through a variety of philosophical traditions (philosophy of mind, logic, philosophy of information so-called, philosophy of science, etc.).

2. PURPOSE OF THE AWARD

2.1 The purpose of this award is to encourage and support scholarship in the philosophy of information.

3. ELIGIBILITY

3.1 The scholarship recipient must meet the following qualifications:

(a) Be an active doctoral student whose primary area of research is directly philosophical, whether the institutional setting is philosophy or another discipline; that is to say, the mode of dissertation research must be philosophical as opposed to empirical or literary study;

(b) Have completed all course work; and

(c) Have had a dissertation proposal accepted by the institution.

3.2 Recipients may receive the award not more than once.

4. ADMINISTRATION

4.1 The Litwin Books Award for Ongoing Doctoral Dissertation Research in the Philosophy of Information is sponsored and administered by Litwin Books, LLC, an independent scholarly publisher.

5. NOMINATIONS

5.1 Nominations should be submitted via email by June 1, to award@litwinbooks.com.

5.2 The submission package should include the following:

(a) The accepted dissertation proposal;

(b) A description of the work done to date;

(c) A letter of recommendation from a dissertation committee member; and

(d) An up-to-date curriculum vitae with current contact information.

6. SELECTION OF THE AWARDEE
6.1 Submissions will be judged on merit with emphasis on the following:

(a) Clarity of thought;

(b) Originality;

(c) Relevance to our time; and

(d) Evidence of good progress toward completion.

7. NOTIFICATION
7.1 The winner and any honorable mentions will be notified via letter by July 1.

ADVISORY BOARD
Jonathan Furner, Graduate School of Education and Information Studies, UCLA

Ron Day, School of Library and Information Science, Indiana University

Melissa Adler, College of Communication and Information, University of Kentucky

PAST WINNERS
2015: Quinn DuPont, of the University of Toronto Faculty of Information, for his dissertation précis titled "Plaintext, Encryption, and Ciphertext: A History of Cryptography and its Influence on Contemporary Society."

2014: Patrick Gavin, of the University of Western Ontario FIMS, for his dissertation propsoal titled "On Informationalized Borderzones: A Study in the Politics and Ethics of Emerging Border Architectures."

2013: Steve McKinlay, of Charles Sturt University, New South Wales, Australia, for his dissertation proposal titled "Information Ethics and the Problem of Reference."

Rory Litwin
Library Juice Academy
Library Juice Press
Litwin Books, LLC
PO Box 188784, Sacramento CA 95818
Tel. 218-260-6115
http://libraryjuice.com/
http://rorylitwin.info/

CALL FOR PAPERS

It is our pleasure to invite all potential authors to submit to the *APA Newsletter on Philosophy and Computers*. Committee members have priority since this is the newsletter of the committee, but anyone is encouraged to submit. We publish papers that tie in philosophy and computer science or some aspect of "computers"; hence, we do not publish articles in other sub-disciplines of philosophy. All papers will be reviewed, but only a small group can be published.

The area of philosophy and computers lies among a number of professional disciplines (such as philosophy, cognitive science, computer science). We try not to impose writing guidelines of one discipline, but consistency of references is required for publication and should follow the *Chicago Manual of Style*. Inquiries should be addressed to the editor, Dr. Peter Boltuc, at pboltu@sgh.waw.pl